SYSTEMATIC OBSERVATION: ENGAGING RESEARCHERS IN THE STUDY OF DAILY LIFE AS IT IS LIVED

EDITED BY: M. Teresa Anguera, Angel Blanco-Villaseñor, Gudberg K. Jonsson, José Luis Losada and Mariona Portell

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

SYSTEMATIC OBSERVATION: ENGAGING RESEARCHERS IN THE STUDY OF DAILY LIFE AS IT IS LIVED

Topic Editors:
M. Teresa Anguera, University of Barcelona, Spain
Angel Blanco-Villaseñor, University of Barcelona, Spain
Gudberg K. Jonsson, University of Iceland, Iceland
José Luis Losada, University of Barcelona, Spain
Mariona Portell, Universitat Autònoma de Barcelona, Spain

Image: Golubovy/Shutterstock.com

Assessment in natural contexts through observation is unquestionably complex. Systematic observation grounded in observational methodology offers a wide range of possibilities to the rigorous study of everyday behavior in their natural context. These possibilities have been enriched in recent decades with the explosion of information and communication technologies. In this eBook we assemble 23 articles from several researchers who have made important contributions to this evolving field. The articles included in this eBook has been organized with a first part on general methodological developments and a second part with methodological contributions

that emphasize different application areas. Considering the enormous possibilities of the systematic observation in the study of daily life, we hope this eBook will be useful to understand innovative applications in different fields.

Citation: Anguera, M. T., Blanco-Villaseñor, A., Jonsson, G. K., Losada, J. L., Portell, M., eds. (2019). Systematic Observation: Engaging Researchers in the Study of Daily Life as It Is Lived. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-962-9

Table of Contents

Editorial: Systematic Observation: Engaging Researchers in the Study of Daily Life as It Is Lived

M. Teresa Anguera[1]*, Angel Blanco-Villaseñor[2], Gudberg K. Jonsson[3], José Luis Losada[2] and Mariona Portell[4]

[1] Faculty of Psychology, Institute of Neurosciences, University of Barcelona, Barcelona, Spain, [2] Faculty of Psychology, University of Barcelona, Barcelona, Spain, [3] University of Iceland, Reykjavik, Iceland, [4] Department of Psychobiology and Methodology of Health Sciences, Faculty of Psychology, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

Editorial on the Research Topic

Systematic Observation: Engaging Researchers in the Study of Daily Life as It Is Lived

The Research Topic *Systematic observation: Engaging researchers in the study of daily life as it is lived* (Section *Quantitative Psychology and Measurement*) faithfully reflects the interest of many researchers to conduct studies based on a methodology that is essentially characterized by being highly flexible and rigorous, and that aims to capture reality as it happens when studying it scientifically.

The analysis of a part of reality is complex, given the multifaceted nature of any aspect of daily life. This complexity is manifested in numerous aspects to be considered, from the initial filtering that must be done to conveniently delimit the part of reality to be studied, to the structuring of the ideal observational design, the construction of a customized instrument that allows to properly channel all the behaviors/elements to be observed around the axes or dimensions around which the studied reality pivots, the materialization of a suitably coded record, the management of the records, data quality control, its subsequent analysis, and the interpretation of the results.

This process is none other than the scientific method, although adapted to the reality of natural situations, in which it is not possible or convenient to apply the control that other methodologies offer, given that the spontaneity of behavior and the habituality of the context are of primary concern. In a schematic way, we could say that in the systematic observation the face and cross of a same coin are focused: "Face" because its proven versatility and adaptability make it extremely interesting and demanded in innumerable situations, and "cross" because the rigor of the own scientific method gives it a very estimable value that prestige such studies, as shown in the scientific community. The fact that *Frontiers in Psychology*, a prestigious scientific journal of high visibility around the world, had accepted the proposal of this Research Topic, gives more body and arguments in favor of *systematic observation* to the scientific community.

The 23 articles that make up the Research Topic *Systematic observation: Engaging researchers in the study of daily life as it is lived* are organized from a substantive point of view in different criteria, although each of the published articles could have been "classified" from several points of view.

METHODOLOGICAL DEVELOPMENTS

Given the incessant development of systematic observation over the last quarter of a century, this Research Topic has been an occasion to complete aspects that required progress, in order to be taken as a point of reference in future studies and new developments.

Basically, we differentiate four different aspects that mark a procedural path, and we consider that they are the following:

On the one hand, the conceptualization, development and analysis possibilities of indirect observation (Anguera et al.), which is strongly emerging in recent years, and revealing a wide field of application. More and more researchers are obtaining texts, either by transcribing oral conduct, or by applying new forms of communication, or by direct writing, and they are provided with a procedure to follow.

The weak current of systematic observation has traditionally been the psychometric, and the article by Chacón-Moscoso et al. focuses, in an applied study, on the measurement of quality of observational studies based on content validity, and taking advantage of the possibilities of the Osterlind index.

One aspect that to date had been absent in systematic observation studies is simulation. Manolov and Losada, offer a computer application developed for this purpose, adaptable to different sampling techniques.

And we can also consider as methodological development the work of Izquierdo and Anguera, centered on the notational development of movement, and for which a structured system of rules and symbols is proposed.

AREAS OF APPLICATION

The studies that we publish in this section stand out for both substantive and procedural aspects within systematic observation, but we have considered that the emphasis that they represent at the level of application areas was the most important:

a) Sport

This field presents special characteristics that make systematic observation extremely suitable and attractive as the scientific procedure to be followed.

At the same time, priority is given to the interest in professional football in most sport articles published in this Research Topic, although it is important to take into account specific aspects of each one. While in Zurutuza et al. the objective is oriented to an analysis by physiological variables and is intended to study the relationship of external and internal training load indicators with the objective and subjective fatigue experienced by semi-professional football players, it has a different purpose in other studies. In Diana et al. the interest lies in knowing how game-location positively affects the secondary and tertiary level of performance, prioritizing the incidence of playing at home, or in an opponent field, or in a neutral field. Casal et al., on the contrary, studied the identification of factors that may allow predicting success in professional football, focusing specifically on ball possession, and using bivariate and multivariate statistical analysis.

The orientation is very different in the works of Castañer et al. and Maneiro and Amatria. In both the aim is the in-depth study of elite football players, from the intensive perspective. The analysis of polar coordinates is used in both, and in that of Castañer et al., in addition, the detection of T-Patterns, which was also used in the Diana et al. article.

Morillo et al. address the study of referees in handball, also taking advantage of the extraordinary possibilities of systematic observation. As in the articles by Castañer et al. and Maneiro and Amatria, they apply polar coordinate analysis.

With the exception of the Zurutuza et al. article, which requires complementing the observation with physiological variables, all the other studies mentioned here constructed a custom observation instrument.

b) Health Psychology

From a broad definition of Health Psychology, three articles were published in the Research Topic.

On one hand, Sanduvete-Chaves et al. have built a scale that allows measuring the quality of the work climate in emergency services, in which there is usually tension due to the responsibility involved in making decisions and the necessary quickness that is required. In addition to indirect observation, we have worked with questionnaires and surveys.

Arias-Pujol and Anguera is a study of clinical psychology in which the interaction between adolescents in a group therapy has been observed. By means of a polar coordinates analysis it has been possible to analyze the conversation in the therapeutic group, starting from a detailed record obtained by means of a customized observation instrument.

Finally, in Cerezo et al. the influence of parental gender (father/mother) is studied in the interaction with children, and also taking into account the gender of the child; we highlight the study of interaction from the framework of nonlinear dynamic systems.

c) Educational Psychology

Systematic observation also has innumerable advantages in studies that revolve around school and learning. In the Research Topic, five articles were published with very different objectives, but with many common elements regarding the procedure.

On the one hand, Rodríguez-Dorta and Borges set out to study the good practices of teachers who attend students with special educational needs. In Escolano-Pérez et al. the systematic observation is complemented with the selective methodology to evaluate the executive functions of preschoolers and analyze their association with later academic skills, using the general linear model from the analytical perspective. Suárez et al. focuses on the teaching of reading by primary school teachers, and, as in Rodríguez-Dorta and Borges and Escolano-Pérez et al., the theory of generalizability is used. The study realized by García-Fariña et al. uses indirect observation, and focuses on the detection of patterns in the verbal behavior of physical education teachers in the school. The article by Santoyo and Mendoza focuses on the study of coercive patterns in the school context, focusing in particular on the description of stability and change in the behavioral patterns of children identified as victims of bullying.

In the five articles of this group, an *ad hoc* observation instrument was elaborated, although in the work of Santoyo and Mendoza it had been presented in a previous study. In Rodríguez-Dorta and Borges, in García-Fariña et al., and in Santoyo and Mendoza, the analytical technique of lag sequential analysis was used, and in Suárez et al. the T-pattern detection and

analysis was used, as in the articles by Castañer et al., Diana et al., Diana et al., and Pic Aguilar et al.

d) Social Psychology

Taking the term Social Psychology in a broad sense, we refer to the Research Topic articles of this group, which are very diverse among themselves.

Diana et al. focus on the study of deception in social interaction; they combine systematic observation with an experiment, and the T-Patterns detection is applied, as in the articles by Castañer et al., Diana et al., Pic Aguilar et al., and Suárez et al.

Cabrera et al. focus on the objective of studying antisocial behavior, which begins in childhood, remains in adolescence, and continues its escalation during adulthood; this paper explores the social interaction patterns of adolescents, with and without risk of committing antisocial behaviors and over 2 years, in a situation of conversational negotiation about conflicting topics.

Pérez-Tejera et al. is a study of Environmental Psychology that focuses on the study of gender differences in the occupation of public parks; the EXOdES instrument was elaborated, and 35,000 co-occurrences of codes were recorded, with the analysis of polar coordinates being used, as in the articles by Castañer et al., Morillo et al., and Maneiro and Amatria.

e) Motor Game and Gaze Direction

Two articles make up this last group:

On the one hand Pic Aguilar et al. focus on the study of motor games, and specifically in triadic ones, in order to know the regularities that are detected from the observable behaviors. The analysis technique used is the detection of T-Patterns, as in the article by Castañer et al., Diana et al. and Diana et al.

And, on the other, Lappi et al. present as objective an expert driver's gaze behavior in natural driving on a real road, without any instruction; and gaze directionality sequences are obtained in the directionality of the gaze.

CONCLUSIONS

In short, the articles included in the Research Topic make up a broad spectrum.

As Editors of this Research Topic, we want to express the satisfaction that comes from having the opportunity to offer the materialization of new studies in the exciting field of systematic observation to the scientific community.

The Research Topic proposal has been motivating, exciting and satisfying, as well as the highest level of acceptance of the originals. Regarding the management, the originals of the 23 articles that make up this Research Topic were sent from December 2016 to November 2017 and were published between April 2017 and November 2018. The time elapsed from submission to publication has ranged between 3 and 21 months (some were delayed due to the difficulty of finding specialist reviewers in the subject), and 100% of the submitted manuscripts were accepted. Now, the average number of views is 3,100.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

And we thank *Frontiers in Psychology* for having trusted in our proposal of the Research Topic *Systematic Observation: Engaging Researchers in the Study of Daily Life as It Is Lived*, as well as the invaluable help given in the management and editing process of the manuscripts throughout of this period.

Check for updates

Indirect Observation in Everyday Contexts: Concepts and Methodological Guidelines within a Mixed Methods Framework

M. Teresa Anguera[1]*, Mariona Portell[2], Salvador Chacón-Moscoso[3,4] and Susana Sanduvete-Chaves[3]

[1] Faculty of Psychology, Institute of Neurosciences, University of Barcelona, Barcelona, Spain, [2] Faculty of Psychology, Department of Psychobiology and Methodology of Health Sciences, Universitat Autònoma de Barcelona, Barcelona, Spain, [3] Facultad de Psicología, Universidad de Sevilla, Seville, Spain, [4] Departamento de Psicología, Universidad Autónoma de Chile, Santiago, Chile

Indirect observation is a recent concept in systematic observation. It largely involves analyzing textual material generated either indirectly from transcriptions of audio recordings of verbal behavior in natural settings (e.g., conversation, group discussions) or directly from narratives (e.g., letters of complaint, tweets, forum posts). It may also feature seemingly unobtrusive objects that can provide relevant insights into daily routines. All these materials constitute an extremely rich source of information for studying everyday life, and they are continuously growing with the burgeoning of new technologies for data recording, dissemination, and storage. Narratives are an excellent vehicle for studying everyday life, and quantitization is proposed as a means of integrating qualitative and quantitative elements. However, this analysis requires a structured system that enables researchers to analyze varying forms and sources of information objectively. In this paper, we present a methodological framework detailing the steps and decisions required to quantitatively analyze a set of data that was originally qualitative. We provide guidelines on study dimensions, text segmentation criteria, *ad hoc* observation instruments, data quality controls, and coding and preparation of text for quantitative analysis. The quality control stage is essential to ensure that the code matrices generated from the qualitative data are reliable. We provide examples of how an indirect observation study can produce data for quantitative analysis and also describe the different software tools available for the various stages of the process. The proposed method is framed within a specific mixed methods approach that involves collecting qualitative data and subsequently transforming these into matrices of codes (not frequencies) for quantitative analysis to detect underlying structures and behavioral patterns. The data collection and quality control procedures fully meet the requirement of flexibility and provide new perspectives on data integration in the study of biopsychosocial aspects in everyday contexts.

Keywords: indirect observation, mixed methods, textual materials, verbal behavior, systematic observation, quantitizing

INTRODUCTION

Psychological science has shown a growing interest in the study of everyday life. New methodologies have been proposed for the within-person study of real-time biopsychosocial aspects in their natural settings (Bolger et al., 2003; Conner and Lehman, 2013; Reis, 2013; Portell et al., 2015b,c). New technologies have made it possible to record spontaneous behavior—that is, behavior that is not elicited by a researcher but forms part of the subject's behavioral repertoire in his or her natural context (see e.g., Mehl et al., 2001). Compared with elicited behavior, spontaneous behavior has the advantage of occurring in a natural context and natural situation, so it is not influenced by extraneous variables such as a non-natural context or social desirability based on researchers' expectations. However, this area of study remains highly complex, particularly when it comes to obtaining quantitative indicators that make it possible to reconstruct the "who," "what," "how," and "when" of events of interest and position these events in the individuals' ecological niche. The difficult task of obtaining quantitative indicators of spontaneous behavior in everyday contexts has been further complicated by the long-standing clash between the qualitative and quantitative paradigms in psychology. Mixed methods research (Johnson et al., 2007) has provided valuable resources for combining qualitative data derived from non-spontaneous behavior (e.g., questionnaire responses) and narrative data derived from natural settings (e.g., a life story). Nevertheless, the merging of qualitative and quantitative perspectives in cases where only spontaneous behavior is of interest has been little explored. In this article, we describe a mixed methods approach grounded in observational methodology (Anguera, 2003) that we believe fills this gap. The proposed approach combines the strengths and offsets the weaknesses of the qualitative and quantitative perspectives.

We present a methodological framework for studying everyday behavior using a rigorous scientific approach based on indirect observation that involves "liquefying" transcribed verbal material or texts from original settings. The process involves the quantification of qualitative data using techniques that are based on order or sequence of events rather than on traditional frequency measures. The approach proposed is perfectly compatible with any guiding theoretical framework whatsoever; this method is not linked to any concrete theoretical model, it offers numerous methodological opportunities, and it has the potential to lead to significant developments in the field of studying everyday behaviors. This approach differs from previous work in this area (Sandelowski, 2001; Sandelowski et al., 2009; Seltzer-Kelly et al., 2012; Bell et al., 2016) in that it analyzes the order and sequence of events. The parameters of frequency (which only indicates the number of occurrences), order (which also provides information about sequence), and duration (which, in addition to the aforementioned information, also indicates the time in conventional units) provide a higher degree of data consistency (Bakeman, 1978). The use of the order parameter, with the introduction of sequentiality, entails added value of extraordinary importance (Sackett, 1980, 1987; Bakeman and Gottman, 1987; Magnusson, 1996, 2000, 2005, 2016; Sánchez-Algarra and Anguera, 2013; Portell et al., 2015a).

The presented liquefying method enables the systematic analysis of minor details that arise in a multitude of situations involving text (e.g., conversations, speeches, diary, or blog entries) with a level of granularity (Schegloff, 2000) that enables these "natural texts" to be analyzed in combination with other contextual data. The approach is applicable to both conventional and new forms of communication (e.g., WhatsApp messages), regardless of format or source. The source may be verbal behavior (informal conversations, focus group discussions, etc.) or documentary material (diaries, narratives, etc.), including in some cases graphic material, such as photographs and drawings.

Most of the solutions proposed to date for transforming text into quantitative data are either qualitative (e.g., ethnographic methods) or quantitative. Our proposal, however, takes a mixed methods approach in which spontaneously generated qualitative material is transformed into quantifiable code matrices.

In this article, we discuss key aspects of our proposed system. We analyze the concepts and meaning of systematic observation and one of its two branches, indirect observation, alongside key concepts of mixed methods research. We also look at types of qualitative data used in indirect observation and describe a methodological framework for building *ad hoc* observation instruments, creating matrices of codes for the data collected, and analyzing data and checking their reliability. Finally, we present a protocol specifically designed for indirect observation with examples from each of the stages in the process.

FROM SYSTEMATIC TO INDIRECT OBSERVATION

Psychologists work in a wide range of fields and subfields that correspond to everyday life situations. To name just a few examples, they are involved in health education programs in nurseries and nursing homes, prosocial programs in primary schools, exercise programs for the elderly, social support programs in neighborhoods, or communities with families of multiple nationalities, AIDS prevention programs for adolescents, support programs for families with a history of child abuse or negligence or families of young car crash victims, relaxation programs for athletes, and social programs in prisons or juvenile correctional institutions. Systematic observation can make important contributions to the study of spontaneous behavior in a vast range of everyday contexts.

Observation is a useful method for collecting, processing, and analyzing information that cannot be studied in the artificial setting of a laboratory. It enables a largely unbiased analysis of everyday behaviors and interactions that occur naturally (Anguera, 2010). Although systematic observation dates back to the 1970s, it has taken on an identity of its own in the last two decades (Anguera, 1979, 2003; Anguera and Izquierdo, 2006; Sánchez-Algarra and Anguera, 2013). It offers both flexibility and rigor as it is built on sound scientific principles, and this combination makes it ideal for use in many fields (Portell et al., 2015b).

Systematic observation differs from ethnography in that its purpose is not to obtain a narrative account of subjective

experiences in a process that requires the participation of the researcher or person being studied. Ethnographic studies require a qualitative approach, but unlike systematic observation, they do not require quantitative analysis and rigorous data quality control. Systematic observation, by contrast, is characterized by highly systematic data collection and analysis, stringent data quality controls, and the merging of qualitative and quantitative methods.

Systematic observation follows the four fundamental stages of scientific research: formulation of a research question, collection of data, analysis of data, and interpretation of results. The wealth of data collected in an observational study provides researchers with the opportunity to capture valuable chunks or snippets of everyday realities, without having to specifically ask for the information (there are no interviews, questionnaires, or psychological tests). In addition, it allows the researcher to study spontaneous behavior in a natural, uncontrolled environment.

Everyday activity in context is the cornerstone of observational studies. It is the source of a rich fabric of information that the psychologist/researcher needs to tap into in order to extract relevant information that is subsequently processed systematically to produce a set of "net" data that can be analyzed both qualitatively and quantitatively.

The study of everyday activity provides insights into the diverse behaviors and events that occur throughout a person's life. It provides thus a privileged vantage point from which to observe changes, but everyday life is a highly complex, dynamic process replete with information that is often not even known to exist (Anguera, 2001). Its study requires the examination of diverse phenomena at different levels of a pyramid-like structure. At the top of the pyramid, psychologists analyze how individuals go about their lives and gradually become familiar with what has shaped their life course. As they move down the pyramid, they discover everyday realities at different levels (family, career, social relationships, hobbies, etc.) and come to understand how these are influenced by interacting factors, such as health, satisfaction of needs, and conflicts.

According to Mucchielli (1974) observation equation, $O = P+I+Pk-B$, observation equals perception plus interpretation plus previous knowledge minus bias. Observation thus is not possible unless what is being observed is perceivable. Perceptibility is a key concept when it comes to differentiating between direct and indirect observation (Anguera, 1979, 2003). In indirect observation, it is always incomplete, and Mucchielli's equation is only partially fulfilled.

In direct observation, perceptibility is considered to be complete when what is being observed (whether *in situ* or through video or audio recordings) can be captured by visual or auditory senses. In anthropology, for example, the subfield concerned with the study of visual representations is known as *visual* rather than *observational* anthropology. Modern-day technology permits maximum levels of precision in visual and auditory perception (Escalera et al., 2009; Bautista et al., 2015) and minimizes the need for interpretation.

Although everyday contexts can take countless shapes and forms, the levels of response (or criteria or dimensions) that can be directly observed are similar. Facial expressions, for

example, can be analyzed by software such as Face Reader, which can distinguish between facial and emotional mimicry. Gestures, in turn, which also have an important role in human communication (Holle et al., 2012; Mashal et al., 2012), even in children (Lederer and Battaglia, 2015) can be effectively analyzed using programs such as NEUROGES+ELAN (Lausberg and Sloetjes, 2009, 2016). Finally, vocal behavior (Russ et al., 2008) can be analyzed using sound analysis software. Non-verbal manifestations, or "expressiveness," are interesting external indicators of a person's emotional state (Rodriguez et al., 2014), although adequate quality control is needed to reduce bias.

While aspects of human communication such as facial expressions, gestures, posture, and voice tone are fully perceivable through visual or auditory channels, they are frequently accompanied by verbal behavior, which has very different characteristics in terms of perceptibility. Indirect observation is an appropriate method for studying both verbal behavior and textual material, whether in the form of transcripts or original material produced by the participants in a study.

Verbal behavior transmits messages and both these and the channels through which they are transmitted can take many shapes and forms. Messages are analyzed differently depending on whether they are spoken or written. Written forms of expression (e.g., self-reports, diaries, biographies) are largely considered to be narratives. Narrative studies have been used in qualitative methodology for many years and have both strengths and shortcomings. One of their main strengths is their adaptability to very different situations and contexts. Narrative studies provide insights into a person's true nature and help to understand their experiences and needs (Riva et al., 2006). They have been used, for example, in a wide range of settings, such as secondary schools (García-Fariña, 2015; García-Fariña et al., 2016), high schools (Tronconi et al., 2018), family gatherings (Gimeno et al., 2006), support groups for patients (Roustan et al., 2013), therapeutic interaction (Blanchet et al., 2005), and group therapy for adolescents (Arias-Pujol and Anguera, 2017). One of their shortcomings is that perceptibility is limited by the documentary nature of the texts, and it is not uncommon for different researchers to draw different conclusions from the same text.

Human communication does not simply refer to the transmission of information. It involves numerous aspects that vary according to content, the people transmitting or receiving the message, their relationship (hierarchy, previous interactions, etc.), the flow of data or metadata, and the interpretative context. In addition, changing lifestyle habits and new technologies have led to new forms of human communication (Bavelas and Chovil, 2000), such as WhatsApp messages and blog posts, extending the traditional dichotomy between verbal and non-verbal behavior established by the classical sociologist Weick (1968, 1985). In a recent study, for example, Radzikowski et al. (2016) analyzed Twitter messages in a quantitative study on the rubella vaccine, and as stated by Hardley (2014, p. 34), "Over many decades, surveillance methods (often termed "indicator based" methods) have been developed and refined to provide disciplined, standardized approaches to acquiring and recording important information. More recently, ubiquitous and

unstandardized data collected from the Internet have been used to gain insight into emerging disease events."

Indirect observation can be considered a valid scientific method (Webb et al., 1966; Anguera, 1991, 2017, in press; Behar, 1993; Morales-Ortiz, 1999; Morales-Sánchez et al., 2014). It uses similar techniques to systematic observation, and as a procedure, it is structurally identical, although there are important differences dictated by the nature of the source data (verbal behavior and text).

Indirect observation involves the analysis of textual material generated either indirectly from transcriptions of audio recordings of verbal behavior in natural settings (e.g., conversation, group discussions) or directly from narratives (e.g., letters of complaint, tweets, forum posts). The addition of seemingly unobtrusive objects can also provide important insights into daily routines. All these materials constitute an extremely rich source of information for studying everyday life, and they are continuously growing with the burgeoning of new technologies for data recording, dissemination, and storage (Morales-Ortiz, 1999; Morales-Sánchez et al., 2014).

Narratives are an excellent vehicle for studying everyday life through indirect observation, and one option for studying them is to apply a procedure for systematizing and structuring the information through *quantitization*. This approach makes it possible to integrate qualitative and quantitative elements.

The data used in indirect observation invariably start out as qualitative and the source material varies according to the level of participation of the person being observed and the nature of the source (textual or non-textual).

Common sources of material used in indirect observation studies include:

- Recordings of verbal behavior as it occurs (normally in mp3 files). There may be single or multiple dialogues and it is essential to clearly distinguish between the different "voices" recorded.
- Transcripts of audio recordings of verbal behavior in a natural setting (Krueger and Casey, 2009). These may involve an individual (speaking, for example, in person or on the telephone), or a group (dyad, triad, focus group, etc.), in which each person can be clearly identified.
- Written texts produced by the participants in a research study. These include texts produced by the participants or those close to them (e.g., letters of complaint, letters to a newspaper, tweets, ads, messages on a mural, instant text messages). A variety of communication channels are possible (e.g., paper, e-mail, WhatsApp).
- Texts transmitted through the Internet, such as e-mails (Björk et al., 2014) and forum posts (Vaimberg, 2010). These constitute an extremely rich source of information and are particularly relevant to psychological interventions.
- Everyday objects related to the research question(s). While objects may appear to have a secondary role in communication, they can provide relevant insights into everyday life as they evoke or facilitate the expression of emotions through micro-valences (Lebrecht et al., 2012). Examples are graphs, paintings, models, and clay

figures. Technological advances have also opened up new opportunities in this area in recent years.

- Graphic material, particularly photographs. These can constitute an extremely rich source of information (Zaros, 2016). A single photograph captures a moment, something static, but a gallery of photographs separated in time can capture the dynamics of an episode or successive episodes in the life of a person, or even a group or institution. This material can be primary (the only source available) or secondary (complementing other sources).
- Unobtrusive objects, also referred to as aggregates (Webb et al., 1966). These may simply be anecdotal, but in some cases they can reveal the existence of certain behaviors, but only after a process of inference involving variable risk. Examples are fingerprints and objects such as cigarette butts or a napkin with notes or drawings left behind in a café.

The above sources of information give rise to a varied set of data that provides empirical evidence and can position specific events and everyday behaviors along a continuum of time. Finally, the information available becomes progressively richer as one gains access to several sources of documentary material.

As mentioned, the material used to collect data in indirect observation is only partly perceivable (Anguera, 1991) and any conclusions made need to be inferred by a researcher drawing from a theoretical framework or taking a position. This is the main challenge in indirect observation. In the system we propose, rigorous application of a carefully designed observation instrument by duly trained observers offers the necessary guarantees of data reliability. Although direct and indirect observation may vary in terms of source material, level of interpretation, and level of participation, the two methods share a scientific procedure that when properly applied can provide quantitative indicators of the processes underlying everyday behavior.

THE CHALLENGES OF MIXED METHODS RESEARCH

Mixed methods research has been increasingly embraced by the scientific community over the past 15 years (Creswell et al., 2003; Johnson et al., 2007; Tashakkori and Teddlie, 2010; Onwuegbuzie and Hitchcock, 2015). The mixed methods approach involves the collection, analysis, and interpretation of qualitative and quantitative data for the same purpose and within the framework of the same study; some authors have even raised the approach to the rank of paradigm. Molina-Azorín and Cameron (2015) acknowledge that mixed methods research is not easy to conduct and requires considerable time and resources. Nonetheless, it is a movement that is gradually gaining supporters. As stated by Leech and Onwuegbuzie (2009) and Onwuegbuzie (2003), mixed methods research lies on a continuum between single-method and fully mixed studies, although the scientific community has yet to agree on which position it holds along this continuum. That said, it is generally agreed that the position will depend on the research objective and the nature of the data, analyses, and level of inference.

Overall, mixed research is largely understood as "a synthesis that includes ideas from qualitative and quantitative research" (Johnson et al., 2007, p. 113). However, this is a very broad framework in which many gaps need to be filled. In the case of indirect observation, the methodological approach must be extremely rigorous as we are dealing with situations in which substantive areas merge with the multiple realities of everyday life.

The exponential growth of mixed methods research in recent decades has generated certain inconsistencies in terms of terminology and definitions. We therefore believe that it is first necessary to clarify the meaning of method/methodology and to discuss the multiple meanings attached to the term "mixed method" before we present our methodological framework for indirect observation.

Greene (2006, p. 93) proposed a broad description of the term "methodology," understood as an inquiry logic that admits different forms of data collection (questionnaires, interviews, observational datasets, etc.), methods of research (experimental, ethnographic, etc.), and related philosophical issues (ontology, epistemology, axiology, etc.). Greene also refers to specific guidelines for practice, which distinguish between methods that obviously vary in terms of design, sampling, data gathering, analysis, etc. We consider that systematic observation fits with Greene's definition of methodology (Anguera, 2003), although we have not always used the term. We also agree with the following statement by Johnson et al. (2007, p. 118): "It is important to keep in one's mind, however, that the word *methods* should be viewed broadly." Accordingly, in the approach we describe in this article, we also consider indirect observation to be a method in the broad sense of the word.

Johnson et al. (2007, p. 123) defined mixed methods research as "the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration" (Johnson et al., 2007, p. 123). They formulated this definition after asking 19 renowned researchers in the field (Pat Bazeley, Valerie Caracelli, Huey Chen, John Creswell, Steve Currall, Marvin Formosa, Jennifer Greene, Al Hunter, Burke Johnsson and Anthony Onwuegbuzie, Udo Kelle, Donna Mertens, Steven Miller, Janice Morse, Isadore Newman, Michael Q. Patton, Hallie Preskill, Margarete Sandelowski, Lyn Shulha, Abbas Tashakkori, and Charles Teddlie) to send in their definition of the term "mixed methods" by e-mail.

We fully agree with the definition proposed by Johnson et al. (2007) and it provided us with the necessary elements to draw up our methodological framework for indirect observation. The success of any mixed methods approach depends on the adequate mixing or integration of qualitative and quantitative elements. Numerous authors have analyzed the term "mixing" in an attempt to provide guidance on the processes required to achieve a seamless result (Bazeley, 2009; O'Cathain et al., 2010; Fetters and Freshwater, 2015). Qualitative and quantitative data can be mixed in three different ways, aptly summed up by Creswell and Plano Clark (2007, p. 7): "There are three ways in which mixing occurs: merging or converging the two datasets by actually bringing them together, connecting the two datasets by having one build on the other, or embedding one data set within the other so that one type of data provides a supportive role for the other data set." For our proposal, we chose the second form: connecting two databases by having one build on the other. According to Sandelowski et al. (2009), this connection can be achieved through transformation, i.e., by quantitizing qualitative data or qualitizing quantitative data. In our indirect observation framework, we transform non-systematic qualitative data into a format suitable for quantitative analysis.

Mixed methods research is marked by a persistent scientific gap that requires powerful solutions rooted in two key challenges in the field of indirect observation. These two challenges, discussed in this article, are (a) how to rigorously transform qualitative textual material derived largely from everyday human communication into matrices of codes, and (b) how to subsequently analyze these codes using quantitative methods suited to the categorical nature of the data in order to uncover the underlying structure. The proposed transformation system breaks away from the classical theoretical framework of mixed methods, which simply involves integrating qualitative and quantitative elements. The key difference is that it contemplates systematic observation, and hence indirect observation, to be a mixed method in itself (Anguera and Hernández-Mendo, 2016; Anguera et al., 2017a).

Integration of qualitative and quantitative elements is the key to any mixed methods approach (Creswell and Plano Clark, 2007; Bazeley, 2009; O'Cathain et al., 2010; Maxwell et al., 2015). Our approach adds another element: the liquefaction of verbal behavior and texts. This process consists of schematically transforming "solid" textual material into "liquid" matrices of codes apt for quantitative analysis (Anguera et al., 2017b; Anguera, in press). The quantitative processing of originally qualitative data with the aim of detecting hidden behavioral patterns or underlying structures, for example, adds an element of robustness to the integration of qualitative and quantitative data, particularly in the case of everyday life events and behaviors.

Talkativeness and text, for example, can now be analyzed within the framework of mixed methods research using frequency counts (Poitras et al., 2015) thanks to the development of reliable—and extremely useful—measures of verbal productivity and the multiple opportunities offered by modern-day technology (Bazeley, 2003, 2006, 2009). Frequency counts, however, are weak and insufficient measures. Considering that "methodological plenitude" (Love, 2006, p. 455) is not always attainable in applied research, the mixed method framework offers new and interesting possibilities for indirect observation.

The combined use of qualitative and quantitative approaches has been tried and tested in multiple studies and has also been analyzed in several systematic reviews (Elvish et al., 2013). In the following sections, we show that it is necessary to start with qualitative inputs and to then quantify these in a process that ensures reliability throughout the various stages.

QUALITATIVE DATASETS IN INDIRECT OBSERVATION

The empirical process in indirect observation starts with the collection of qualitative data. While the characteristics and standards that guarantee quality are perfectly outlined in the literature on quantitative methodology, the same cannot be said of qualitative methodology. Qualitative methodology offers enormous flexibility, but interpretations on content and form vary and are not free of controversy. Content provides personal and interpersonal information, which stems from experiences that are temporally unstable and highly influenced by the context and versatility of the moment. As for form, the tools used to support indirect observation (narratives, biographies, self-reports, life stories, in-depth interviews, etc.) cause doubt and distrust in many researchers, who, in the absence of standardized tools, question their stability and consistency.

Much has been written about the forms used to structure narratives (e.g., Hurwitz et al., 2004; De Fina and Georgakopoulo, 2015; Riessman, 2015), and qualitative data can be gathered using many tools, including interviews (e.g., Riera et al., 2015), biographies (e.g., Lindqvist et al., 2014), children's vignettes (e.g., Jackson et al., 2015), focus group vignettes (e.g., Brondani et al., 2008), telephone interviews (e.g., Björk et al., 2014), self-reports (e.g., Coutinho et al., 2014), focus group recordings (e.g., McLean et al., 2011), and participant observation (e.g., Caddick et al., 2015). In our case we are specifically interested in qualitative datasets within the framework of indirect observation. Although systematic observation dates back to the 1970s, it has taken on an identity of its own in the last two decades (Anguera, 2003; Anguera and Izquierdo, 2006; Sánchez-Algarra and Anguera, 2013; Anguera et al., 2017a). Indirect observation shares many of the characteristics previously described for systematic observation, namely, highly systematic data collection and analysis, strict data quality controls, and an approach that requires the merging of qualitative and quantitative techniques.

A METHODOLOGICAL FRAMEWORK FOR LIQUEFYING TEXT

In these next sections, we are going to describe, and illustrate with examples, the stages and sub-stages involved in an indirect observation study. We will focus largely on the extraction and transformation of information from textual material produced using conventional or newer channels of communication in a variety of formats (handwritten letters, reports, transcriptions of group meetings, and interviews, etc.), irrespective of origin (e.g., informal conversations or focus group discussions or documentary material).

Extracting information on human behavior from text and transforming it into suitably systematized and organized categorical data, without loss of key information, is a major challenge in the Behavioral Sciences. In addition, the process must offer sufficient scientific and ethical guarantees and produce results in a format that can be rigorously processed using any of a range of quantitative techniques available for analyzing categorical data.

Our text-liquefying process consists of six stages: (1) specification of study dimensions, (2) establishment of segmentation criteria to divide the text into meaningful units, (3) building of a purpose-designed observation instrument, (4) coding of information, (5) data quality control, and (6) quantitative analysis of data. **Table 9** presents detailed steps and guidelines for the "liquefication" of indirect observations. Each of the steps will be explicated within the following sections.

Specification of Study Dimensions

In systematic observation, and by extension, indirect observation, the term "dimension," also known as level of response (Weick, 1968) or criterion, refers to a distinguishable facet related to the research objective. Dimensions are generally derived from a theoretical framework (e.g., the seminal work of (Weick, 1985) in the field of social interaction), but they can also be created *ex novo* based on experience or expertise. In the latter case, they must always be justified.

Studies can be one-dimensional or multidimensional. It is not uncommon for researchers to start off with a single dimension and then gradually add others as they delve deeper into the theoretical framework. Below are examples of dimensions and theoretical frameworks used in three indirect observation studies. In the first case, a study of disruptive behavior and communication difficulties in adolescents participating in group communication therapy, Arias-Pujol and Anguera (2004) proposed the dimensions verbal and non-verbal behavior, derived from the corresponding interpersonal theoretical framework (Danzinger, 1982; Gale, 1991; Poyatos, 1993). In the second case, Vaimberg (2010), on studying a psychotherapy group in which participants were able to write what they wanted on an online forum at any time over 3 years, chose the following dimensions: in-person, otherness, emotionality, thoughtfulness, positivity, and realism. The theoretical framework was built from work by various authors (e.g., Winniccott, 1979; Bion, 1985; McDougall, 1991; Lévy, 1995). In the third case, which was a recent study of teacher-led discourse in physical education built on the theoretical framework of the Teaching Games for Understanding model (originally proposed by Bunker and Thorpe, 1982) and work on discourse strategies by Coll and Onrubia (2001), García-Fariña et al. (2016) proposed nine dimensions: exploration and activation of previous knowledge, attribution of positive meaning by students, progressive establishment of increasingly expert and complex representations of subject matter, interactivity segment, message structure, extralinguistic resources, task type, destination of message, and location of session.

Specification of Segmentation Criteria to Create Textual Units

The second step toward liquefying a text is to define the segmentation criteria to divide the text into meaningful units. This process is known as "unitizing." Although initially proposed by Dickman (1963) and Birdwhistell (1970), Krippendorff (2013, p. 84) defined unitizing as "the systematic distinctions with

TABLE 1 | Vignette showing the segmentation of a text (transcribed from a conversation) into units.

S1. The truth is that I sometimes doubt whether I like basketball that much [U1], even though I have already devoted 15 years of my life to the game [U2].

S2. But you started as a young boy [U3], when you were given the possibility of playing as a junior at school [U4].

S1. That moment was very important for me [U5], as I got carried away with the enthusiasm [U6] and I couldn't go for a day without playing [U7]. Then, when I finished secondary school, I got the opportunity to join the club where I am now [U8] and to dedicate myself in body and soul to basketball [U9].

S2. Did you think back then about what this decision would entail? [U10].

S1. I couldn't tell you exactly…[U11] I think I was somewhat confused [U12], as on the one hand I wanted to study industrial engineering, probably influenced by my father and my uncle [U13], but on the other, the fact that I was valued, without being particularly tall [11], was a golden dream [U15]. I think that I was living between real life and the dream…[U16] And I accepted straight away [U17], although after talking it through with my parents, uncle and brother [U18]. They gave me some opinions and advice [U19], but left the final decision up to me [U20].

a continuum of otherwise undifferentiated text—documents, images, voices, websites, and other observables—that are of interest to an analysis, omitting irrelevant matter but keeping together what cannot be divided without loss of meaning." This definition suggests that it would be logical to first segment the text into primary criteria within the main study dimension and then establish secondary criteria for the other dimensions (e.g., voices, gestural behavior, etc.).

Krippendorff (2013) suggested segmenting text using orthographic, syntactic, contextual, and inter-speaker criteria. In this last case, each intervention by an individual is considered a unit. This is a very useful approach for analyzing interactions between various people. We propose using the inter-speaker criterion as the primary criterion and subsequently establishing secondary criteria (subunits) for verbal or written interventions containing various syntactic elements (phrases).

In cases with several dimensions, such as verbal behavior accompanied by gestures, postures, or exchange of looks, verbal behavior, as the most perceivable behavior, could be established as the primary criterion. The other behaviors could then be segmented into subunits as appropriate. In very specialized cases, however, we consider that the above level of segmentation is insufficient. The initial segmentation stage is crucial as the categories that will be created in the next stage will directly determine the content of the dataset for analysis. Where possible, test runs or pilot studies should be performed first. **Table 1** shows how a conversation between two anonymous speakers is segmented into units.

Building an Indirect Observation Instrument

Indirect observation studies, like systematic observation studies (Anguera, 2003; Anguera and Izquierdo, 2006; Sánchez-Algarra and Anguera, 2013; Portell et al., 2015a) require a purpose-built observation instrument to systematically code the information that will form the subsequent datasets.

Observation instruments can be built using category systems, a field format system, a combination of these systems, or rating scales (Anguera et al., 2007). One-dimensional studies use category systems and rating scales, while multidimensional studies use field formats or field formats combined with category systems. To build a category system, there must be a theoretical framework, and to build a rating scale, it must be possible to grade the corresponding dimensions ordinally. In addition, the

category system must fulfill the requirements of exhaustivity and mutual exclusion, and each category must be accurately defined.

The field format is built by creating a catalog of mutually exclusive behaviors for each dimension. As it is not exhaustive, the catalog is left open and is therefore considered to be in a permanent state of construction. While not required, a theoretical framework is recommendable for field format systems.

Observation instruments combining a field format system with category systems are becoming increasingly common. This combination is possible when some or all of the dimensions in the field format have a theoretical framework and the object of research is atemporal (i.e., it is not a process).

To simplify matters, it is highly recommendable to code both categories and dimensions using letters, numbers, or symbols. If A, B, C, and D are categories in a category system, i.e., fulfilling the requirements of exhaustivity and mutual exclusion (e.g., A = XX, B = XX, C = XX, and D = XX), then the notation would be CS (category system) = {A B C D}. If A, B, C, and D are behaviors in an open catalog, i.e., they are mutually exclusive but not exhaustive (e.g., A = XX, B = XX, C = XX, and D = XX), the notation would be Catalogue = A B C D…

Guidelines for Coding Information

Observational datasets created from narratives (Crawford, 1992; Gabriel, 2004; Tuttas, 2015) have wide applications in many everyday life situations. However, before qualitative inputs from human communication can be transformed into quantitative data, it is first necessary to decide how to organize the heterogeneous information available. This process can be extremely complex as it is necessary to bring together data from very different sources, and very possibly, different points in time (Duran et al., 2007). The first step is to correctly record and code the data, and this is where the *ad hoc* observation instrument becomes invaluable. As started by Bradley et al. (2007, p. 1,761), "coding provides the analyst with a formal system to organize the data, uncovering and documenting additional links within and between concepts and experiences described in the data."

If the sources have been carefully selected, they will all contribute to creating a stockpile of information on the behaviors or actions of all those involved in the communication process being analyzed (e.g., therapists, participants, supervisors…).

The system for processing narratives or bodies of texts is quite similar to that used in discourse analysis (Calsamiglia and

TABLE 2 | Tabular structure for creating a code matrix.

		Dimensions (i)			
		D1 [Verbal behavior]	D2 [Vocal behavior]	...	Di
Textual units (n)	U1 [Initial greeting]	Code … [Informal expressions]	Code … [Raised voice with laughter]		Code …
	U2 [Activity to carry out]	Code … [We have to solve the math problems]	Code … [Normal voice]		
	...				
	Un	Code …	Code …	...	Code …

Each row of the matrix contains a series of boxes that are completed with codes corresponding to each textual unit (fragments of text from indirect observation). The columns, in turn, contain the different dimensions, or criteria, of the observation instrument. The codes come from the ad hoc observation instrument and may correspond to behaviors from a field format catalog or to categories in an observation instrument based on category systems only or on category systems combined with a field format. By way of illustration, we have added in brackets the first two dimensions (verbal behavior and vocal behavior), a simulation of the first units and an indication of the behaviors produced (which will be coded).

Tusón, 1999), although the information retrieved is richer and more diverse. Once the necessary quality controls are in place, the information can be managed and processed systematically within an empirical research setting that ensures replicability. Examples of texts used for this purpose are interviews, speeches, and conversations (Sidnell and Stivers, 2013). These may be a specific audience, a single speaker or several (with turn-taking), words in isolation, or, when direct and indirect observation are combined, words accompanied by tone/pitch, gestures, facial expressions, posture, objects, etc (Fischer et al., 2012).

Once the study dimensions have been selected (section Specification of Study Dimensions) and the text has been segmented into units (section Specification of Segmentation Criteria to Create Textual Units) and the behaviors coded using the *ad hoc* observation instrument (section Building an Indirect Observation Instrument), the data can be transformed into a series of complete or incomplete code matrices containing purely qualitative information (Anguera, 2017, in press; Anguera et al., 2017b). This transformation is achieved by organizing the dimensions into columns and adding the behavioral units to the corresponding rows, achieving thus a "liquid" text, ready for quantitative analysis (**Table 2** contains an example).

Table 3a shows a hypothetical example of data extracted from a text in a one-dimensional study using an observation instrument with category systems, using a simulated example of the diary of a patient with endogenous depression. **Table 3b**, in turn, shows the results for a combined field format-category system instrument from a multidimensional study, using a simulated example of an oral mediation situation involving a conflict between the parties A and B, with the assistance of the mediator C. These matrices of codes (**Table 3a** is atypical as it has just one column due to the single dimension analyzed) show how the qualitative data have been structured.

Additional sources of information, such as drawings, sounds, or photographs can be incorporated simply by adding new dimensions. Although this is still a relatively new concept, it is perfectly feasible with today's advanced coding systems (Saldaña, 2013) and technological possibilities (e.g., Bazeley, 2003, 2006, 2009; Crutcher, 2003, 2007; Holtgraves and Han, 2007; Romero et al., 2007; Dam and Kaufmann, 2008; Taylor et al., 2015). In the ATLAS.ti (v.7) qualitative data analysis program, for example, the text coding feature can be used to supplement the information entered with an object or an audio or video recording.

Researchers now have access to a multitude of software programs that facilitate their work. For those working with indirect observation, the CAQDAS platform (AQUAD6, ATLAS.ti, MAXqda2, NUDIST, NVivo, etc.) offers numerous programs for segmenting and coding text, and there are also open-access programs, such as T-LAB (http://tlab.it/en/presentation.php), IRAMUTEQ (www.iramuteq.org), and those created by the Italian group GIAT (www.giat.org). Numerous considerations are necessary when extracting information from text using content analysis techniques. Content analysis programs have traditionally favored the processing of large, mostly qualitative, bodies of texts, graphs, and audio and video material. The analysis uncovers relational structures (families, networks, etc.) that are relatively stable, or at least appear to be, and are always determined by the choices of the researcher. Nowadays, however, powerful software programs can analyze multiple sources of information to produce code matrices (Vaimberg, 2010) that are of enormous value for analyzing human communication in many fields.

Two programs can be used for both direct and indirect observation. These are HOISAN (Hernández-Mendo et al., 2012) (http://www.menpas.com), which is open-access and is available in several languages (English, Spanish, Portuguese, French) that can be selected from the tab *Archivos (Files)*, and TRANSANA (http://www.transana.com).

QUANTITATIVE PROCESSING OF CODE MATRICES

Rigorous Data Quality Control

The issue of data quality in indirect observation has been widely debated in the literature, with a particular focus on reliability and validity, and concerns have led many psychologists and researchers working in this area to modify their approaches. Both intraobserver and interobserver agreement are important measures of reliability, but they are not the only ones. While reliability is necessary, it alone does not guarantee the validity of a dataset (Krippendorff, 2013).

Krippendorff (2013) was the first author to insist on rigorous data quality control as a requirement for the quantification of data resulting from indirect observation. Thanks to his

TABLE 3 | (a,b) Hypothetical examples of a code matrix derived from a text.

a	b

a

Diary of a patient diagnosed with endogenous depression:

SC = {A B C D}

A: Expressions of sorrow or sadness
B: Expressions of self-perceived improvement
C: Expressions of self-perceived worsening (situation of hopelessness)
D: Expressions of joy at having overcome the problem

[This is an exhaustive and mutually exclusive system of categories, constructed from a theoretical framework (Altimir et al., 2010; Dagnino et al., 2012; Krause et al., 2016)]

Units	Categ.
U1	A
U2	B
U3	D
U4	A
U5	C
U6	A
U7	B
U8	D
U9	A
U10	D
U11	A
U12	B
U13	C
U14	A
U15	B
U16	D
U17	A
U18	B
U19	D
U20	C

b

Oral mediation situation involving a conflict between the parties A and B, with the assistance of the mediator C:
E = {E1 E2 E3 E4 E5}

F = F1 F2 F3 F4 F5 F6 F7 F8 …

G = {G1 G2 G3}

H = H1 H2 H3 H4 …

Dimension E: Verbal behavior
E1: Facilitating elements (greeting, courtesy routines, etc.)
E2: Focused on the crux of the issue
E3: Related to secondary aspects
E4: Neutral sentences not related to the conflict
E5: Conflictive elements (insults, mockery, etc.)
[This is an exhaustive and mutually exclusive system of categories]

Dimension F: Vocal conduct
F1: Shouting
F2: Speaking in an annoyed tone
F3: Speaking loudly
F4: Speaking while crying
F5: Speaking normally
F6: Speaking softly
F7: Whispering
F8: Silence
[This is a catalog of behaviors; as such, it is an open list and additional codes can be added]

Dimension G: Interacting parties
G1: Party A
G2: Party B
G3: Mediator
[This is an exhaustive and mutually exclusive system of categories]

Dimension H: Expression of displeasure/disagreement
H1: Shaking head to indicate "no"
H2: H1 plus hands clasped
H3: H2 plus bulging eyes
H4: H3 plus clenched jaw
[This is a catalog of behaviors; as such, it is an open list and additional codes can be added]

Units	Dim. E	Dim. F	Dim. G	Dim. H
U1	E1		G1	H1
U2	E2	F5	G2	H3
U3	E2	F5	G2	H2
U4	E3	F5	G1	H2
U5	E3	F1	G2	
U6	E3	F1	G2	
U7	E5	F1	G3	
U8	E5	F1	G1	
U9		F5	G1	
U10		F5	G1	H1
U11	E1	F5	G2	H2
U12	E1		G1	
U13	E1	F4	G3	H1
U14	E2	F2	G3	
U15	E2	F2	G3	H1

The columns correspond to the dimensions and the rows to the units into which they were segmented. Codes on the same row reflect concurrent behaviors. The codes are defined in the ad hoc instrument designed for the study.

TABLE 4 | Example of datasets used to calculate intraobserver canonical agreement.

Occasion 1	Occasion 2	Occasion 3	Canonical agreement
E1 G1 H3	E1 G1 H3	E1 G1 H3	0.84
E2 F5 G2 H3	E2 F5 G2 H3	E2 F5 G2 H3	Result: Satisfactory agreement
E2 F5 G2 H2	E2 F5 G2 H2	E3 F5 G2 H2	
E3 F5 G1 H2	E3 F5 G1 H2	E3 F5 G1 H2	
E3 F1 G2 H4	E3 G2 H4	E3 F1 G2 H4	
E3 F1 G2	E3 F1 G2	E3 F1	
E5 F1 G3	E5 F1 G3	E5 F1 G3	
E5 F1 G1	E1 F1 G1	E1 F1 G1	
F5 G1	F5 G1	F5 G1	
F5 G1 H1	F5 G1 H1	F5 G1 H1	
E1 F5 G2 H1	E1 F5 G2 H1	E1 F5 G2 H1	
E1 G1	E1 G1	E1 G1	

In such cases, the same verbal behavior or textual material must be coded by the same observer, using the same indirect observation instrument, on three separate occasions, separated by at least a week. The data in the first column are from **Table 3b**.

TABLE 5 | (a) The first row shows the simple frequency counts for the data from **Table 3a**. The matrix below shows the transition frequencies for the given behavior A with the conditional behaviors shown at the head of each column. The different lags are shown by rows. **(b)** The first row shows the unconditional probabilities while the rows below show the conditional probabilities.

	a					b			
Lag	A	B	C	D	Total	A	B	C	D
	7	5	3	5	20	0.35	0.25	0.15	0.25
1	0	5	1	1	7	0	**0.71**	0.14	0.14
2	2	0	1	4	7	0.28	0	0.14	**0.57**
3	4	2	1	0	7	**0.57**	**0.28**	0.14	0
4	0	2	2	2	6	0	**0.33**	**0.33**	**0.33**
5	4	0	0	1	5	**0.8**	0	0	0.2

Bold values are significative (upper that respective unconditional probabilities).

contributions in this area, there are now methodological tools in place to demonstrate the quality of such data. The two main quantitative measures for testing the reliability of data from direct observation (behaviors) and indirect observation (texts) are (a) coefficients of agreement between two observers who separately code behaviors using the same dataset and observation instrument and (b) coefficients of agreement based on correlation. Numerous coefficients exist for quantitatively verifying the quality of data in a wide range of situations. One widely used measure in indirect observation is Krippendorff's canonical agreement coefficient, which is an adaptation of Cohen's kappa coefficient for analyzing three or more datasets. It can be calculated in HOISAN. Another option for use in situations with different sources of variation is generalizability theory (Blanco-Villaseñor, 2001; Escolano-Pérez et al., 2017).

A more qualitative method, the consensus agreement method (Anguera, 1990), is gaining increasing recognition in indirect observation and other studies. In this method, at least three observers work together to discuss and agree on the most suitable code for each unit from the observation unit. This method has obvious advantages, as it produces a single dataset and frequently results in a better observation instrument thanks to the detection of possible gaps and shortcomings. While it offers significant guarantees of quality, however, it also carries risks. An observer may defer to the decisions of a more senior or "expert" colleague, for example, and the need to agree can also give rise to frictions or conflicts. The results of the consensus agreement method can be complemented by quantitative measures of agreement (Arana et al., 2016).

There has been much debate in the field of psychology about the extent to which adherence to a particular theoretical framework may influence agreement between observers. To overcome this potential problem, Pope et al. (2000) proposed using observers from different backgrounds to analyze the data. Such an approach, however, would require even more rigorous quality control measures given the greater difficulty of reaching agreement.

Table 4 shows the canonical agreement coefficient calculated in HOISAN for the data in **Table 3b**, combined with two other sets of data recorded for the same section of text by the same observer and with the same instrument, but at different moments.

Quantitative Processing of Code Matrices

Once the text has been liquefied and the necessary data controls performed, the researcher now has access to a series of code matrices perfectly suited for analysis using different techniques.

The novel nature of our proposal is that we do not study frequency counts, which, despite their serious limitations, were the only measure of quantification used in observation studies for decades.

Over the last 15 years, our group has prioritized three analytical techniques that are particularly well-suited to processing qualitative data in both systematic observation (Blanco-Villaseñor et al., 2003) and indirect observation studies. These are lag sequential analysis, polar coordinate analysis, and T-pattern detection. All three techniques are based on statistical calculations and therefore provide the necessary guarantees of replicability and robustness.

Lag Sequential Analysis

Lag sequential analysis, which works with code matrices (see example in **Table 5**), is used to detect behavioral patterns that show the structure of interactive episodes (Bakeman, 1978, 1991; Bakeman and Gottman, 1987; Bakeman and Quera, 1996, 2011). The analysis can be performed prospectively (looking forward in time from a given moment) or retrospectively (looking backwards) using positive or negative lag counts. A behavior, for example, with a lag count of +2 would correspond to a behavior that occurs 2 positions after the behavior(s) of interest, while one with a lag count of −2 would correspond to a behavior that occurs 2 positions before the behavior(s) of interest.

The analysis can be applied to part of a session, to a complete session, to parts of different sessions (e.g., the first few minutes of a series of sessions), or to series of complete sessions. The technique thus offers enormous flexibility in terms of addressing different research questions. Two types of data can be used:
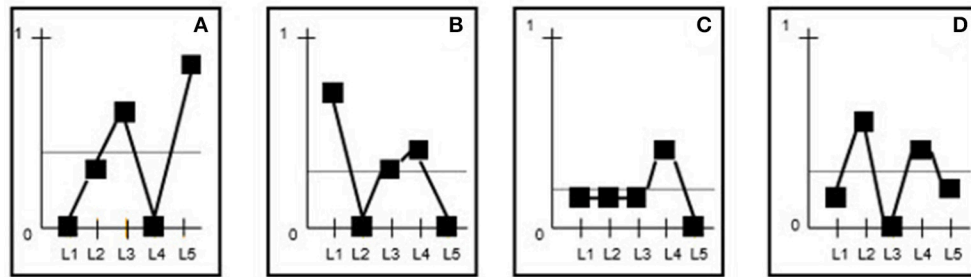
FIGURE 1 | (A–D) The lags are shown on the X-axis and the probabilities on the Y-axis. Based on the results from **Table 5b**, the values corresponding to the unconditional probabilities (first row) are indicated by the horizontal line parallel to the X-axis (e.g., 0.35 for category A). Also shown are the values for each of the conditional probabilities for each category and lag. These values are linked by a (generally uneven) line for each category. The horizontal line parallel to the X-axis represents the upper limit for the effect of chance. Accordingly, any conditional probabilities in the subsequent lags that are higher than the unconditional probability for the corresponding category are significant and hence form part of the behavioral pattern.
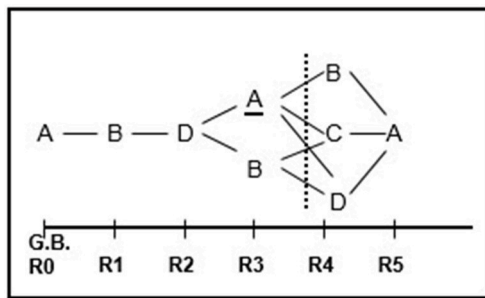
FIGURE 2 | Behavioral pattern extracted after assigning significant conditional behaviors (behaviors with a conditional probability greater than the unconditional probability) to each lag. The behavior pattern extracted from the presented illustration exhibits a regularity consisting of expressions of sorrow or sadness being followed by expressions of self-perceived improvement and these expressions, in turn, being followed by joy at having overcome the problem. From there, the pattern bifurcates, leading either to the initial situation of sorrow and sadness or to expressions of self-perceived worsening.

TABLE 6 | (a) Formula for calculating the corrected unconditional (expected) probability. **(b)** Table showing the probabilities from **Table 5b** with the addition of the corrected conditional probabilities in the second row (bold values).

a	b			
	A	**B**	**C**	**D**
$p_{esp\ corr} = p_{esp} + Z_\alpha \sigma$	0.35	0.25	0.15	0.25
$\sigma = \sqrt{\dfrac{p_{esp}(1 - p_{esp})}{N}}$	**0.56**	**0.43**	**0.31**	**0.43**
	0	**0.71**	0.14	0.14
	0.28	0	0.14	**0.57**
	0.57	**0.28**	0.14	0
	0	**0.33**	**0.33**	**0.33**
	0.8	0	0	0.2

These correspond to the upper limit of the confidence interval built around the unconditional probability values, with p < 0.05.

data for which only the order of occurrence of concurrent behaviors has been recorded, using any of the free software programs available SDIS-GSEQ v. 4.1.2 (Bakeman and Quera, 2011), GSEQ5 (Bakeman and Quera, 2011), or HOISAN v. 1.6.3.3 (Hernández-Mendo et al., 2012), and data for which both order and duration have been recorded (SDIS-QSEQ and GSEQ5). Lag sequential analysis has been successfully applied in many indirect observation studies conducted over the past 25 years (e.g., Martínez del Pozo, 1993; Arias-Pujol and Anguera, 2004; Cuervo, 2014).

Using the data from **Table 3a** again, we illustrate how to manually calculate the results for the first, and simple, part of the lag sequential analysis process. The first step is to create tables for the matching frequencies and probabilities (**Tables 5a,b**) for category A (in our example, expressions of sorrow or sadness), which, according to the hypothesis applied, is the given behavior (the behavior of interest). In row 1, for example, A has a frequency count of 0 because this code does not occur again; B (expressions of self-perceived improvement) has a count of 5 because it occurs after A on five occasions (units 2, 7, 12, 15, and

18); C (expressions of self-perceived worsening) has a count of 1 because it only occurs after A on one occasion (unit 5), similarly to D (expressions of joy at having overcome the problem) (unit 10). In row 2, in turn, A has a count of 2 because it occurs on two occasions (units 6 and 11) in the second position after the given behaviors (units 4 and 9, respectively); B has a count of 0 because it does not occur in the second position after the given behavior; and C has a count of 1 because it occurs just once (unit 13) in the second position after the given behavior (unit 11), and so on.

The data are analyzed to search for behavioral patterns, with consideration of some or all of the other behaviors, known as target behaviors, to see if they form part of the pattern(s) detected.

The information for each of the categories is shown on a graph with the lags on the X-axis and the probability values (ranging between 0 and 1) on the Y-axis. Each of the four **Figures 1A–D**, shows the value of the unconditional probability (the line parallel to the Y-axis) and the points corresponding to the conditional probability of each lag.

Based on this simple visual output and considering all the statistically significant categories at each lag (i.e., the categories with a conditional probability value greater than that of the
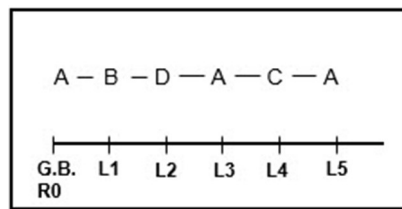
FIGURE 3 | Optimized corrected behavioral pattern following construction of a confidence interval around the unconditional probabilities. The corrected pattern reveals the typical alternation seen in patients with endogenous depression.
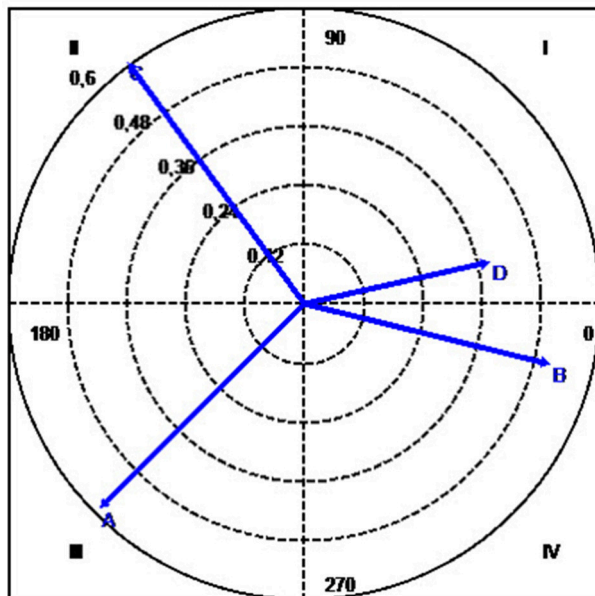
FIGURE 4 | Polar coordinate map showing the vectors for the categories A (focal category), B, C, and D. As indicated in the legend of **Table 8**, A is the focal behavior and expressions of sorrow or sadness activate expressions of self-perceived improvement (Quadrant IV) and joy at having overcome the problem (Quadrant I). The focal behavior is not self-generating (Quadrant III). Additionally, expressions of sorrow or sadness do not generate self-perceived worsening (Quadrant II), although self-perceived worsening does generate the focal behavior.

TABLE 7 | Adjusted residuals and corresponding Z-values from the polar coordinate analysis with A as the focal behavior or category and B, C, and D as the conditional behaviors.

Lag	Categories			
	A	**B**	**C**	**D**
ADJUSTED RESIDUALS				
−5	1.94	−0.187	−0.996	−1.236
−4	−1.936	0.826	2.148	−0.413
−3	1.253	−1.557	−1.019	0.934
−2	−0.125	2.49	−1.112	−1.689
−1	−2.121	−1.861	2.121	2.605
0	4.359	−1.77	−1.282	−1.77
+1	−2.121	3.207	0	−0.744
+2	−0.125	−1.689	−0.078	1.9
+3	1.253	0.596	−0.165	−1.789
+4	−1.936	0.826	0.537	0.826
+5	1.94	−1.764	−1.139	0.706
Z-VALUES				
−5	1.94	−0.187	−0.966	−1.236
−4	−1.936	0.826	2.148	−0.413
−3	1.253	−1.557	−1.019	0.934
−2	−0.125	−2.49	−1.112	−1.689
−1	−2.121	−1.861	2.121	2.605
+1	−2.12	3.207	0	−0.74
+2	−0.13	−1.69	−0.08	1.9
+3	1.253	0.596	−0.165	−1.789
+4	−1.836	0.826	0.537	0.826
+5	1.94	−1.764	−1.139	0.706

The analysis was performed in HOISAN.

TABLE 8 | Polar coordinate analysis results showing the length and angle of the different vectors, the quadrant in which each vector is located, and the Z_{sum} values (Cochran, 1954) from the prospective and retrospective perspectives.

Catergory	Quadrant	Prospective	Retrospective	Length	Angle
SC_A	III	−0.44	−0.44	0.63	225
SC_B	IV	0.53	−0.13	0.54	346.19
SC_C	II	−0.38	0.52	0.65	125.79
SC_D	I	0.4	0.09	0.41	12.6

In the presented situation, A is the focal behavior, so the results show how expressions of sorrow or sadness activate expressions of self-perceived improvement (Quadrant IV) or joy at having overcome the problem (Quadrant I). The focal behavior is not self-generating (Quadrant III). Additionally, expressions of sorrow or sadness do not generate self-perceived worsening (Quadrant II), although self-perceived worsening does generate the focal behavior.

unconditional probability), we extracted the behavioral pattern shown in **Figure 2**. The strength of patterns is assessed using interpretative rules (Bakeman and Gottman, 1987). In the example provided, the first lag that is followed by another lag containing significant categories is considered to be the last lag (max lag) in the pattern (lag 3 in the example).

The robustness of the pattern must then be further strengthened by building a confidence interval around the conditional probabilities, for which only the upper limit is needed. This upper limit is used to determine whether a given category will form part of the pattern at the lag being analyzed, as the conditional probability obtained has to be higher than unconditional probability. The lower limit, by contrast, will always be lower than the unconditional probability and as such, will never be significant. Application of this confidence interval increases the requirements for statistical significance for the categories at each lag, resulting in a more robust corrected pattern.

The results obtained by applying the formula corresponding to the corrected expected or unconditional probability (shown in **Table 6a**) are presented in **Table 6b**, which is an extension of **Table 5b**.
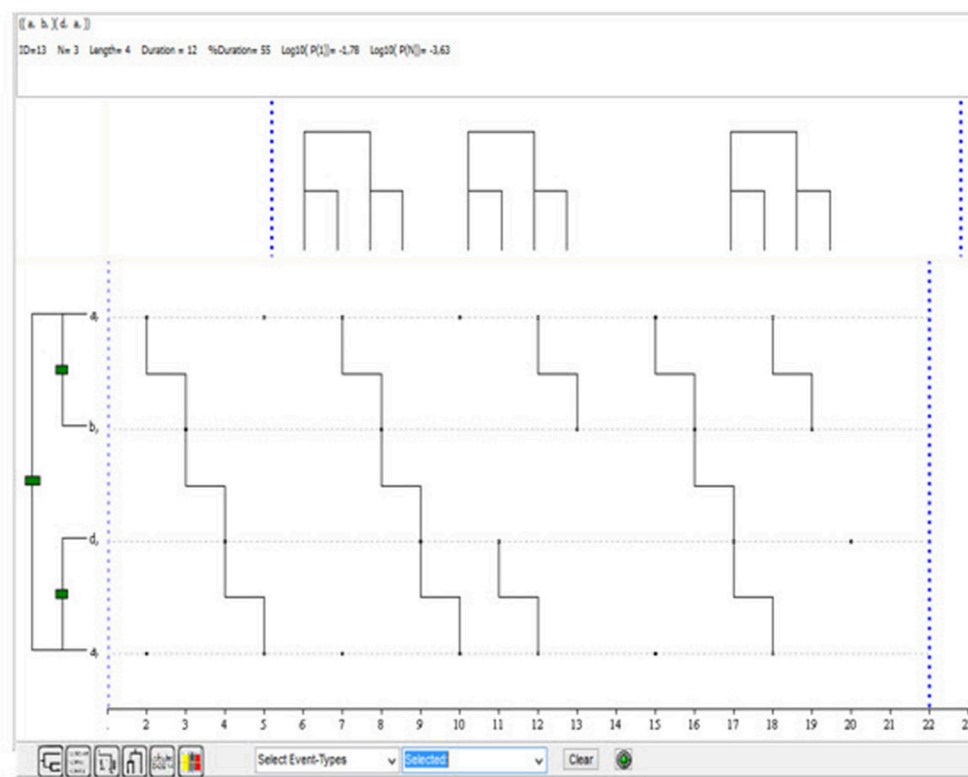
FIGURE 5 | First of the 13 T-patterns detected in the data from **Table 3a** ($p < 0.05$).

A second optimization step involving the calculation of adjusted residuals or hypergeometric Z-values (Allison and Liker, 1982) is also possible but cannot be done manually.

Figure 4 shows the corrected behavioral pattern extracted from the data in **Table 6b**. As shown, it is different to the uncorrected pattern shown in **Figure 3**. Note that in both cases, A, the given behavior, is statistically associated with B at the first lag and D at the second lag.

Lag sequential analysis is the first of the three key techniques we use in our text-liquefying approach to indirect observation. It has been widely used in systematic observation studies from a range of areas published in journals listed in the Journal Citations Report (JCR) (e.g., Gimeno et al., 2006; Lapresa et al., 2013; Roustan et al., 2013).

Polar Coordinate Analysis

Polar coordinate analysis, which was proposed by Sackett (1980), combines adjusted residuals from lag sequential analysis and the Z_{sum} statistic (Cochran, 1954). This statistic provides a representative value for a series of independent values (adjusted residuals at different prospective or retrospective negative lags) to produce prospective and retrospective Z_{sum} values. Sackett (1980) recommended using the same number of prospective and retrospective lags. Based on experience to date (Sackett, 1987; Anguera and Losada, 1999), we suggest analyzing at least five prospective lags and five retrospective lags (-5 to $+5$).

The results of the computation determine the quadrant in which the different vectors are located and indicate their respective lengths and angles (Sackett, 1980). Vectors provide information on the nature of the relationship (prospective/retrospective activation/inhibition) between a focal behavior, which is equivalent to a given behavior in lag sequential analysis, and other categories of interest, known as conditional behaviors. The concept of genuine retrospectivity (Anguera, 1997) was introduced at a later stage to improve the classic concept of retrospectivity. The genuine retrospective approach considers negative lags from a backwards rather than a forwards perspective, i.e., it looks at what happened from lag 0 back to lag -5 rather than from lag -5 to lag 0.

Adjusted residuals, Z-values, and vector length and angles can all be computed in the open-access software program HOISAN (v. 1.6.3.3) (Hernández-Mendo et al., 2012), which also includes a feature to produce the results in graph form.

The meaning of the vectors (see below) varies according to the quadrant in which they are located, and the position of a vector in one quadrant or another is determined by the combination of positive or negative signs on the prospective and retrospective Z_{sum} values. In quadrant I ($+ +$), the focal and conditional behaviors activate each other; in quadrant II ($- +$), the focal behavior inhibits and is activated by the conditional behavior; in quadrant III ($- -$), the focal and conditional behaviors inhibit each other; and in quadrant IV ($+ -$), the focal behavior activates and is inhibited by the conditional behavior. The length of

the vectors indicates the strength (statistical significance) of the association between the focal and conditional behaviors.

To illustrate briefly how the technique works, we used the data from **Table 3a** to produce a vector map showing the relationships between A, the focal behavior (in our example, expressions of sorrow or sadness), and categories B (expressions of self-perceived improvement), C (expressions of self-perceived worsening), and D (expressions of joy at having overcome the problem), the conditional behaviors. **Table 7** shows the values for the adjusted residuals and corresponding Z_{sum} values, while **Table 8** shows the length and angle of the vectors for each of the conditional behaviors. The corresponding vectors are shown in **Figure 4**.

The strongest association detected for the focal behavior A (apart from with itself) was with B (in quadrant IV, with a vector length of 0.54), followed by D (quadrant I, with a vector length of 0.41). Although A and C have the longest vector (0.65), the fact that C is located in quadrant II (because its angle is 125.79°) means that A inhibits rather than activates C. C does not appear because its excitatory activity was insignificant.

Readers can find numerous examples of the application of polar coordinate analysis in a wide range of fields in direct observation (e.g., Gorospe and Anguera, 2000; Herrero Nivela, 2000; Anguera et al., 2003; Castañer et al., 2016, 2017; López et al., 2016; Aragón et al., 2017; Morillo et al., 2017; Santoyo et al., 2017; Suárez et al., 2018), and more recently indirect observation (e.g. Arias-Pujol and Anguera, 2017).

T-Pattern Detection

T-pattern detection was proposed and developed by Magnusson (1996, 2000, 2005, 2016). It involves the use of an algorithm that calculates the temporal distances between behaviors and analyzes the extent to which the critical interval remains invariant relative to the null hypothesis that each behavior is independently and randomly distributed over time. It needs data, in the form of code matrices, for which the duration of each co-occurrence has been recorded. Microanalyses of data are also possible and very useful (Anguera, 2005). The software program, Theme (v. 6 Edu), features different settings that can be modified to obtain complementary results that, analyzed together, can provide a greater understanding of interactive transitions over time. Theme is an open-access software program that provides all the necessary features for analyzing data and presenting the results graphically as dendrograms or tree diagrams.

As with lag sequential and polar coordinate analysis, we have also used the data from **Table 3a** to illustrate the use of T-pattern detection. It should be noted that the method applied is rather unconventional, as the temporal distance parameter was set at 1 in all cases.

Figure 5 shows the first of the 13 T-patterns obtained ($p < 0.05$). Note that despite the small size of the dataset, Theme detected a primary relationship between A and B (between expressions of sorrow or sadness and expressions of self-perceived improvement) and A and D (between expressions of sorrow or sadness and expressions of joy at having overcome the problem), as shown graphically in **Figure 5**.

Examples of the application of T-pattern detection can be found in studies by Castañer et al. (2013), Diana et al. (2017), Lapresa et al. (2013), and Sarmento et al. (2015) in direct observation and by Blanchet et al. (2005) and Baraud et al. (2016) in indirect observation.

Complementary Use of Techniques

Although the specifics of lag sequential analysis, polar coordinate analysis, and T-pattern detection differ, all three techniques serve to analyze and increase understanding of the internal structure of verbal or textual material derived from indirect observation. In addition, they can be applied to the same data to provide complementary insights and unveil invisible structures hidden within data. Their relevance is even greater in indirect observation studies where data have traditionally been analyzed from a purely qualitative perspective.

The convergence of results from three different quantitative approaches is a cause for celebration in a field such as indirect observation, where studies to date have largely relied on frequency counts or on qualitative approaches, which of course have their merits but are prone to considerable subjectivity bias.

There is growing interest in combining these techniques to gain a greater understanding of behavioral patterns that remain hidden to the naked eye. Two recent examples can be found in the studies of Santoyo et al. (2017) and Tarragó et al. (2017).

ADAPTED METHODOLOGICAL PROCEDURE FOR CONDUCTING AN INDIRECT OBSERVATION STUDY BASED ON TEXT LIQUEFACTION

We have presented a structured procedure detailing the successive stages of the method we propose for studying verbal behavior and/or textual material in an indirect observation study (**Table 9**). Our aim was not to offer a general approach to systematic observation from the perspective of indirect observation, as guidelines already exist for the reporting of systematic studies within observational methodology (Portell et al., 2015a). Our aim rather was to introduce the reader to the key concepts of indirect observation studies and provide step-by-step guidance on how to perform such a study. The procedure we propose is summarized in **Table 9** and has already been applied in studies from different fields (Vaimberg, 2010; García-Fariña et al., 2016; Arias-Pujol and Anguera, 2017).

CONCLUSIONS AND LIMITATIONS

Within the broad framework of mixed methods, we have presented indirect observation as a structured method consisting of different steps designed to guarantee scientific rigor. The method consists of the quantitization of qualitative data derived from verbal or textual material to produce code matrices which, following appropriate organization and rigorous quality control procedures, can be analyzed using robust, rigorous, and objective

TABLE 9 | Procedure for conducting an indirect observation study based on liquefying a text.

Methodological action	Description
Guidelines for Indirect Observation (Liquefying a Text)	
Define the research question	Focus the question on aspects that can be clearly delimited.
Delimit the source(s) of verbal or textual information using clearly specified criteria (setting, participants, situations, etc.)	Take all necessary decisions about the sources of data, such as type, volume, time frame, geographic location (if relevant) vs. online, etc.
Specify the study dimensions	After a detailed analysis of the theoretical framework, decide on the dimensions (or facets) of the research question.
Establish the text segmentation criteria	If the study is multidimensional, choose the primary dimension and define the segmentation criteria accordingly. This step will influence the segmentation of all the other dimensions, which will be considered secondary for this purpose.
Build an *ad hoc* observation instrument	Consider the number of dimensions, the existence or not of a theoretical framework, and the temporal nature of the subject of study (process vs. atemporal situation). Validate the coding process.
Code the data and create code matrices	Apply the codes from the observation instrument to the data to transform or 'liquefy' the primary material (verbal behavior or textual material) into matrices of codes suited for quantitative analysis.
Computerize the codes	Depending on the features of the software programs available, convert the dataset into a computerized format.
Merge/divide the code matrices in accordance with specific research questions	These data block management operations must be highly flexible as it may sometimes be necessary to jointly analyze several code matrices or to analyze parts of the same matrix separately.
Apply rigorous data quality controls	Rigorous data controls prior to the analysis of the data are essential for preventing possible biases from skewing the results. Such controls are necessary as the datasets are prone to subjectivity bias. Each set of textual units should be coded at least three times (by the same observer or by different observers).
Analyze the data quantitatively using a suitable technique or techniques	Once the quality of the dataset has been confirmed, the code matrices can be analyzed quantitatively to reveal underlying structures in the form of significant associations between codes. Choose the most appropriate technique depending on the aim of the analysis. A) Use lag sequential analysis to discover behavioral patterns that occur more often than would be expected by chance. B) Use polar coordinate analysis to obtain a vector map showing the prospective and retrospective activating or inhibitory relationships between a focal behavior and other behaviors of interest. C) Use T-pattern detection to uncover temporal relationships between categories based on the time distance between successive occurrences of each code.
Analyze convergent or complementary results if various techniques have been used	Compare and analyze similarities detected using different techniques and explore possible explanations for divergent results.
Evaluate the strengths and weaknesses of the study	The detection of strengths will allow you to argue the objectivity, rigor, and robustness of the study. Consider potential weaknesses by critically appraising the methodology, studying the literature, and evaluating the consistency of the theoretical framework.
Interpret your results	Interpret your results by analyzing them in the light of your research question(s) and considering similarities and differences reported by related studies.
Conduct a thorough and up-to-date literature review (although this is mentioned as the last step, relevant literature should be investigated and read throughout the study)	Conduct an exhaustive preliminary literature review and then apply rigorous filters as appropriate.

techniques. In a sense, we liquefy the text into a form suitable for quantitative analysis.

Although the materials that support direct and indirect observation are different, the methodological proposal described in this paper shows that both forms of observation share a systematic procedure in which adequately trained observers apply a robust, reliable purpose-designed observation instrument to produce quantitative indicators of the many processes underlying everyday behavior. The main strengths of our approach are that it enables the merging of data from different sources and offers the possibility of taking advantage of the continuous advances in information and communication technologies to study aspects of biopsychosocial behavior in everyday contexts. There are two main limitations. On the one hand, the dimensions in an indirect observation study depend largely on a theoretical framework and a conceptual

framework, and these may be lacking. On the other hand, observation instruments comprising category systems, either alone or combined with a field format, also require a theoretical framework. However, the proposed approach has the advantage of allowing all data obtained from narratives to be included in the study, even those which do not fit with the theoretical framework or are contradictory. In fact, the validation of the coding process entails, among other things, checking that no new information has been added, that no information has been eliminated, and that the meaning of the information has not been altered. In this way, there is no omission of information that could lead to bias. This information can be included using bottom-up or top-down processes (Anguera, 1991; Anguera et al., 2007), in other words, the narratives are categorized on the basis of the chosen theoretical framework (top-down) and the theoretical framework is adapted on the basis of the narratives given (bottom-up). An

exclusively quantitative study would entail the loss of sensitive and relevant information about the spontaneous behavior, as it would require excluding all variables not envisaged in the chosen theoretical framework. Hence our insistence on the enormous potential of mixed methods research, which suitably integrates both qualitative and quantitative elements.

This work presented a novel approach, based on sequence of occurrence, for transforming qualitative data into quantitative data that can be analyzed using robust quantitative techniques. Additionally, it is important to note that it is possible, at any time during the analysis, to return from the quantitative data to the narrative data. As a result, this approach presents advantages of both qualitative and quantitative methods, at the same time it covers weaknesses of both methods.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Allison, P. D., and Liker, J. K. (1982). Analyzing sequential categorical data on dyadic interaction: a comment on Gottman. *Psychol. Bull.* 91, 393–403. doi: 10.1037/0033-2909.91.2.393

Altimir, C., Krause, M., de la Parra, G., Dagnino, P., Tomicic, A., Valdés, N., et al. (2010). Clients', therapists', and observers' agreement on the amount, temporal location, and content of psychotherapeutic change and its relation to outcome. *Psychother. Res.* 20, 472–487. doi: 10.1080/10503301003705871

Anguera, M. T. (1979). Observational typology. *Qual. Quant.* 13, 449–484.

Anguera, M. T. (1990). "Metodología observacional [Observational methodology]," in *Metodología de la Investigación en Ciencias del Comportamiento*, eds J. Arnau, M. T. Anguera, and J. Gómez (Murcia: Universidad de Murcia), 125–238.

Anguera, M. T. (1991). La metodología observacional en evaluación de programas [Observational methodology in program evaluation]. *Rev. Mex. Anal. Conducta* 17, 121–145.

Anguera, M. T. (1997). "From prospective patterns in behavior to joint analysis with a retrospective perspective," in *Colloque sur Invitation Méthodologie d'Analyse des Interactions Sociales*. Paris: University of Paris V, Sorbonne.

Anguera, M. T. (2001). "Hacia una evaluación de la actividad cotidiana y su contexto: presente o futuro para la metodología? [Towards an evaluation of everyday activity and its context: present or future for methodology? Lecture of admission to the Royal Europan Academy of Doctors in 1999]," in *Estrategias de Evaluación y Medición del Comportamiento en Psicología [Strategies of Evaluation and Measurement of the Behavior in Psychology]*, eds A. Bazán Ramírez and A. Arce Ferrer (México: Instituto Tecnológico de Sonora y Universidad Autónoma de Yucatán), 11–86.

Anguera, M. T. (2003). "Observational methods (general)," in *Encyclopedia of Psychological Assessment*, Vol. 2, ed R. Fernández-Ballesteros (London: Sage), 632–637.

Anguera, M. T. (2005). "Microanalysis of T-patterns. Analysis of symmetry/asymmetry in social interaction," in *The Hidden Structure of Social Interaction. From Genomics to Culture Patterns*, eds L. Anolli, S. Duncan, M. Magnusson, and G. Riva (Amsterdam: IOS Press), 51–70.

Anguera, M. T. (2010). Posibilidades y relevancia de la observación sistemática por el profesional de la Psicología [Possibilities and relevance of systematic observation for pyschology professionals]. *Papeles Psicol.* 31, 122–130.

Anguera, M. T. (2017). "Transiciones interactivas a lo largo de un proceso de desarrollo: complementariedad de análisis [Interactive transitions throughout a development process. A complementary analysis approach]," in *Mecanismos Básicos de Toma de Decisiones: Perspectivas Desde las Ciencias del Comportamiento y Del Desarrollo*, ed C. Santoyo (México: CONACYT 178383/UNAM), 179–213.

Anguera, M. T. (in press). "Is it possible to perform "Liquefying" actions in conversational analysis? The detection of structures in indirect observation," in *The Temporal Structure of Multimodal Communication*, ed L. Hunyadi (New York, NY: Springer).

Anguera, M. T., Camerino, O., Castañer, M., Sánchez-Algarra, P., and Onwuegbuzie, A. J. (2017a). The specificity of observational studies in physical activity and sports sciences: moving forward in mixed methods research and proposals for achieving quantitative and qualitative symmetry. *Front. Psychol.* 8:2196. doi: 10.3389/fpsyg.2017.02196

Anguera, M. T., and Hernández-Mendo, A. (2016). Avances en estudios observacionales en Ciencias del Deporte desde los mixed methods [Advances in observational studies in Sport Sciences from a mixed methods approach]. *Cuad. Psicol. Deporte* 16, 17–30.

Anguera, M. T., and Izquierdo, C. (2006). "Methodological approaches in human communication. From complexity of situation to data analysis," in *From Communication to Presence. Cognition, Emotions and Culture towards the Ultimate Communicative Experience*, eds G. Riva, M. T. Anguera, B. K. Wiederhold, and F. Mantovani (Amsterdam: IOS Press), 203–222.

Anguera, M. T., Jonsson, G. K., and Sánchez-Algarra, P. (2017b). Liquefying text from human communication processes: a methodological proposal based on t-pattern detection. *J. Multimodal. Comm. Stud.* 4, 10–15.

Anguera, M. T., and Losada, J. L. (1999). "Reducción de datos en marcos de conducta mediante la técnica de coordenadas polares [Data reduction in behavioral frameworks through polar coordinate analysis]," in *Observación de la Conducta Interactiva en Situaciones Naturales: Aplicaciones*, ed M. T. Anguera (Barcelona: E.U.B.), 163–188.

Anguera, M. T., Magnusson, M. S., and Jonsson, G. K. (2007). Instrumentos no estándar [Non-standard instruments]. *Avances Med.* 5, 63–82.

Anguera, M. T., Santoyo, C., and Espinosa, M. C. (2003). "Evaluating links intensity in social networks in a school context through observational designs," in *Culture, Environmental Action and Sustainability*, eds R. García Mira, J. M. Sabucedo Cameselle, and J. Romay Martínez (Göttingen: Hogrefe and Huber), 286–298.

Aragón, S., Lapresa, D., Arana, J., Anguera, M. T., and Garzón, B. (2017). An example of the informative potential of polar coordinate analysis: Sprint tactics in elite 1500 m track events. *Meas. Phys. Educ. Exerc. Sci.* 21, 26–33. doi: 10.1080/1091367X.2016.1245192

Arana, J., Lapresa, D., Anguera, M. T., and Garzón, B. (2016). Ad hoc procedure for optimising agreement between observational records. *Anal. Psicol.* 32, 589–595. doi: 10.6018/analesps.32.2.213551

Arias-Pujol, E., and Anguera, M. T. (2004). Detección de patrones de conducta comunicativa en un grupo terapéutico de adolescentes [Detection of communicative behavior patterns in an adolescent group therapy group]. *Acción Psicol.* 3, 199–206. doi: 10.5944/ap.3.3.513

Arias-Pujol, E., and Anguera, M. T. (2017). Observation of interactions in adolescent group therapy: a mixed methods study. *Front. Psychol.* 8:1188. doi: 10.3389/fpsyg.2017.01188

Bakeman, R. (1978). "Untangling streams of behavior: sequential analysis of observation data," in *Observing Behavior, Vol. 2: Data Collection and Analysis Methods*, ed G. P. Sackett (Baltimore, MD: University of Park Press), 63–78.

Bakeman, R. (1991). From lags to logs: advances in sequential analysis. *Rev. Mex. Anal. Conducta* 17, 65–83.

Bakeman, R., and Gottman, J. M. (1987). "Applying observational methods: a systematic view," in *Handbook of Infant Development*, ed J. D. Osofsky (New York, NY: Wiley), 818–853.

Bakeman, R., and Quera, V. (1996). *Análisis de la Interacción. Análisis Secuencial con SDIS y GSEQ [Analysis of Interaction. Sequential Analysis with SDIS and GSEQ]*. Madrid: Ra-Ma.

Bakeman, R., and Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge: Cambridge University Press.

Baraud, I., Deputte, B. L., Pierre, J.-S., and Blois-Heulin, C. (2016). "Informative value of vocalizations during multimodal interactions in red-capped mangabeys," in *Discovering Hidden Temporal Patterns in Behavior and Interactions: T-Pattern Detection and Analysis with THEME,* eds M. S. Magnusson, J. K. Burgoon, M. Casarrubea, and D. McNeill (New York, NY: Springer), 255–278.

Bautista, M. A., Hernández-Vela, A., Escalera, S., Igual, L., Pujol, O., Mora, J., et al. (2015). A gesture recognition system for detecting behavioral patterns of ADHD. *IEEE Trans. Cybern.* 46, 136–147. doi: 10.1109/TCYB.2015.2396635

Bavelas, J. B., and Chovil, N. (2000). Visible acts of meaning. An integrated message model of language in face-to-face dialogue. *J. Lang. Soc. Psychol.* 19, 163–194. doi: 10.1177/0261927X00019002001

Bazeley, P. (2003). "Computerized data analysis for mixed methods research," in *Handbook of Mixed Methods in Social and Behavioral Research*, eds A. Tashakkori, and C. Teddlie (Thousand Oaks, CA: Sage), 385–422.

Bazeley, P. (2006). The contribution of computer software to integrating qualitative and quantitative data and analyses. *Res. Sch.* 13, 64–74.

Bazeley, P. (2009). Integrating data analyses in mixed methods research. *J. Mix. Methods Res.* 3, 203–207. doi: 10.1177/1558689809334443

Behar, J. (1993). "Observación y análisis de la producción verbal de la conducta [Observation and analysis of the verbal production of behavior]," in *Metodología Observacional en la Investigación Psicológica. Vol. 1. Fundamentación*, ed M. T. Anguera (Barcelona: P.P.U.), 331–389.

Bell, K., Fahmy, E., and Gordon, D. (2016). Quantitative conversations: the importance of developing rapport in standardised interviewing. *Qual. Quant.* 50, 193–212. doi: 10.1007/s11135-014-0144-2

Bion, W. R. (1985). *Experiencias en Grupos [Experiences in Groups]*. Buenos Aires: Paidós.

Birdwhistell, R. L. (1970). *Kinesics and Context: Essays and Body Motion Communication*. Philadelphia, PA: University of Pennsylvania Press.

Björk, A. B., Sjöström, M., Johansson, E. E., Samuelson, E., and Umefjord, G. (2014). Women's experiences of Internet-based or postal treatment for stress urinary incontinence. *Qual. Health Res.* 24, 484–493. doi: 10.1177/1049732314524486

Blanchet, A., Batt, M., Trognon, A., and Masse, L. (2005). "Language and behaviour patterns in a therapeutic interaction sequence," in *The Hidden Structure of*

Social Interaction. From Genomics to Culture Patterns*, eds L. Anolli, S. Duncan, M. S. Magnusson, and G. Riva (Amsterdam: IOS Press), 124–140.

Blanco-Villaseñor, A. (2001). Generalizabilidad de observaciones uni y multifaceta: estimadores LS y ML [Generalizability of mono and multifaceted observations: LS and ML estimators]. *Metodol. Cienc. Comport.* 3, 161–193.

Blanco-Villaseñor, A., Losada, J. L., and Anguera, M. T. (2003). Analytic techniques in observational designs in environment behavior relation. *Medio Ambient. Comport. Hum.* 4, 111–126.

Bolger, N., Davis, A., and Rafaeli, E. (2003). Diary methods: capturing life as it is lived. *Annu. Rev. Psychol.* 54, 579–616. doi: 10.1146/annurev.psych.54.101601.145030

Bradley, E. H., Curry, L. A., and Devers, K. J. (2007). Qualitative data analysis for health services research: developing taxonomy, themes, and theory. *Health Serv. Res.* 42, 1758–1772. doi: 10.1111/j.1475-6773.2006.00684.x

Brondani, M. A., MacEntee, M. I., Bryant, S. R., and O'Neill, B. (2008). Using written vignettes in focus groups among older adults to discuss oral health as a sensitive topic. *Qual. Health Res.* 18, 1145–1153. doi: 10.1177/1049732308320114

Bunker, D., and Thorpe, R. (1982). A model for the teaching of games in secondary schools. *Bull. Phys. Educ.* 18, 5–8.

Caddick, N., Smith, B., and Phoenix, C. (2015). Male combat veterans' narratives of PTSD, masculinity, and health. *Sociol. Health Illn.* 37, 97–111. doi: 10.1111/1467-9566.12183

Calsamiglia, H., and Tusón, A. (1999). *Las Cosas del Decir. Manual de Análisis del Discurso [Things About Saying. A Discourse Analysis Manual]*. Barcelona: Ariel.

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Canton, A., and Hileno, R. (2016). Goal scoring in soccer: a polar coordinate analysis of motor skills used by Lionel Messi. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Fernandes, T., and Hileno, R. (2017). Mastery in goal scoring, T-pattern detection and polar coordinate analysis of motor skills used by Lionel Messi and Cristiano Ronaldo. *Front. Psychol.* 8:741. doi: 10.3389/fpsyg.2017.00741

Castañer, M., Camerino, O., Anguera, M. T., and Jonsson, G. K. (2013). Kinesics and proxemics communication of expert and novice PE teachers. *Qual. Quant.* 47, 1813–1829. doi: 10.1007/s11135-011-9628-5

Cochran, W. G. (1954). Some methods for strengthening the common $\chi 2$ tests. *Biometrics* 10, 417–451. doi: 10.2307/3001616

Coll, C., and Onrubia, J. (2001). Estrategias discursivas y recursos semióticos en la construcción de sistemas de significados compartidos entre profesor y alumnos [Discursive strategies and semiotic resources in the construction of shared meaning systems between teachers and students]. *Invest. Esc.* 45, 21–31.

Conner, T. S., and Lehman, B. J. (2013). "Getting started. Lauching a study in daily life," in *Handbook of Research Methods for Studying Daily Life*, eds M. R. Mehl and T. S. Conner (New York, NY: The Guilford Press), 89–105.

Coutinho, J., Ribeiro, E., Sousa, I., and Safran, J. D. (2014). Comparing two methods of identifying alliance rupture events. *Psychotherapy* 51, 434–442. doi: 10.1037/a0032171

Crawford, T. H. (1992). The politics of narrative form. *Lit. Med.* 11, 147–162. doi: 10.1353/lm.2011.0234

Creswell, J. W., and Plano Clark, V. L. (2007). *Designing and Conducting Mixed Methods Research, 3rd Edn 2017*. Thousand Oaks, CA: Sage.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., and Hanson, W. E. (2003). "Advanced mixed methods research designs," in *Handbook of Mixed Methods in Social and Behavioral Research*, eds A. Tashakkori, and C. Teddlie (Thousand Oaks, CA: Sage), 209–240.

Crutcher, R. J. (2003). A computer-aided digital audio recording and encoding system for improving the encoding of verbal reports. *Behav. Res. Methods* 35, 263–268. doi: 10.3758/BF03202551

Crutcher, R. J. (2007). CAPAS 2.0: a computer tool for coding transcribed and digitally recorded verbal reports. *Behav. Res. Methods* 39, 167–174. doi: 10.3758/BF03193145

Cuervo, J. J. (2014). *Comunicación y Perdurabilidad en Parejas Viables. Estudio Observacional de Caso Múltiple [Communication and Durability in Viable Couples. A Multiple Observational Case Study]*. Ph.D. thesis, Universidad Autónoma de Barcelona: Barcelona.

Dagnino, P., Krause, M., Pérez, J. C., Valdés, N., and Tomicic, A. (2012). The evolution of communicative intentions during change episodes and throughout the therapeutic process. *Res. Psychother.* 15, 75–86.

Dam, G., and Kaufmann, S. (2008). Computer assessment of interview data using latent semantic analysis. *Behav. Res. Methods* 40, 8–20. doi: 10.3758/BRM.40.1.8

Danzinger, K. (1982). *Comunicación Interpersonal [Interpersonal Communication]*. México: Manual Moderno.

De Fina, A., and Georgakopoulo, A. (eds.). (2015). *The handbook of narrative analysis.* Malden: Wiley-Blackwell.

Diana, B., Zurloni, V., Elia, M., Cavalera, C. M., Jonsson, G. K., and Anguera, M. T. (2017). How game location affects soccer performance: T-pattern analysis of attack actions. *Front. Psychol.* 8:1415. doi: 10.3389/fpsyg.2017.01415

Dickman, H. R. (1963). "The perception of behavioral units," in *The Stream of Behavior*, ed R. Barker (New York, NY: Appleton-Century-Crofts), 23–41.

Duran, N. D., McCarthy, P. M., Graesser, A. C., and McNamara, D. S. (2007). Using temporal cohesion to predict temporal coherence in narrative and expository texts. *Behav. Res. Methods* 39, 212–223. doi: 10.3758/BF031 93150

Elvish, R., Lever, S. J., Johnstone, J., Cawley, R., and Keady, J. (2013). Psychological interventions for careers of people with dementia: a systematic review of quantitative and qualitative evidence. *Counsell. Psychother. Res. J.* 13, 106–125. doi: 10.1080/14733145.2012.739632

Escalera, S., Martínez, R. M., Vitriá, J., Radeva, P., and Anguera, M. T. (2009). Automatic dominance detection in dyadic conversations. *Escr. Psicol.* 3, 41–45.

Escolano-Pérez, E., Herrero-Nivela, M. L., Blanco-Villaseñor, A., and Anguera, M. T. (2017). Systematic observation: relevance of this approach in preschool executive function assessment and association with later academic skills. *Front. Psychol.* 8:2031. doi: 10.3389/fpsyg.2017.02031

Fetters, M. D., and Freshwater, D. (2015). The $1 + 1 = 3$ integration challenge. *J. Mix. Methods Res.* 9, 115–117. doi: 10.1177/1558689815581222

Fischer, A. H., Becker, D., and Veenstra, L. (2012). Emotional mimicry in social context: the case of disgust and pride. *Front. Psychol.* 3:475. doi: 10.3389/fpsyg.2012.00475

Gabriel, Y. (2004). "The voice of experience and the voice of the expert - Can they speak to each other?," in *Narrative Research in Health and Illness,* eds B. Hurwitz, T. Greenhalgh, and V. Skultans (Malden, MA: Blackwell), 168–185.

Gale, J. E. (1991). *Conversation Analysis of Therapeutic Discourse*. Norwood, OH: Ablex Publishing Corporation.

García-Fariña, A. (2015). *Análisis del Discurso Docente Como Recurso Metodológico del Profesorado de Educación Física en la Etapa de Educación Primaria [Analysis of Teacher-Led Discourse as a Methodological Resource for Primary School Physical Education Teachers]*. Ph.D. thesis, Universidad de La Laguna, Tenerife.

García-Fariña, A., Jiménez Jiménez, F., and Anguera, M. T. (2016). Análisis observacional del discurso docente del profesorado de educación física en formación a través de patrones comunicativos [Observational analysis of teaching discourse physical education training teachers through communicative patterns]. *Cuad. Psicol. Deporte* 16, 171–182.

Gimeno, A., Anguera, M. T., Berzosa, A., and Ramírez, L. (2006). Detección de patrones interactivos en la comunicación de familias con hijos adolescentes [Detection of interactive communication patterns in families with adolescent children]. *Psicothema* 18, 785–790.

Gorospe, G., and Anguera, M. T. (2000). Modificación de la técnica clásica de coordenadas polares mediante un desarrollo distinto de la retrospectividad: aplicación al tenis [Modification to the classic polar coordinate technique using a distinct concept of retrospectivity: an application to tennis]. *Psicothema* 12(Suppl. 2), 279–282.

Greene, J. C. (2006). Toward a methodology of mixed methods social inquiry. *Res. Sch.* 13, 93–98.

Hardley, D. M. (2014). Using social media and Internet data for public health surveillance: the importance of talking. *Milbank Q.* 92, 34–39. doi: 10.1111/1468-0009.12039

Hernández-Mendo, A., López-López, J. A., Castellano, J., Morales-Sánchez, V., and Pastrana, J. L. (2012). Hoisan 1.2: programa informático para uso en metodología observacional [Hoisan 1.2: software for observational methodology]. *Cuad. Psicol. Deporte* 12, 55–78. doi: 10.4321/S1578-84232012000100006

Herrero Nivela, M. L. (2000). Utilización de la técnica de coordenadas polares en el estudio de la interacción infantil en el marco escolar [Use of the technic of polar coordinates in the study of the children interaction in the school room]. *Psicothema* 12, 292–297.

Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., and Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Front. Psychol.* 3:74. doi: 10.3389/fpsyg.2012.00074

Holtgraves, T., and Han, T. L. (2007). A procedure for studying online conversational processing using a chat bot. *Behav. Res. Methods* 39, 156–163. doi: 10.3758/BF03192855

Hurwitz, B., Greenhalgh, T., and Skultans, V. (eds.). (2004). *Narrative Research in Health and Illness*. London: British Medical Journal Books and Blackwell.

Jackson, M., Harrison, P., Swinburn, B., and Lawrence, M. (2015). Using a qualitative vignette to explore a complex public health issue. *Qual. Health Res.* 25, 1395–1409. doi: 10.1177/1049732315570119

Johnson, R. B., Onwuegbuzie, A. J., and Turner, L. A. (2007). Toward a definition of mixed methods research. *J. Mix. Methods Res.* 1, 112–133. doi: 10.1177/1558689806298224

Krause, M., Altimir, C., Pérez, J. C., Echávarri, O., Valdés, N., and Strasser, K. (2016). Therapeutic verbal communication in change episodes: a comparative microanalysis of linguistic basic forms. *Estud. Psicol.* 37, 514–547. doi: 10.1080/02109395.2016.1227575

Krippendorff, K. (2013). *Content Analysis. An Introduction to Its Methodology, 3rd Edn*. Thousand Oaks, CA: Sage.

Krueger, R. A., and Casey, M. A. (2009). *Focus Groups. A Practical Guide for Applied Research*. Los Angeles, CA: Sage.

Lapresa, D., Arana, J., Anguera, M. T., and Garzón, B. (2013). Comparative analysis of the sequentiality using SDIS-GSEQ and THEME: a concrete example in soccer. *J. Sports Sci.* 31, 1687–1695. doi: 10.1080/02640414.2013. 796061

Lausberg, H., and Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES–ELAN system. *Behav. Res. Methods* 41, 841–849. doi: 10.3758/BRM.41.3.841

Lausberg, H., and Sloetjes, H. (2016). The revised NEUROGES-ELAN system: an objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture. *Behav. Res. Methods* 48, 973–993. doi: 10.3758/s13428-015-0622-z

Lebrecht, S., Bar, M., Barrett, L. F., and Tarr, M. J. (2012). Micro-valences. Perceiving affective valence in everyday objects. *Front. Psychol.* 3:107. doi: 10.3389/fpsyg.2012.00107

Lederer, S. H., and Battaglia, D. (2015). Using signs to facilitate vocabulary in chldren with language delays. *Infants Young Child.* 28, 18–31. doi: 10.1097/IYC.0000000000000025

Leech, N. L., and Onwuegbuzie, A. J. (2009). A typology of mixed methods research designs. *Qual. Quant.* 43, 265–275. doi: 10.1007/s11135-007-9105-3

Lévy, P. (1995). *Qué es lo Virtual? [What is Virtual?]*. Barcelona: Paidós.

Lindqvist, O., Threlkeld, G., Street, A. F., and Tishelman, C. (2014). Reflections on using biographical approaches in end-of-life-care: dignity therapy as example. *Qual. Health. Res.* 25, 40–50. doi: 10.1177/1049732314549476

López, J., Valero, A., Anguera, M. T., and Díaz, A. (2016). Disruptive behavior among elementary students in physical education. *Springerplus* 5:1154. doi: 10.1186/s40064-016-2764-6

Love, A. C. (2006). History, scientific methodology, and the "squishy" sciences. *Perspect. Biol. Med.* 49, 452–456. doi: 10.1353/pbm.2006.0042

Magnusson, M. S. (1996). Hidden real-time patterns in intra- and inter-individual behavior. *Eur. J. Psychol. Assess.* 12, 112–123. doi: 10.1027/1015-5759.12.2.112

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behav. Res. Methods Instrum. Comput.* 32, 93–110. doi: 10.3758/BF03200792

Magnusson, M. S. (2005). "Understanding social interaction: discovering hidden structure with model and algorithms," in *The Hidden Structure of Social Interaction. From Genomics to Culture Patterns,* eds L. Anolli, S. Duncan, M. S. Magnusson, and G. Riva (Amsterdam: IOS Press), 4–22.

Magnusson, M. S. (2016). "Time and self-similar structure in behavior and interactions: from sequences to symmetry and fractals," in *Discovering Hidden Temporal Patterns in Behavior and Interactions: T-Pattern Detection and Analysis with THEME,* eds M. S. Magnusson, J. K. Burgoon, M. Casarrubea, and D. McNeill (New York, NY: Springer), 3–35.

Martínez del Pozo, M. (1993). *Análisis del "Proceso de Elaboración del Duelo" en el Hipertenso Esencial: Estudio Empírico de Sus Fases Mediante Observación Sistemática [Analysis of the "Elaboration of Mourning" in Individuals with Essential Hypertension: An Empirical Study of Elaboration Phases through Systematic Observation]*. Ph.D. thesis, University of Barcelona, Barcelona.

Mashal, N., Solodkin, A., Dick, A. S., Chen, E. E., and Small, S. L. (2012). A network model of observation and imitation if speech. *Front. Psychol.* 3:84. doi: 10.3389/fpsyg.2012.00084

Maxwell, J. A., Chmiel, M., and Rogers, S. E. (2015). "Designing integration in multimethod and mixed methods research," in *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry*, eds S. N. Hesse-Biber, and R. B. *Johnson* (New York, NY: Oxford University Press), 223–239.

McDougall, J. (1991). *Teatros del Cuerpo [Body Theaters]*. Madrid: Julián Yébenes.

McLean, M., Cleland, J. A., Worrell, M., and Vögele, C. (2011). "What am I going to say here?" The experiences of doctors and nurses communicating with patients in a cancer unit. *Front. Psychol.* 2:339. doi: 10.3389/fpsyg.2011.00339

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., and Price, J. H. (2001). The Electronically Activated Recorder (EAR): a device for sampling naturalistic daily activities and conversations. *Behav. Res. Methods Instrum. Comput.* 33, 517–523. doi: 10.3758/BF03195410

Molina-Azorín, J. F., and Cameron, R. (2015). "History and emergent practices of mixed and multiple methods in business research," in *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry,* eds S. Hesse-Biber and R. B. Johnson (New York, NY: Oxford University Press), 466–485.

Morales-Ortiz, M. (1999). "La observación indirecta: una aplicación al análisis de textos [Indirect observation applied to text analysis]," in *Observación en la Escuela: Aplicaciones*, ed M. T. Anguera (Barcelona: Ediciones Universidad de Barcelona), 299–307.

Morales-Sánchez, V., Pérez-López, V., and Anguera, M. T. (2014). Tratamiento metodológico de la observación indirecta en la gestión de organizaciones deportivas [Indirect observational methodology in managing sports services]. *Rev. Psicol. Deport.* 23, 201–207.

Morillo, J. P., Reigal, R. E., Hernández-Mendo, A., Montaña, A., and Morales-Sánchez, V. (2017). Decision-making by handball referees: design of an *ad hoc* observation instrument and polar coordinate analysis. *Front. Psychol.* 8:1842. doi: 10.3389/fpsyg.2017.01842

Mucchielli, R. (1974). *L'observation Psychologique et Psychosociologique [Psychological and Psychosociological Observation]*. Paris: E.S.F.

O'Cathain, A., Murphy, E., and Nicholl, J. (2010). Three techniques for integrating data in mixed methods studies. *BMJ* 341:c4587. doi: 10.1136/bmj.c4587

Onwuegbuzie, A. J. (2003). Effect sizes in qualitative research: a prolegomenon. *Qual. Quant.* 37, 393–409. doi: 10.1023/A:1027379223537

Onwuegbuzie, A. J., and Hitchcock, J. H. (2015). "Advanced mixed analysis approaches," in *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry*, eds S. Hesse-Biber and R. B. Johnson (New York, NY: Oxford University Press), 275–295.

Poitras, J., Hill, K., Hamel, V., and Pelletier, F. B. (2015). Managerial mediation competency: a mixed-method study. *Negot. J.* 31, 105–115. doi: 10.1111/nejo.12085

Pope, C., Ziebland, S., and Mays, N. (2000). Qualitative research in health care. Analysing qualitative data. *BMJ* 320, 114–116. doi: 10.1136/bmj.320. 7227.114

Portell, M., Anguera, M. T., Chacón-Moscoso, S., and Sanduvete-Chaves, S. (2015a). Guidelines for reporting evaluations based on observational methodology. *Psicothema* 27, 283–289. doi: 10.7334/psicothema2014.276

Portell, M., Anguera, M. T., Hernández-Mendo, A., and Jonsson, G. K. (2015b). Quantifying biopsychosocial aspects in everyday contexts: an integrative methodological approach from the behavioral sciences. *Psychol. Res. Behav. Manag.* 8, 153–160. doi: 10.2147/PRBM.S82417

Portell, M., Anguera, M. T., Hernández-Mendo, A., and Jonsson, G. K. (2015c). The legacy of Brunswik's representative design in the 21st century: methodological innovations for studying everyday life. *Brunswik Soc. Newslett.* 30, 38–40.

Poyatos, F. (1993). *Paralanguage. A Linguistic and Interdisciplinary Approach to Interactive Speech and Sound*. Amsterdam: John Benjamins Publishing Company.

Radzikowski, J., Stefanidis, A., Jacobsen, K. H., Croitoru, A., Crooks, A., and Delamater, P. L. (2016). The measles vaccination narrative in Twitter: a quantitative analysis. *JMIR Public Health Surveill.* 2:e1. doi: 10.2196/publichealth.5059

Reis, H. T. (2013). "Why researchers should think "Real-World." A conceptual rationale," in *Handbook of Research Methods for Studying Daily Life*, eds M. R. Mehl and T. S. Conner (New York, NY: The Guilford Press), 3–21.

Riera, A., Ocasio, A., Tiyyagura, G., Krumeich, L., Ragins, K., Thomas, A., et al. (2015). Latino caregiver experiences with asthma health communication. *Qual. Health Res.* 25, 16–26. doi: 10.1177/1049732314549474

Riessman, C. K. (2015). Ruptures and sutures: time, audience and identity in an illness narrative. *Sociol. Health Illn.* 37, 1055–1071. doi: 10.1111/1467-9566.12281

Riva, G., Anguera, M. T., Wiederhold, B. K., and Mantovani, F. (eds.). (2006). *From Communication to Presence. Cognition, Emotions and Culture Towards the Ultimate Communicative Experience*. Amsterdam: IOS Press.

Rodriguez, C. S., Spring, H. J., and Rowe, M. (2014). Nurse's experiences of communicating with hospitalized, suddenly speechless patients. *Qual. Health. Res.* 25, 168–178. doi: 10.1177/1049732314550206

Romero, P., Cox, R., du Boulay, B., Lutz, R., and Bryant, S. (2007). A methodology for the capture and analysis of hybrid data. A case study of program debugging. *Behav. Res. Methods* 39, 309–317. doi: 10.3758/BF03193162

Roustan, M., Izquierdo Rodríguez, C., and Anguera Argilaga, M. T. (2013). Sequential analysis of an interactive peer support group. *Psicothema* 25, 396–401. doi: 10.7334/psicothema2012.93

Russ, J. B., Gur, R. C., and Bilker, W. B. (2008). Validation of affective and neutral sentence content for prosodic testing. *Behav. Res. Methods* 40, 935–939. doi: 10.3758/BRM.40.4.935

Sackett, G. P. (1980). "Lag sequential analysis as a data reduction technique in social interaction research," in *Exceptional Infant. Psychosocial Risks in Infant-Environment Transactions*, eds D. B. Sawin, R. C. Hawkins, L. O. Walker, and J. H. Penticuff (New York, NY: Brunner/Mazel), 300–340.

Sackett, G. P. (1987). "Analysis of sequential social interaction data: some issues, recent developments, and a causal inference model," in *Handbook of Infant Development, 2nd Edn,* ed J. D. Osofsky (New York, NY: Wiley), 855–878.

Saldaña, J. (2013). *The Coding Manual for Qualitative Researchers*. Los Angeles, CA: Sage.

Sánchez-Algarra, P., and Anguera, M. T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: impact of recording and coding among predominating perspectives. *Qual. Quant.* 47, 1237–1257. doi: 10.1007/s11135-012-9764-6

Sandelowski, M. (2001). Real qualitative researchers do not count: the use of numbers in qualitative research. *Res. Nurs. Health* 24, 230–240. doi: 10.1002/nur.1025

Sandelowski, M., Voils, C. I., and Knafl, G. (2009). On quantitizing. *J. Mix. Methods Res.* 3, 208–222. doi: 10.1177/1558689809334210

Santoyo, C., Jonsson, G. K., Anguera, M. T., and López-López, J. A. (2017). Observational analysis of the organization of on-task behavior in the classroom using complementary data analyses. *Anal. Psicol.* 33, 497–514. doi: 10.6018/analesps.33.3.271061

Sarmento, H., Bradley, P. S., Anguera, M. T., Polido, T., Resende, R., and Campaniço, J. (2015). Quantifying the offensive sequences that result in goals in elite futsal matches. *J. Sports Sci.* 34, 621–629. doi: 10.1080/02640414.2015.1066024

Schegloff, E. (2000). On granularity. *Annu. Rev. Sociol.* 26, 715–720. doi: 10.1146/annurev.soc.26.1.715

Seltzer-Kelly, D., Westwood, S. J., and Pena-Guzmán, D. M. (2012). A methodological self-study of quantitizing: negotiating meaning and revealing multiplicity. *J. Mix. Methods Res.* 6, 258–274. doi: 10.1177/1558689811425798

Sidnell, J., and Stivers, T. (eds.). (2013). *The Handbook of Conversation Analysis*. New York, NY: Wiley and Sons.

Suárez, N., Sánchez-López, C. R., Jiménez, J. E., and Anguera, M. T. (2018). Is reading instruction evidence-based? Analyzing teaching practices using T-Patterns. *Front. Psychol.* 9:7. doi: 10.3389/fpsyg.2018.00007

Tarragó, R., Iglesias, X., Lapresa, D., Anguera, M. T., Ruiz-Sanchís, L., and Arana, J. (2017). Analysis of diachronic relationships in successful and unsuccessful behaviors by world fencing champions using three complementary techniques. *Anal. Psicol.* 33, 471–485. doi: 10.6018/analesps.33.3. 271041

Tashakkori, A., and Teddlie, C. (eds.). (2010). *The Sage Handbook of Mixed Methods in Social and Behavioral Research, 2nd Edn.* Thousand Oaks, CA: Sage.

Taylor, K., Thorne, S., and Oliffe, J. L. (2015). It's a sentence, not a word. Insights from a keyword analysis in cancer communication. *Qual. Health Res.* 25, 110–121. doi: 10.1177/1049732314549606

Tronchoni, H., Izquierdo, C., and Anguera, M. T. (2018). *Interacción Participativa en las Clases Magistrales: Fundamentacion y Construcción de un Instrumento de Observación [Participatory Interaction in Lectures: Theoretical Framework and Construction of an Observation Instrument]*. Publicaciones. Facultad de Educación y Humanidades del Campus de Melilla.

Tuttas, C. A. (2015). Lessons learned using web conference technology for online focus group interviews. *Qual. Health Res.* 25, 122–133. doi: 10.1177/1049732314549602

Vaimberg, R. (2010). *Psicoterapias Tecnológicamente Mediadas [Technology-Mediated Psychotherapy]*. Ph.D. thesis, Universidad de Barcelona, Barcelona.

Webb, E. T., Campbell, D. T., Schwartz, R. D., Sechrest, L., and Grove, J. B. (1966). *Nonreactive Measures in the Social Sciences*. Boston, MA: Houghton Mifflin.

Weick, K. E. (1968). "Systematic observational methods," in *Handbook of Social Psychology, Vol. 2, 2nd Edn.*, eds G. Lindzey, and E. Aronson (Reading, MA: Addison-Wesley), 357–451.

Weick, K. E. (1985). "Systematic observational methods," in *Handbook of Social Psychology, Vol. 1, 3rd Edn.*, eds G. Lindzey, and E. Aronson (Reading, MA: Addison-Wesley), 567–634.

Winniccott, D. W. (1979). *Realidad y Juego [Reality and Play]*. Barcelona: Gedisa.

Zaros, A. A. (2016). Retratos de una comunidad religiosa: sobre la memoria y las fotos familiares de la comunidad armenia en Padua [Portraits of a religious community. About the memory and family photos of Padua's Armenian community]. *Rev. Cult. Relig.* 10, 88–106.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Psychology

Preliminary Checklist for Reporting Observational Studies in Sports Areas: Content Validity

Salvador Chacón-Moscoso [1,2]*, Susana Sanduvete-Chaves [1], M. Teresa Anguera [3], José L. Losada [3], Mariona Portell [4] and José A. Lozano-Lozano [1,2]

[1] HUM-649 Innovaciones Metodológicas en Evaluación de Programas, Departamento de Psicología Experimental, Facultad de Psicología, Universidad de Sevilla, Sevilla, Spain, [2] Departamento de Psicología, Universidad Autónoma de Chile, Santiago de Chile, Chile, [3] Faculty of Psychology, Institute of Neurosciences, University of Barcelona, Barcelona, Spain, [4] Department of Psychobiology and Methodology of Health Sciences, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

Observational studies are based on systematic observation, understood as an organized recording and quantification of behavior in its natural context. Applied to the specific area of sports, observational studies present advantages when comparing studies based on other designs, such as the flexibility for adapting to different contexts and the possibility of using non-standardized instruments as well as a high degree of development in specific software and data analysis. Although the importance and usefulness of sports-related observational studies have been widely shown, there is no checklist to report these studies. Consequently, authors do not have a guide to follow in order to include all of the important elements in an observational study in sports areas, and reviewers do not have a reference tool for assessing this type of work. To resolve these issues, this article aims to develop a checklist to measure the quality of sports-related observational studies based on a content validity study. The participants were 22 judges with at least 3 years of experience in observational studies, sports areas, and methodology. They evaluated a list of 60 items systematically selected and classified into 12 dimensions. They were asked to score four aspects of each item on 5-point Likert scales to measure the following dimensions: representativeness, relevance, utility, and feasibility. The judges also had an open-format section for comments. The Osterlind index was calculated for each item and for each of the four aspects. Items were considered appropriate when obtaining a score of at least 0.5 in the four assessed aspects. After considering these inclusion criteria and all of the open-format comments, the resultant checklist consisted of 54 items grouped into the same initial 12 dimensions. Finally, we highlight the strengths of this work. We also present its main limitation: the need to apply the resultant checklist to obtain data and, thus, increase quality indicators of its psychometric properties. For this reason, as relevant actions for further development, we encourage expert readers to use it and provide feedback; we plan to apply it to different sport areas.

Keywords: checklist, reporting, observational studies, sports area, content validity, experts, Osterlind index

INTRODUCTION

Observational studies are mainly based on systematic observation, understood as an organized recording and quantification of behavior in its natural context (Anguera, 1979, 1996, 2003). These types of studies involve a low level of intervention (Chacón-Moscoso et al., 2013). Observational studies present the following important advantages compared to those with a medium or high level of intervention (Portell et al., 2015), such as: (1) they can be adapted to any situation in any setting and (2) they do not need standardized measurement instruments because the context of the study is prioritized and, as a consequence, the use of ad hoc instruments is accepted.

Observational studies are commonly used in many areas, such as social (Anguera and Redondo, 1991; Santoyo and Anguera, 1992), psycho-pedagogical (Moya et al., 2012; Herrero-Nivela et al., 2014), clinical (Roustan et al., 2013; Arias et al., 2015), or sport (Weigelt and Memmert, 2012; Anguera and Hernández-Mendo, 2014) studies. The concrete field of observational methodology, as applied to sports, currently benefits from the advanced development of statistical analyses and specific software to study men and women's sports-related behaviors in order to obtain indicators to improve their performance (Anguera and Hernández-Mendo, 2015; Anguera et al., 2017, 2018). For example, (1) sequential analysis of behaviors using SDIS-GSEQ software (Bakeman and Quera, 2011) has been developed to establish model sequences by high-level sportsmen and women (Castelão et al., 2015); (2) the use of polar coordinate analysis by means of HOISAN software (Hernández-Mendo et al., 2012) enables the study of interrelations between different categories of observational tools in different sports, such as tae kwon do (López-López et al., 2015), handball (Sousa et al., 2015), or soccer (Castellano and Hernández-Mendo, 2003; Castañer et al., 2016); and (3) T-pattern analysis using Theme software (Magnusson, 1996, 2000) can be applied to discover hidden structures in the observed behavior that are not directly visible in elite climbing (Arbulu et al., 2016), futsal (hard-court soccer, Sarmento et al., 2016), synchronized swimming (Iglesias et al., 2015), or bouts epee (fencing, Tarrag et al., 2015).

Although observational studies are frequently used and their utility in different contexts has been widely proven, a tool to measure the reporting quality of these types of studies does not exist, nor does a specific one for sports areas (Portell et al., 2015). This lack causes important consequences for observational studies in sports areas: (1) the author's report is the unique information we usually have about primary studies (Altman et al., 2001; Grimshaw et al., 2006; Cornelius et al., 2009). As authors do not have a checklist for reporting, transparency may be affected, and important information for assessing the quality of the study and, therefore, its degree of risk of bias, may be omitted (Portell et al., 2015). (2) Authors who want to publish these kinds of studies do not have a checklist to confirm that all the important elements were considered in the study and included in the report, and reviewers of these same studies lack a useful tool for determining the indicators to consider when accepting or rejecting their publication in a scientific journal (Chacón-Moscoso et al., 2016).

Checklists to measure the quality of the reporting of primary studies in general, without specifying the design type, have previously been published, e.g., by the Journal Article Reporting Standards (JARS) (American Psychological Association, 2010). In addition, as a consequence of the differences existing across designs, checklists with the same purpose but for specific study designs have been published (Portell et al., 2015; Chacón-Moscoso et al., 2016). For example, for high-intervention designs (randomized control trials), we have the Consolidated Standards of Reporting Trials (CONSORT) (Schulz et al., 2010); for epidemiological studies, such as cohort, case-control, and cross-sectional studies, we have the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement (von Elm et al., 2007); (3) for intensive repeated measurements in naturalistic settings, we have the Guidelines for Reporting Momentary Studies (Stone and Shiffman, 2002); (4) for qualitative studies, we have the Guidelines for Qualitative Research Methodologies (Blignault and Ritchie, 2009); and (5) for mixed methods, we have the Guidelines for Conducting and Reporting Mixed Research for Counselor Researchers (Leech and Onwuegbuzie, 2010). The current standard for low-intervention studies is the Guidelines for Reporting Evaluations Based on Observational Methodology (GREOM) (Portell et al., 2015; included in the EQUATOR library: http://www.equator-network.org/reporting-guidelines/guidelines-for-reporting-evaluations-based-on-observational-methodology/) which, combined with JARS, provides a general view of the structural characteristics of observational designs that must be considered for evaluation in low-intervention situations without specifying any concrete area.

The aim of this work is to develop a checklist to measure the quality of the reporting of sports-related observational studies. This checklist will further clarify the general guidelines presented in the GREOM; determine the main quality indicators of the reporting of sports-related observational studies; serve as a useful tool for authors conducting and publishing observational studies in this area, as well as for reviewers making decisions for publications; and present indicators of the representativeness (REP), relevance (REL), utility (U), and feasibility (F) of the developed tool to measure the quality of the reporting based on a content validity study.

METHODS

Participants

Twenty-two of the 43 potential candidates who were contacted opted to participate in the study, resulting in a participation rate of 51.2%. The inclusion criteria for eligibility were a minimum of 3 years of experience in observational studies, sports areas, and methodology. The sample participants were between 31 and 70 years old [mean (M) = 45.9, standard deviation (SD) = 11.3], including 17 men (77.3%) and 5 women (22.7%). Their years of experience in methodology were between 3 and 44 (M = 17.1, SD = 11.7). Their years of experience in observational studies ranged from 3 to 35 (M = 13.6, SD = 9.3). Finally, their years of experience in sports areas ranged between 3 and 40 (M = 19.9, SD = 10).

The main dedication and sports interest of these experts was in the area of physical education (8 participants, 36.4% of the sample), understood as the set of disciplines that aim to develop the human body through sports participation and encourages psychomotor learning in a game-like setting or through movement exploration. This is a commonly introduced subject in primary and secondary education curricula (Woodward, 2016). The second most frequent area of interest was high-performance sports (five participants, 22.7%), referring to the type of activity (individual and team) in competition contexts (Harenberg et al., 2016). The third area of interest was sport initiation (4 participants, 18.2%), the process by which a person makes contact with new experiences in a physical activity or sport (Thomas et al., 2015). The areas of health, sports, and physical education (a set of educational, sporting, and organizational practices to promote well-being and health; Williams and Macdonald, 2015) and adapter sports (sports practices of people with some kind of physical and/or psychological disability; Park and Sinelnikov, 2016) were chosen by two participants each (9.1%). Finally, one participant (4.5%) chose technology, defined as the tools aimed to improve athletes' sports performance in order to set personal records and, thus, be more competitive (Hardcastle et al., 2015).

Instruments

Appendix I in Supplementary Material presents the instrument we designed to enable content validity experts to determine the main aspects of sports-related observational studies. It is composed of 60 items representing 12 dimensions from the GREOM (Portell et al., 2015): (1) *Extrinsic characteristics* (1 item); (2) *Objectives delimitation* (6 items); (3) *Observational design* (3 items); (4) *Participants* (9 items); (5) *Context-setting* (11 items); (6) *Observational instrument* (7 items); (7) *Recording instrument* (6 items); (8) *Data* (3 items); (9) *Parameter specification* (2 items); (10) *Observational sampling* (6 items); (11) *Data quality control* (5 items); and (12) *Data analysis* (1 item).

For the content validity study, four 5-point Likert scales (Sanduvete-Chaves et al., 2013) were associated with each item to be assessed by the experts referring to four different aspects with respect to its dimension: (1) REP referred to the degree to which each item represented the dimension to which it had been assigned; (2) REL was defined as the extent to which each item was important or highlighted something of the dimension in which it was included; (3) U referred to the extent to which each item was useful to evaluate the dimension to which it was assigned; and (4) F was defined as the possibility of recording information about each item. Additionally, an open-format item (comments) was available for experts who wished to propose something new, such as improving the writing of an item or exchanging it for something more appropriate.

This instrument was available in two formats: the Internet format using Google Drive Forms and a paper version. Microsoft Excel was used for the data analysis.

Procedure

Ethical approval and written informed consent were not needed for our study, as the participants were experts, a non-vulnerable group, and the information gathered was professional opinions about the adequacy of different items used to report observational studies, without medical, clinical, or personal implications.

Item Selection and Assignment to Dimensions

We delimited the main dimensions of observational studies and a list of items to measure those dimensions based on three information sources: (1) A systematic review (Chacón-Moscoso et al., 2016) was conducted of 12 databases that were of interest due to their content (Web of Science, Scopus, Springer, EBSCO Online, Medline, CINAHL, EconLit, MathSciNet, Current Contents, Humanities Index, ERIC, and PsycINFO). We found 548 different ways to measure methodological quality in primary studies. From this total, some of the tools were general reporting standards not specific to any particular research design (e.g., Zaza et al., 2000; American Educational Research Association, 2006; American Psychological Association, 2010; Möhler et al., 2012), while others were specific reporting standards for research designs with some similarities to observational designs (Stone and Shiffman, 2002; Tong et al., 2007; Blignault and Ritchie, 2009; Pluye et al., 2009; Leech and Onwuegbuzie, 2010). (2) The GREOM (Portell et al., 2015) represents the specific guidelines for developing observational studies. As an illustrative example of the GREOM's high degree of influence over the list of items gathered, apart from the common structure, we can see the direct correspondence in dimensions of the present checklist 6 *Observational instrument,* 7 *Recording instrument,* and 9 *Parameters specification,* according to section B2 *Instruments* of GREOM, including the guidelines 7 *Observation instrument*, 8 *Primary recording parameters,* and 9 *Recording instruments.* (3) The final dimension involves sports-related observational studies found in the previously cited databases (Anguera and Hernández-Mendo, 2015).

Two coders separately studied the degree of agreement in the items dimension assigned and intercoder reliability (Nimon et al., 2012; Stolarova et al., 2014) was studied by calculating Cohen's κ (Cohen, 1960). Any disagreements were resolved by consensus.

Content Validity Study

Once the 60 items were selected and assigned to one of the 12 dimensions chosen, the experts were asked, through 5-point rating scales, about the REP, REL, U, and F of each item with respect to its dimension.

The instrument was sent to experts in English (Appendix I in Supplementary Material) or Spanish (Supplementary Material), depending on their native language. We provided the access link to the instrument in Google Drive by sending an email to the potential experts that satisfied the participant's inclusion criteria. Fifteen days later, we reminded the participants that the instrument was available to be fulfilled in the same link. After another 15 days, we made the last call for answers in the same way. After a final 15 days, the application was definitively closed. As part of the final gathering stage, the same instrument was available in paper format for all of the participants (who worked in observational studies, methodology, and sports areas) at the VII European Congress of Methodology, held in Palma de Mallorca (Spain) in July 2016. Throughout the entire process, the information was gathered anonymously.

After gathering the information, the Osterlind index of congruence (Osterlind, 1998) was calculated for each item and each aspect measured (REP, REL, U, and F). The formula used was.

$$I_{ik} = \frac{(N-1)\sum\limits_{j=1}^{n} X_{ijk} + N \sum\limits_{j=1}^{n} X_{ijk} - \sum\limits_{j=1}^{n} X_{ijk}}{2(N-1)n}$$

where N = number of dimensions (12 in this case), X_{ijk} = score provided by each expert to each item referred to each aspect measured, and n = number of experts. The scores were provided in a 5-point Likert scale (-1 = strongly disagree, -0.5 = disagree, 0 = neither agree nor disagree, 0.5 = agree, and 1 = strongly agree) instead of the classical one with 3 points, to make the achievement of high results slightly difficult, as the 5-point version is more conservative (Revised Osterlind Index, Sanduvete-Chaves et al., 2013).

The results of the previous formula ranged from -1 to $+1$. Minus one implied a total agreement among the experts, indicating that all answered that they disagreed strongly; 1 meant a total agreement among the experts, positioning all in strong agreement; and 0 represented the highest possible disagreement among the experts.

Based on the criteria (Osterlind, 1998), items that obtained a score of 0.5 or higher on the four aspects measured were included in the final version of the checklist for reporting observational studies.

RESULTS

The assignment of the 60 items selected to the 12 dimensions made by two independent researchers obtained a degree of consensus of $\kappa = 0.76$ ($p < 0.001$) and a 95% confidence interval (CI) of [0.646, 0.874]. This result can be considered appropriate (Landis and Koch, 1977).

Forty-three experts were contacted by email to fulfill the content validity questionnaire. A total of 14 experts answered via Google Drive. Two participants sent their responses after the first call for answers, five participants answered in the second round, and seven additional experts gave their opinions in the final round. Additionally, eight experts fulfilled the questionnaire in paper format during the VII European Congress of Methodology (July 2016). The total number of answers gathered was 22. According to Prieto and Muñiz (2000), a number of experts ranging from 10 to 30 through a systematic procedure can be considered a moderate sample size.

Table 1 presents the Osterlind indexes obtained for each item referring to REP, REL, U, and F. Fifty-three items met the criterion of having a result of 0.5 or higher in these four aspects. Only seven items were removed because they did not meet this criterion: those in dimension 4 referred to the participants, items 12 (cultural background), 13 (socio-economic level), 17 (differential exclusion of participants), and 18 (participants' allocation); those in dimension 5 referred to the context (setting), item 24 (number of non-observable periods); those in dimension 6 referred to the observational instrument, item 34 (criteria

that lead to the catalogs and categories systems); and those in dimension 9 referred to the parameters specification, item 48 (parameters fitting). The removed items appear in bold text in Table 1.

Analyzing all of the items as a whole and taking into account that the possible results ranged from -1 to 1, we found that, in REP, $Mdn = 0.71$ ($SD = 0.14$), with the values ranging from 0.3 to 0.98; in REL, $Mdn = 0.69$, $SD = 0.16$, range = 0.21–1; in U, $Mdn = 0.69$, $SD = 0.15$, range = 0.33–0.98; finally, in F, $Mdn = 0.73$, $SD = 0.14$, range = 0.28–0.93.

Table 2 presents the open-format comments made by the experts and the actions developed in order to follow their advice. From a total of 22 different comments, all were followed with the exception of one (item 11), to which we made only a partial change. Four comments did not imply changes because they referred to items excluded by the Osterlind index results.

All of the comments presented were provided by only one expert, except those referring to the graduation of the answers for some dichotomous items, which were proposed by five experts.

Appendix II in Supplementary Material presents the final version of the checklist for reporting sports-related observational studies after making the changes derived from the results of the Osterlind indexes and the experts' open-format comments. One proposal provided in open format was to add one more item. Originally, the instrument presented 60 items, and 7 were removed due to the Osterlind indexes, resulting in the inclusion of 54 items in the final version.

DISCUSSION

We propose a 54-item and 12-dimension checklist to measure the reporting quality of observational studies in sports areas. Its use by authors and reviewers may contribute to the increased transparency of these studies, as it lists the main aspects to consider and delimit when designing, executing, or evaluating observational studies in sports areas. The importance of this checklist resides in its exclusivity, considering that no other tool with this same purpose exists in the literature. There are other checklists available with the same objective as our proposal, that is, to measure the quality of reporting, although to be applied in other contexts (e.g., in orthopedics, Mundi et al., 2008) and other kinds of designs (e.g., in orthopedics and randomized control trials, Chan and Bhandari, 2007). Additionally, there are checklists in sports (e.g., Arnold and Schilling, 2017), but in designs different from observational studies (Anguera et al., 2018; as guidelines created for this methodology, readers can see the GREOM included in the EQUATOR library: http://www.equator-network.org/reporting-guidelines/guidelines-for-reporting-evaluations-based-on-observational-methodology/). On other occasions, we find checklists applied to similar designs (e.g., STROBE for epidemiological studies, von Elm et al., 2007), although not exactly for observational studies understood as an organized recording and quantification of behavior in its natural context.

These checklists present some characteristics in common with our proposal, such as the format (closed-ended questions) or

TABLE 1 | Osterlind indexes obtained for each item in representativeness (REP), relevance (REL), utility (U), and feasibility (F).

Item[a]	REP	REL	U	F
DIMENSION 1. EXTRINSIC CHARACTERISTIC				
1. Publication type	0.90	0.93	0.94	0.93
DIMENSION 2. OBJECTIVES DELIMITATION				
2. Problem delimitation, sport	0.77	0.93	0.84	0.85
3. Problem delimitation, general objectives	0.93	1	0.95	0.93
4. Problem delimitation, specific objectives	0.93	0.90	0.88	0.85
5. Reference to theoretical framework	0.71	0.81	0.79	0.74
6. Specification of response levels	0.82	0.83	0.80	0.78
7. Specification of participation degree	0.70	0.76	0.70	0.73
DIMENSION 3. OBSERVATIONAL DESIGN				
8. Specification of observational design for each objective	0.86	0.83	0.93	0.90
9. Justification of the observational design	0.81	0.79	0.73	0.73
10. Sequence data are obtained	0.50	0.79	0.70	0.60
DIMENSION 4. PARTICIPANTS				
11. Age	0.50	0.58	0.66	0.66
12. Cultural background	**0.34**	**0.21**	**0.33**	**0.33**
13. Socio-economic level	**0.30**	**0.36**	**0.33**	**0.28**
14. Sport modality	0.55	0.64	0.65	0.80
15. Professionalism	0.60	0.62	0.60	0.58
16. Global exclusion of participants	0.66	0.68	0.58	0.60
17. Differential exclusion of participants	**0.57**	**0.40**	**0.42**	0.52
18. Participants' allocation	**0.57**	**0.50**	**0.42**	**0.40**
19. Activity type	0.68	0.62	0.68	0.75
DIMENSION 5. CONTEXT (SETTING)				
20. Place (location)	0.61	0.62	0.63	0.83
21. Social impact of the activity	0.59	0.52	0.53	0.63
22. Time frame	0.66	0.67	0.73	0.75
23. Session acceptance criteria	0.68	0.64	0.70	0.65
24. Number of non-observable periods	0.66	0.55	0.55	**0.45**
25. Duration of non-observable periods	0.73	0.71	0.75	0.73
26. Total results indication	0.73	0.69	0.65	0.68
27. Partial results indication	0.59	0.69	0.65	0.63
28. Observational unit adjustment	0.61	0.69	0.65	0.60
29. Observational units delimiting, denominating and definable	0.70	0.71	0.73	0.68
30. Global/molecular units' degree justified	0.91	0.95	0.93	0.88
DIMENSION 6. OBSERVATIONAL INSTRUMENT				
31. Type of observational instrument	0.86	0.79	0.82	0.73
32. Instrument appropriate to the design	0.80	0.74	0.68	0.75
33. Justification of instrument type according to the observational design	0.66	0.69	0.65	0.68
34. Criteria that led to the catalogs and categories systems	0.64	**0.43**	0.55	0.50
35. Requirements to categorize from a certain criterion	0.64	0.60	0.55	0.63

(Continued)

TABLE 1 | Continued

Item[a]	REP	REL	U	F
36. Availability of full coding manual	0.80	0.79	0.68	0.75
37. Observational instrument adequate to the study context	0.75	0.74	0.75	0.78
DIMENSION 7. RECORDING INSTRUMENT				
38. Software utilization as user	0.70	0.68	0.70	0.65
DIMENSION 8. DATA				
39. Software type used to record	0.68	0.64	0.63	0.70
40. Observational recording	0.80	0.81	0.75	0.80
41. Software used to record	0.91	0.93	0.98	0.93
42. Software used for data quality control	0.91	0.90	0.93	0.90
43. Software used for data analysis	0.82	0.83	0.85	0.83
44. Type of data according to Bakeman (1978)	0.75	0.62	0.68	0.75
45. Type of data according to Bakeman (1983)	0.70	0.60	0.55	0.65
46. Data management	0.66	0.60	0.53	0.53
DIMENSION 9. PARAMETER SPECIFICATION				
47. Parameter type (the most complex)	0.64	0.52	0.55	0.50
48. Parameter fitting (the most complex)	**0.48**	**0.48**	**0.45**	0.53
DIMENSION 10. OBSERVATIONAL SAMPLING				
49. Observational period	0.75	0.57	0.68	0.74
50. Sessions periodicity	0.75	0.60	0.70	0.73
51. Number of sessions	0.86	0.71	0.78	0.74
52. Starting session criterion	0.75	0.76	0.68	0.68
53. Ending session criterion	0.80	0.74	0.73	0.78
54. Within-session sampling	0.70	0.57	0.65	0.58
DIMENSION 11. DATA QUALITY CONTROL				
55. Agreements	0.67	0.60	0.58	0.50
56. Concordance	0.93	0.93	0.90	0.85
57. Within-session reliability	0.86	0.79	0.80	0.75
58. Between-session reliability	0.84	0.76	0.78	0.75
59. Generalizability theory application	0.76	0.70	0.71	0.66
DIMENSION 12. DATA ANALYSIS				
60. Data analysis developed	0.98	0.90	0.89	0.89

REP, representativeness; REL, relevance; U, utility; F, feasibility. One item is considered appropriate when the values obtained in the four aspects measured (REP, REL, U, and F) are at least 0.5. We marked the Osterlind indexes under 0.5 and removed the items in bold text.
[a]Items appear in abbreviated form; the full items can be consulted in Appendix I (Spanish version in Supplementary Material).

the capacity to detect relevant information that has not been reported. Nevertheless, they differ in content, not only due to the sport context [e.g., item 14, *Sport modality: (1) Individual sport, (2) Team sport*; or item 15, *Professionalism: (1) Professionals, (2) Semi-professionals, (3) Sportsmen/women in training stage*], but also due to the casuistic of the observational design [e.g., item 9, *Justification of the observational design: (1) No, (2) Yes*; or item 24, *Number of non-observable periods*].

The main strength of this work is that the content validity study was developed through a clear, careful, and explicit process, so it presented a high degree of reproducibility. In this way, we were able to define a list of items based on different sources of information: a systematic review, the GREOM as

the theoretical framework and the basis for the 12 delimited dimensions and content in several dimensions (illustrated in the correspondence between dimensions 6 *Observational instrument,* 7 *Recording instrument,* and 9 *Parameters specification* of the presented checklist and guidelines 7 *Observation instrument,* 8 *Primary recording parameters,* and 9 *Recording instruments,* corresponding to section B2 *Instruments* of the GREOM) (Portell et al., 2015), and published observational studies in sports areas. We provided the full list of items assessed by the experts in English (Appendix I in Supplementary Material) and Spanish (Supplementary Material). We determined the inclusion criterion a priori; we reported the Osterlind index for all of the items in the four aspects measured (**Table 1**). We

TABLE 2 | Open-format comments provided by experts and actions taken as a consequence.

Item[a]	Comment	Action[b]
DIMENSION 2. OBJECTIVES DELIMITATION		
2. Problem delimitation, sport	Remove the alternative *psychomotor skills* and add *modified/reduced sport games*	Done
3. Problem delimitation, general objectives	Introduce one more option between *no* and *yes*	We introduced the intermediate option *yes, but unclearly*
4. Problem delimitation, specific objectives	Introduce one more option between *no* and *yes*	We introduced the intermediate option *yes, but unclearly*
5. Reference to theoretical framework	Remove the alternative *Regulation*	Done
6. Specification of response levels	Introduce one more option between *no* and *yes*	We introduced the intermediate option *yes, but unclearly*
DIMENSION 3. OBSERVATIONAL DESIGN		
9. Justification of the observational design	Introduce one more option between *no* and *yes*: *partial/incomplete*	Done
DIMENSION 4. PARTICIPANTS		
11. Age	Establish 10-year intervals from 19s	Based on basketball classification, we included one more category (*under 20s*). We did not include more intervals because, in the absolute category (senior), there is usually no age restriction
12. Cultural background	Introduce one more option between *no* and *yes*	Removed due to Osterlind index results
13. Socio-economic level	Introduce one more option between each pair of alternatives	Removed due to Osterlind index results
14. Sport modality	Related to this, add a new item: opposition (yes/no)	Done
15. Professionalism	Include the options *amateur* and *retired from professional sport*	Done
16. Global exclusion of participants	Include it in dimension 11, Data quality control	Done
17. Differential exclusion of participants	Include it in dimension 11, Data quality control	Removed due to Osterlind index results
18. Participants' allocation	It is not applicable in most observational studies. Include it in dimension 11, Data quality control	Removed due to Osterlind index results
19. Activity type	Include *physical exercise*	Done
DIMENSION 5. CONTEXT (SETTING		
21. Social impact of the activity	Change the options to low/medium/high	Done
28. Observational unit adjustment	Add the option *according to the player*	Done
DIMENSION 6. OBSERVATIONAL INSTRUMENT		
37. Observational instrument adequate to the study context	Include an intermediate option: *partial*	Done
DIMENSION 7. RECORDING INSTRUMENT		
41. Software used to record	Add *SportsCode*	Done
42. Software used for data quality control	Include this item in dimension 11 *Data quality control*	Done
43. Software used for data analysis	Include this item in dimension 12 *Data analysis*	Done

[a]*Items appear in abbreviated form; the full items can be consulted in Appendix I (Spanish version in Supplementary Material). Only the items that received some comment have been included in* **Table 2**.
[b]*Changes resulting from these comments are presented in the final version of the checklist (Appendix II).*

objectively applied the previously established inclusion criterion and transcribed all of the open-format comments provided by the experts and each action we executed in answer to each comment (**Table 2**). After considering the Osterlind indexes and open-format comments, we presented the final version of the checklist for reporting sports-related observational studies (Appendix II in Supplementary Material).

Additionally, we obtained adequate results for the fitness item dimension with respect to four aspects: REP, REL, U, and F, which provides a quality indicator of the content validity in favor of the use of the resulting tool as appropriate. The resultant checklist is expected to be extensively useful, as it can be applied to any sports area.

On the other hand, the main limitation we found in the checklist obtained is that it supposes a preliminary proposal in which further development is needed to increase the quality indicators of its psychometric properties. For this purpose, we encourage and urge expert readers to improve our final version checklist (Appendix II in Supplementary Material) with their comments or results regarding its application.

Additionally, we plan to apply the checklist to different sports areas in order to demonstrate that it is an adequate measurement instrument independent of the sport context and to develop an intercoder reliability study to locate discrepancies across the independent coding of a high number of studies (more than 40) by two different previously trained

coders. We consider this proposal as open and in progress, as we will continue to consider additional comments for the improvement of the checklist that we might receive by experts.

Taking this work as the basis, we plan to develop a scale to measure methodological quality in sports-related observational studies. This checklist can serve as a guideline for measuring the reporting quality of these studies because it lists the main aspects to consider when designing, executing, and evaluating a sports-related observational study. We can also recommend concrete actions to increase the methodological quality of these studies.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Declaration on bioethics and human rights, UNESCO, 2005 with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Ethics Committee, Universidad Autónoma de Chile.

AUTHOR CONTRIBUTIONS

All of the authors contributed to documenting, designing, drafting, and writing the manuscript, and revised it for important theoretical and intellectual content. Additionally, all of the authors provided final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00291/full#supplementary-material

REFERENCES

Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., et al. (2001). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann. Intern. Med.* 134, 663–694. doi: 10.7326/0003-4819-134-8-200104170-00012

American Educational Research Association (2006). Standards for reporting on empirical social science research in AERA publications. *Educ. Res. 35*, 33–40. doi: 10.3102/0013189X035006033

American Psychological Association (2010). *Publication Manual of the American Psychological Association, 6th Edn.* Washington, DC: American Psychological Association.

Anguera, M. T. (1979). Observational typology. *Qual. Quant.* 13, 449–484.

Anguera, M. T. (1996). Introduction. *Eur. J. Psychol. Assess.* 12, 87–88. doi: 10.1027/1015-5759.12.2.87

Anguera, M. T. (2003). "Observational methods (General)," in *Encyclopedia of Behavioral Assessment*, Vol. 2, ed R. Fernández-Ballesteros (London: Sage), 632–637.

Anguera, M. T., and Hernández-Mendo, A. (2014). Metodología observacional y psicología del deporte: estado de la cuestión [Observational methodology and sport psychology: state of the art]. *Rev. Psicol. Dep.* 23, 103–109. Available online at: http://www.rpd-online.com/article/view/v23-n1-anguera-hernandez-mendo/pdf_es

Anguera, M. T., and Hernández-Mendo, A. (2015). Técnicas de análisis en estudios observacionales en ciencias del deporte [Analyses techniques in observartional studies in sport science]. *Cuadern. Psicol. Dep.* 15, 13–30. doi: 10.4321/S1578-84232015000100002

Anguera, M. T., and Redondo, S. (1991). La evaluación de la intervención penitenciaria [The evaluation of the penitentiary intervention]. *Delincuen. Delinq.* 3, 245–289.

Anguera, M. T., Camerino, O., Castañer, M., Sánchez-Algarra, P., and Onwuegbuzie, A. J. (2017). The specificity of observational studies in physical activity and sports sciences: Moving forward in mixed methods research and proposals for achieving quantitative and qualitative symmetry. *Front. Psychol.* 8:2196. doi: 10.3389/fpsyg.2017.02196

Anguera, M. T., Portell, M., Chacón-Moscoso, S., and Sanduvete-Chaves, S. (2018). Indirect observation in everyday contexts: concepts and methodological guidelines within a mixed methods framework. *Front. Psychol.* 9:13. doi: 10.3389/fpsyg.2018.00013

Arbulu, A., Lapresa, D., Usabiaga, O., and Castellano, J. (2016). Detección y aplicación de T-patterns en la escalada de élite [Detection and application of T-patterns in elite climbing]. *Cuadern. Psicol. Dep.* 16, 95–102. Available online at: http://hdl.handle.net/10201/48609

Arias, E., Fieschi, E., Miralbell, J., Castellò, C., Soldevila, A., Anguera, M. T., et al. (2015). Uso de la imitación en psicoterapia psicoanalítica con un niño con Trastorno del Espectro Autista grave: efectos en la interacción social recíproca [Use of imitation in psychoanalytic psychotherapy with a child with severe spectrum disorder autism: effects on reciprocal social interaction]. *Rev. Psicopatol. Salud Mental Niño Adolesc.* 25, 9–20.

Arnold, B. L., and Schilling, B. K. (2017). *Evidence-Based Practice in Sport and Exercise: A Guide to Using Research.* Philadelphia, PA: Davis Company.

Bakeman, R. (1978). "Untangling streams of behavior: sequential analysis of observation data," in *Observing Behavior, Data Collection and Analysis Methods*, Vol. 2, ed G. P. Sackett (Baltimore, MD: University of Park Press), 63–78.

Bakeman, R. (1983). Computing lag sequential analysis statistics: the ELAG program. *Behav. Res. Methods Instrum.* 15, 530–535. doi: 10.3758/BF03203700

Bakeman, R., and Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge: Cambridge University Press.

Blignault, I., and Ritchie, J. (2009). Revealing the wood and the trees: reporting qualitative research. *Health Promot. J. Austr.* 20, 140–145. doi: 10.1071/HE09140

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Canton, A., and Hileno, R. (2016). Goal scoring in soccer: a polar coordinate analysis of motor skills used by Lionel Messi. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Castelão, D. P., Garganta, J., Afonso, J., and da Costa, I. T. (2015). Análise seqüencial de comportamentos ofensivos desempenhados por seleções nacionais de futebol de alto rendimento [sequential analysis of attack behaviors carried out by high-performance national teams in soccer]. *Rev. Brasil. Ciênc. Esp.* 37, 230–236. doi: 10.1016/j.rbce.2015.05.001

Castellano, J., and Hernández-Mendo, A. (2003). El análisis de coordenadas polares para la estimación de relaciones en la interacción motriz en fútbol [The analysis of polar coordenates to estimate the relationships in the motrix interaction in soccer]. *Psicothema* 15, 569–574. Available online at: http://www.psicothema.com/pdf/1109.pdf

Chacón-Moscoso, S., Sanduvete-Chaves, S., and Sánchez-Martín, M. (2016). The development of a checklist to enhance methodological quality in intervention programs. *Front. Psychol.* 7:1811. doi: 10.3389/fpsyg.2016.01811

Chacón-Moscoso, S., Sanduvete-Chaves, S., Portell, M., and and Anguera, M. T. (2013). Reporting a program evaluation: needs, program plan, intervention, and decisions. *Int. J. Clin. Health Psychol.* 13, 58–66. doi: 10.1016/S1697-2600(13)70008-5

Chan, S., and Bhandari, M. (2007). The quality of reporting of orthopaedic randomized trials with use of a checklist for nonpharmacological therapies. *J. Bone Joint Surg. Am.* 89, 1970–1978. doi: 10.2106/jbjs.f.01591

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46.

Cornelius, V. R., Perrio, M. J., Shakir, S. A. W., and Smith, L. A. (2009). Systematic reviews of adverse effects of drug interventions: a survey of their conduct and reporting quality. *Pharmacoepidemiol. Drug Saf.* 18, 1223–1231. doi: 10.1002/pds.1844

Grimshaw, J., Eccles, M., Thomas, R., MacLennan, G., Ramsay, C. R., Fraser, C., et al. (2006). Toward evidence-based quality improvement. evidence (and its limitations) of the effectiveness of guideline dissemination and implementation strategies 1966-1998. *J. Gen. Internal Med.* 21(Suppl. 2), 14–20. doi: 10.1111/j.1525-1497.2006.00357.x

Hardcastle, S. J., Tye, M., Glassey, R., and Hagger, M. S. (2015). Exploring the perceived effectiveness of a life skills development program for high-performance athletes. *Psychol. Sport Exerc.* 16, 139–149. doi: 10.1016/j.psychsport.2014.10.005

Harenberg, S., Riemer, H. A., Jarraman, E., and Dorsch, K. D. (2016). Coaches' perspectives of intrateam competition in high performance sport teams. *Int. Sport Coach. J.* 3, 156–169. doi: 10.1123/iscj.2016-0056

Hernández-Mendo, A., López-López, J. A., Castellano, J., Morales-Sánchez, V., and Pastrana, J. L. (2012). Hoisan 1.2: programa informático para uso en metodología observacional [Hoisan 1.2: software for observational methodology]. *Cuadern. Psicol. Dep.* 12, 55–78. doi: 10.4321/S1578-84232012000100006

Herrero-Nivela, M. L., Escolano-Pérez, E., Anguera, M. T., and Acero-Ferrero, M. (2014). Evaluación mediante observación de la motricidad en niños de Educación Infantil [evaluation by means of observation of motor skills in children in early childhood education]. *Int. J. Dev. Educ. Psychol.* 5, 469–476. doi: 10.17060/ijodaep.2014.n1.v5.708

Iglesias, X., Rodríguez-Zamora, L., Chaverri, D., Clapés, P., Rodríguez, F. A., and Anguera, M. T. (2015). Diversificación de patrones en rutinas de solo en natación sincronizada de alto nivel [Patterns of solo routines in high-performance synchronized swimming]. *Cuadern. Psicol. Dep.* 15, 89–98. doi: 10.4321/S1578-84232015000100009

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310

Leech, N. L., and Onwuegbuzie, A. J. (2010). Guidelines for conducting and reporting mixed research in the field of counseling and beyond. *J. Counsel. Dev.* 88, 61–69. doi: 10.1002/j.1556-6678.2010.tb00151.x

López-López, J. A., Menescardi, C., Estevan, I., Falcó, C., and Hernández-Mendo, A. (2015). Technical-tactical analysis in Taekwondo with polar coordinates through software HOISAN. *Cuadern. Psicol. Dep.* 15, 131–142. doi: 10.4321/S1578-84232015000100013

Magnusson, M. S. (1996). Hidden real-time patterns in intra- and inter-individual behavior. *Eur. J. Psychol. Assess.* 12, 112–123. doi: 10.1027/1015-5759.12.2.112

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behav. Res. Methods Instrum. Comp.* 32, 93–110. doi: 10.3758/BF03200792

Möhler, R., Bartoszek, G., Köpke, S., and Meyer, G. (2012). Proposed criteria for reporting the development and evaluation of complex interventions in healthcare (CReDECI): guideline development. *Int. J. Nurs. Stud.* 49, 40–46. doi: 10.1016/j.ijnurstu.2011.08.003

Moya, J., de Armas, M., and Anguera, M. T. (2012). Els problemes de comportament en els nens i adolescents de Catalunya: estudi descriptiu del trastorn per dèficit d'atenció / trastorn de conducta. Una reflexió epistemológica [The behavioral problems in children and adolescents in Catalonia: a descriptive study of attention deficit disorder / behavior disorder. an epistemological approach]. *Àmb. Psicopedag.* 35, 8–15. Available online at: http://www.psiara.cat/view_article.asp?id=4218

Mundi, R., Chaudhry, H., Singh, I., and Bhandari, M. (2008). Checklists to improve the quality of the orthopaedic literature. *Indian J. Orthop.* 42, 150–164. doi: 10.4103/0019-5413.40251

Nimon, K., Zientek, L. R., and Henson, R. K. (2012). The assumption of a reliable instrument and other pitfalls to avoid when considering the reliability of data. *Front. Psychol.* 3:102. doi: 10.3389/fpsyg.2012.00102

Osterlind, S. J. (1998). *Constructing Tests Items*. Boston, MA: Kluwer Academic Publishers.

Park, C. W., and Sinelnikov, O. A. (2016). Changing college students' perception on disability through adapted sports. *Res. Q. Exerc. Sport* 87, 1. doi: 10.1080/02701367.2016.1217725

Pluye, P., Gagnon, M. P., Griffiths, F., and Johnson-Lafleur, J. (2009). A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *Int. J. Nurs. Stud.* 46, 529–546. doi: 10.1016/j.ijnurstu.2009.01.009

Portell, M., Anguera, M. T., Chacón-Moscoso, S., and Sanduvete-Chaves, S. (2015). Guidelines for reporting evaluations based on observational methodology. *Psicothema* 27, 283–289. doi: 10.7334/psicothema2014.276

Prieto, G., and Muñiz, J. (2000). Un modelo para evaluar la calidad de los test utilizados en España. *Pap. Psicól.* 77, 65–75. Available online at: http://www.redalyc.org/articulo.oa?id=77807709

Roustan, M., Izquierdo, C., and Anguera, M. T. (2013). Sequential analysis of an interactive peer support group. *Psicothema* 25, 396–401. doi: 10.7334/psicothema2012.93

Sánchez-Meca, J. (1997). "Methodological issues in the meta-evaluation of correctional treatment," in *Advances in Psychology and Law. International Contributions*, eds S. Redondo, V. Garrido, J. Pérez, and R. Barberet (New York, NY: Walter de Gruyter), 486–498.

Sánchez-Meca, J., and Ato, M. (1989). "Meta-análisis: una alternativa metodológica a las revisiones tradicionales de la investigación [Meta-analysis: a methodological alternative to traditional research reviews]," in *Tratado de Psicología General*, eds J. Arnau and H. Carpintero (Madrid: Alhambra), 617–669.

Sanduvete-Chaves, S., Chacón-Moscoso, S., Sánchez-Martín, M., and Pérez-Gil, J. A. (2013). The revised osterlind index. a comparative analysis in content validity studies. *Acc. Psicol.* 10, 19–26. doi: 10.5944/ap.10.2.11821

Santoyo, C., and Anguera, M. T. (1992). El hacinamiento como contexto: estrategias metodológicas para su análisis [Crowding as context: analytical and methodological strategies]. *Psicothema* 4, 551–569.

Sarmento, H., Bradley, P., Anguera, M. T., Polido, T., Resende, R., and Campaniço, J. (2016). Quantifying the offensive sequences that result in goals in elite futsal matches. *J. Sports Sci.* 34, 621–629. doi: 10.1080/02640414.2015.1066024

Schulz, K. F., Altman, D. G., and Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ* 340, 698–702. doi: 10.1136/bmj.c332

Sousa, D. J., Nunes-Prudente, J., Sequeira, P., López-López, J. A., and Hernández-Mendo, A. (2015). Analysis of 2vs. 2 situations in men's European handball championship 2012: application of polar coordinates technique. *Cuadern.*

Psicol. Dep. 15, 181–193. Available online at: http://revistas.um.es/cpd/article/view/223391/173611

Stolarova, M., Wolf, C., Rinker, T., and Brielmann, A. (2014). How to assess and compare inter-rater reliability, agreement and correlation of ratings: an exemplary analysis of mother–father and parent–teacher expressive vocabulary rating pairs. *Front. Psychol.* 5:509. doi: 10.3389/fpsyg.2014.00509

Stone, A. A., and Shiffman, S. (2002). Capturing momentary, self-report data: a proposal for reporting guidelines. *Ann. Behav. Med.* 24, 236–243. doi: 10.1207/S15324796ABM2403_09

Tarrag,ó, R., Iglesias, X., Michavila, J. J., Chaverri, D., Ruiz-Sanchis, L., and Anguera, M. T. (2015). Análisis de patrones en asaltos de espada de alto nivel [Patterns analysis in high-performance bouts epee]. *Cuadern. Psicol. Dep.* 15, 151–160. doi: 10.4321/S1578-84232015000100015

Thomas, J. R., Silverman, S., and Nelson, J. (2015). *Research Methods in Physical Activity*. Champaign. IL: Human Kinetics.

Tong, A., Sainsbury, P., and Craig, J. (2007). COnsolidated criteria for REporting Qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *J. Qual. Health Care* 19, 349–357. doi: 10.1093/intqhc/mzm042

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., and Vandenbroucke, J. P. (2007). The strengthening the reporting of observational studies in epidemiology (STROBE) statement. guidelines for reporting observational studies. *Epidemiology* 18, 800–804. doi: 10.1136/bmj.39335.541782.AD

Weigelt, M., and Memmert, D. (2012). Goal–side selection in soccer penalty kicking when viewing natural scenes. *Front. Psychol.* 3:312. doi: 10.3389/fpsyg.2012.00312

Williams, B. J., and Macdonald, D. (2015). Explaining outsourcing in health, sport and physical education. *Sport Educ. Soc.* 20, 57–72. doi: 10.1080/13573322.2014.914902

Woodward, K. (2016). *Embodied Sporting Practices: Regulating and Regulatory Bodies*. London: Springer.

Zaza, S., Wright-De Agüero, L. K., Briss, P. A., Truman, B. I., Hopkins, D. P., Hennessy, M. H., et al. (2000). Data collection instrument and procedure for systematic reviews in the guide to community preventive services. *Am. J. Prev. Med.* 18, 44–74. doi: 10.1016/S0749-3797(99)00122-1

frontiers
in Psychology

Simulation Theory Applied to Direct Systematic Observation

Rumen Manolov* and José L. Losada

Department of Social Psychology and Quantitative Psychology, Faculty of Psychology, University of Barcelona, Barcelona, Spain

Observational studies entail making several decisions before data collection, such as the observational design to use, the sampling of sessions within the observational period, the need for time sampling within the observation sessions, as well as the observation recording procedures to use. The focus of the present article is on observational recording procedures different from continuous recording (i.e., momentary time sampling, partial and whole interval recording). The main aim is to develop an online software application, constructed using R and the Shiny package, on the basis of simulations using the alternating renewal process (a model implemented in the *ARPobservation* package). The application offers graphical representations that can be useful to both university students constructing knowledge on Observational Methodology and to applied researchers planning to use discontinuous recording in their studies, because it helps identifying the conditions (e.g., interval length, average duration of the behavior of interest) in which the prevalence of the target behavior is expected to be estimated with less bias or no bias and with more efficiency. The estimation of frequency is another topic covered.

Keywords: direct observation, time sampling, alternating renewal process, prevalence, interval recording

INTRODUCTION

Observation as a means of gathering data has been and is still present across disciplines and contexts related to psychological processes, including clinical psychology (Langer et al., 2016), work-related behaviors (Beck et al., 2016), family interactions (Dishion et al., 2016), social competence in childhood (Vaughn et al., 2016), sports (Castañer et al., 2016), primatology (Dolado et al., 2016), and ethology in general (Pasquaretta et al., 2016). Observation is also the most frequently used means for gathering data in single-case designs in which the behavior of individuals usually takes place in free-operant contexts (Pustejovsky, 2015). In the present text, the focus is put on *direct* observation, which is considered direct in two senses (Fassnacht, 1982): there is nothing between observer and observed (e.g., no interview or questionnaire is used) and records are compiled immediately after the observation session. In that sense, Ayres and Gast (2010) distinguish direct observation from automated-quantitative recording (that does not require human observers) and direct measurement of permanent products (such as exams or reports elaborated by the participants).

In the following sections we present an example of an observational study, in the context of which we illustrate the decisions that need to be made when conducting such an investigation: (a) choose observational designs; (b) choose what to code; (c) decide whether time sampling is required; (d) choose an observational recording procedure. Afterward, we focus on the latter point;

specifically, we describe the method used for performing the simulations for studying how well prevalence and frequency of the target behavior are estimated in different observational recording procedures. We comment on the way in which the results of the simulations are implemented into interactive graphs, how these graphs can be used and what their main results are.

An Example

In an observational study, the aim is to focus on spontaneous behavior taking place in the natural environment of this target behavior and without modifications being introduced by the researcher. Specifically, the context of the example is Attention Deficit Hyperactivity Disorder (ADHD), due to its relatively high and maintained prevalence across countries and decades (Polanczyk et al., 2014). Moreover, the diagnostic criteria for ADHD are largely based on directly observable behaviors (American Psychiatric Association, 2013).

The aim of the study is to obtain initial information about a class of students, for whom the teacher claims that the number of interruptions and inappropriate behaviors is excessive, according to his perception. Specifically, the objective is defined as estimating the proportion of time in which the students are involved in off-tasks behaviors and on-task behaviors. Subsequent evaluation is planned for future research assessing whether the relative duration of off-tasks behaviors is excessive and whether they are systematically related to any of the students for which there is a suspicion by the teacher that they might present problems with deficit of attention or impulsivity.

Decision #1: Choose an Observational Design

The design in Observational Methodology is the strategy determining the course of action or sequence of decisions about how to collect, organize and analyze the data, always subordinate to the objectives of the study (Anguera et al., 2001). The purpose of an observational design is to identify the axes of time (when to record: in a continuous or discontinuous way?), behaviors (what to record: one or several target behaviors?) and subjects (who to record: one or several participants?) involved in an investigation, in order to be able to propose the best strategy in an observation situation.

In the math class studied there are 10 students. According to the subject axis, a nomothetic design (Allport, 1942; Anguera et al., 2001) is used, given that all children are observed. According to the behavioral axis, a multidimensional design is used, given that there are several different specific behaviors coded as "off-task" or "on-task" (see the "Decision #2: Choose What to Code" subsection). According to the time axis (see **Figure 1**) and the inter-sessional criterion, the design is a "tracking" one (also referred as "follow-up" design), as several sequential sessions are to be recorded. According to the time axis and the intra-sessional criterion, time sampling has to be used, as discussed in the subsection entitled "Decision #3: Decide Whether Time Sampling Is Required." The beginning and end of the observational sessions (i.e., the uninterrupted time of recording) are defined according to the duration of the math classes.

Decision #2: Choose What to Code

Systematization of the recordings consists in expressing in observable terms all the information contained in behaviors or events, in order to improve objectivity. The behavioral units (i.e., the minimal behavioral manifestation that is considered meaningful) can be distinguished according to their duration, being either "states" (longer units, for which duration matters) or "events" (brief events, for which duration is not recorded; Altmann, 1974). Additionally, it is possible to distinguish the behavioral units according to their content, being "structural" (a physical movement or location, defined in time and space), "functional" (consequence of the structural units on the physical or social environment), or "causal" (causes of the structural units). Finally, the behavioral units can be classified according to their degree of abstraction, leading to "molecular" categories based on Weick's (1968) response levels: verbal, vocal, gestures, and proxemics behavior or to "molar" categories (complex combinations of these response levels with a greater degree of abstraction, implying a certain amount of inference about the intentions).

For instance, Ardoin and Martens (2000) adapted Barkley's (1990) Restricted Academic Situation and distinguished the following categories: off-task (interruption of the child's attention from the task to engage in another behavior such as breaking eye contact with the worksheet), fidgeting (repetitive, purposeless movement of the legs, feet, arms, hands, fingers, buttocks, or trunk), vocalization (verbal noises), plays with objects (touching objects not directly related to the task, desk or child's own body), and out of seat (child's buttocks breaking contact with the seat). Similarly, yet slightly different, Stahr et al. (2006) mention as examples of "off-task behavior" the repetitive pencil tapping, head or leg shaking and fidgeting, drawing, gazing around class; leaving the assigned instructional area, and making audible vocalizations not related to the instructional task. Stahr et al. (2006) define "on-task behavior" as attending to or participating in instructional activities as requested by classroom staff (e.g., looking at the teacher while she was instructing, doing or attempting the assigned task, seeking assistance, and following directions). Therefore, the on-task and off-task behaviors refer to different response levels (i.e., they are "molar" categories), coded according to their relation to the academic task taking place at any given moment. Moreover, the focus is put on the function of the behavior rather than its location or the specific movement in any part of the body; thus, the units are "functional". Finally, whereas some of the specific instances of on-task behavior can be "events" (e.g., shifting the gaze from the book to the blackboard), the "on-task behavior" category itself is rather a "state," given that it is expected to have a certain duration.

Decision #3: Decide whether Time Sampling Is Required

In the running example, carrying out the observational study involving the direct presence of observers in the environment

FIGURE 1 | Representation of intersessional and intrasessional time sampling.
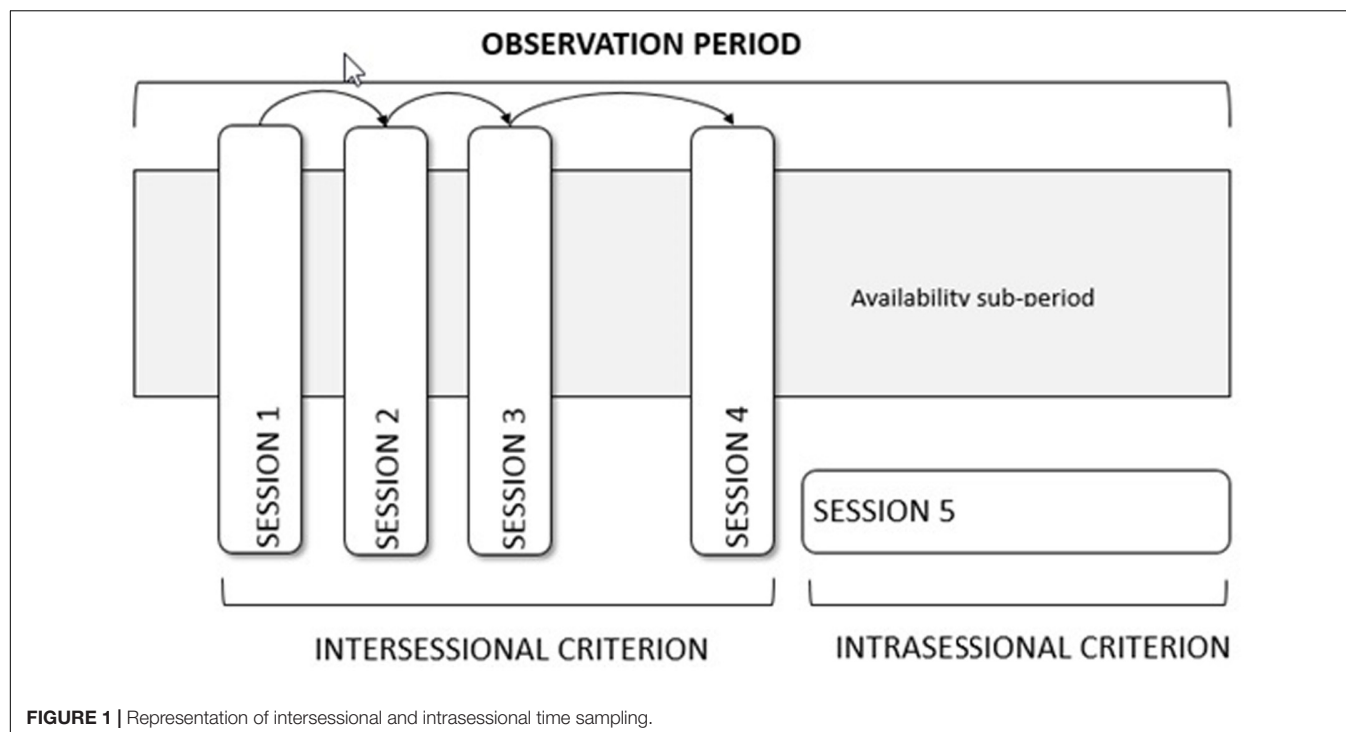
would require an authorization from the school. One approach would be "recording activated by transitions" (RAT), in which the observer is coding every transition from one category to another, optionally recoding duration times as well, without any time-related divisions of the observation session. However, a RAT would require video parents' authorization for videotaping. Therefore, time sampling would be required. When the recording rule is conceptualized as "recording activated by units of time" (RAUT), the observation session is divided into many short intervals in which an observer determines if an event occurs (Barlow et al., 2009). These intervals are usually of constant duration, although in some cases intervals with variable duration are also possible (Test and Heward, 1984; Ayres and Gast, 2010). The main types of observational recording procedures that follow a RAUT rule are momentary time sampling (MTS, in which only the category taking place at the end of the time interval is recorded), partial interval recording (PIR, in which any category appearing at any point during the time interval is recorded) and whole interval recording (WIR, in which an occurrence is recorded only in case one category takes place throughout the whole interval) (Arrington, 1943; Hutt and Hutt, 1970; Cooper et al., 2007). In terms of taxonomies, Suen and Ary (1989) refer to PIR and WIR as "semi-continuous" recording and to MTS as "discrete" recording, whereas other authors (e.g., Rapp et al., 2011) refer to MTS, PIR, and WIR as "discontinuous" recording. The main features of the MTS, PIR, and WIR are described in **Table 1**.

Opting for MTS, PIR, or WIR as feasible alternatives to continuous recording is justified on the basis that all these recording procedures have been commonly used in a variety of disciplines (e.g., Mudford et al., 2009, report that discontinuous

recording was used in 45% of the articles reviewed; Adamson and Wachsmuth, 2014, report that MTS was used in 9% of the articles using direct observation and a time-based system like PIR or WIR was used in 48% of the studies, versus 39% using an event-based code). Moreover, MTS, PIR, and WIR may inform about whether a behavior is likely to occur at the beginning, mid, or end of an observation period, which cannot be assessed via event coding only.

Given that there are several participants to be observed, this can be achieved using multifocal sweep sampling, and more specifically, its alternating variant. This within-session sampling of focal participants takes places as follows. The observational session lasting for 100 min is divided into ten 10-min fractions. In the first fraction, during the 1st minute participant 1 is observed, during the 2nd minute participant 2 is observed, and so forth up to participant 10 being observed during the 10th minute. In the second fraction, during the 1st minute participant 2 is observed, during the 2nd minute participant 3 is observed, and so forth up to participant 10 being observed during the 9th minute and participant 10 being observed in the 10th minute. The sequence of observing the participants continues accordingly up to the 10th fraction in which during the 1st minute participant 10 is observed, during the 2nd minute participant 1 is observed, and so forth up to participant 9 being observed during the 10th minute. This alternating multifocal sweep sampling (represented on **Figure 2**) ensures that all individuals are observed in all fractions and, additionally, that all individuals are observed in different parts of the fractions (i.e., not always at the beginning or at the end). Subsequently, it is necessary to choose the interval length and the specific observational recording procedure to use (see next section).

TABLE 1 | Main features of the observational recording procedures following a recording activated by units of time (RAUT) rule.

Feature	Momentary time sampling	Partial interval sampling	Whole interval sampling
Recording rule	Code the category that occurs at the end of the interval	Code any category that occurs during any moment within the interval	Code the category that occurs during the whole interval or as per Hutt and Hutt (1970) code the predominant category
Need for observer attention during the whole interval	No, only at the end of the interval	Yes, unless all categories in the coding scheme already took place at least once	Yes, unless the category present since the beginning of the interval stops occurring before its end
Minimum number of categories that can be coded in an interval	1	1	0
Maximum number of categories that can be coded in an interval	1	As many as categories present in the coding scheme	1
Coding of several occurrences within a single interval	Coded as one occurrence, only if taking place at the end of the interval; otherwise, 0	Coded as one occurrence.	Coded as zero occurrences, assuming that a non-occurrence takes place in between
Coding of a single occurrence spreading over two intervals	Coded as one occurrence, assuming that it takes place at the end of the interval; coded as two occurrences if it last until the end of the second interval	Coded as two occurrences	Coded as zero occurrences, unless it takes place during the whole interval (coded as 1 or 2).

Decision #4: Choosing A Raut Observational Recording Procedure

There are three reasons why the choice of a discontinuous recording procedure is important. First, they are expected (and have been shown) to present random or systematic errors, due to the fact that these procedures do not record the frequency and duration of each category (Gardenier et al., 2004). Second, it has been shown (Rapp et al., 2011) that the type of observation recording procedure used on the same real behavioral stream is related to the degree of interobserver agreement (IOA). This finding suggests that high values of IOA are not necessarily the result of high concordance between data collected by two independent observers, but could also stem from procedural features. Third, inaccuracy of MTS, PIR, and WIR in estimating count and duration also has an effect on subsequent analyses performed for giving an answer to the research question of interest (e.g., see Ledford et al., 2015, for results related to estimating effects in single-case designs; Barlow et al., 2009). Accordingly, there have been efforts to propose effect size indices, whose values do not depend on the observation recording procedure (Pustejovsky, 2015).

The factors that have been related to the presence of error are: (a) the type of time sampling method used (Powell et al., 1977; Simpson and Simpson, 1977; Murphy and Goodall, 1980; Green et al., 1982; Gardenier et al., 2004; Alvero et al., 2007; Rapp et al., 2008; Devine et al., 2011); (b) the length of the intervals used (Dunbar, 1976; Leger, 1977; Powell et al., 1977; Mansell, 1985; Mudford et al., 1990; Alvero et al., 2007, 2011; Rapp et al., 2008; Devine et al., 2011) and (c) factors related to the categories of interest, such as its frequency (McDowell, 1973; Powell et al., 1977; Murphy and Goodall, 1980; Green et al., 1982; Gardenier et al., 2004; Alvero et al., 2011) and duration (Murphy and Goodall, 1980; Sanson-Fisher et al., 1980; Green et al., 1982; Ary and Suen, 1986). In general, it has been observed that when the duration of the interval (τ) is small relative to the duration of the category and the spaces

FIGURE 2 | Graphical representation of within-session sampling of participants to be the focus of observation: alternating multifocal sweep.
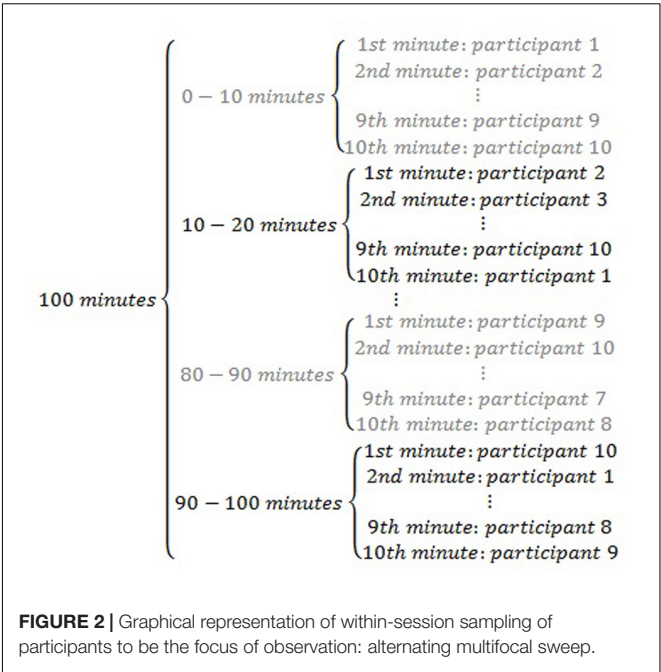
between categories, the estimates will be more precise (Suen and Ary, 1989). The number of factors and the number references provided suggests that choosing a discontinuous recording procedure and, additionally, choosing an interval length are not necessarily straightforward tasks. The interactive graphs we created and implemented in a web page are intended to provide guidance for this specific decision in the process of conducting an observational study.

Aim of the Article

Given that observation is commonly present in research and it is also included in the curricula of university majors such as Psychology and Educational Sciences, it is important to illustrate

the conditions (e.g., interval length, average duration and prevalence of the behavior of interest) in which MTS, PIR, and WIR are expected to perform well when estimating estimate the frequency and prevalence of the behavior of interest. Specifically, we here describe the development of interactive graphs available in a free web page, with the aim to make accessible to students and applied researchers the complex simulation evidence, taking into consideration several factors at a time.

METHOD

Justification of the Need for Simulation

Simulations offer several advantages over the analysis of observational records obtained from real situations. First, simulations entail knowing the truth about the parameters of the underlying process from which the observed behavioral streams arise. More concretely, the researcher can specify the average duration of the behavior each time that it occurs (i.e., how long or how short are the individual occurrences, on average) or the proportion of time that it takes place (i.e., what is the prevalence of the behavior). Second, an evaluation of the sampling methods with a simulation eliminates the error attributable to the human observer. The possible error produced in the records by the observer can be indirectly attributed to a series of variables such as biological factors, psycho-social factors, reaction time, motivation, behavior perceptibility (Repp et al., 1976; Tyler, 1979; Green et al., 1982; Saudargas and Zanolli, 1990; Murphy and Harrop, 1994; Taylor et al., 2012). Third, the measurement error can be quantified either in terms of absolute error values (i.e., difference between estimated and actual durations) or in terms of relative error values (i.e., the difference expressed as a proportion of the actual durations of the events; this is the option we followed here).

Data Generation Model

For generating the behavioral stream of occurrences and their duration we used the alternating renewal process (ARP) model (Pustejovsky and Runyon, 2014), implemented in the *ARPobservation* package for R (R Core Team, 2016). ARP treats both the length of behavioral events and the interim times (i.e., interresponse time between events) as random quantities (Pustejovsky and Swan, 2015).

The review of simulation studies performed by Pustejovsky and Runyon (2014) showed that most studies followed a procedure that agrees with the ARP model, whereas others mostly followed a *random onset model* in which the point of onset for a behavioral event is chosen at random repeatedly, on the basis of a predetermined duration per occurrence, and usually avoiding overlaps (e.g., Ledford et al., 2015). Another procedure followed in previous research (Rapp et al., 2011) is to use real data gathered via continuous recording and then to convert this data to interval measures on the basis of MTS, PIR, or WIR.

The main advantage of the ARP model and the *ARPobservation* package is that it mimics the actual process in which there is first a behavioral stream and then data are

gathered according to a predetermined procedure (continuous recording, MTS, PIR, or WIR). Moreover, the ARP model offers great flexibility in simulating behavioral streams with different characteristics (Pustejovsky and Runyon, 2014).

The assumptions of the ARP model include (Pustejovsky and Swan, 2015): the event duration times corresponding to the same observation session are assumed to be identically distributed; the interim times corresponding to the same observation session are assumed to be identically distributed[1]; the length of the next event or interim time does not depend on the sequence of events leading up to it; there is a constant probability that an event is occurring at any given point in time during the observation session (i.e., the behavior stream is in equilibrium).

Data Generation Parameters

The following are the relevant simulation parameters that describe the main characteristics of the observational situation:

(a) Session duration: set to 10, 20, 30, and 60 min. Previous studies included sessions of 10 min (Ledford et al., 2015), or 10, 15, and 30 min (Rapp et al., 2011). According to the review performed by Pustejovsky and Runyon (2014) the range of session durations is between 10 and 300 min, with most common lengths being 30 or 60 min.

(b) The prevalence of the behavior of interest (π) is defined as proportion of duration with respect to whole observation session length. We used the range from 5 to 95% in steps of 5%. Ledford et al. (2015) varied prevalence from 10 to 70%, Pustejovsky and Runyon (2014) provided an illustration with values from 1 to 99%, in steps of 1%, and the prevalence from Rapp et al.'s (2011) real data sets ranged from 10 to 93%.

(c) The incidence per time unit is defined as the average number of times that a behavior occurs, for instance, per minute. Pustejovsky and Runyon (2014) provided an illustration with values ranging from 0.1 to 0.5. Incidence is not manipulated directly in the *ARPobservation* package; we rather tallied the occurrences and divided the sum by the observation session length, measured in minutes. Pustejovsky and Runyon (2014) define incidence, within the ARP framework, as being equal to $1/(\mu + \lambda)$, where μ is the average event duration and λ is the average interim time. In our simulations, incidence ranged from 0.1 to 3.2 (according to the prevalence of the behavior) for $\mu = 18$ s and from 1.5 to 28.6 (according to the prevalence of the behavior) for $\mu = 2$ s.

(d) Average event duration (μ), also referred to as mean bout duration or average "duration per occurrence" (DPO): ranging from 2 to 120 s in our simulation. Comparatively, Ledford et al. (2015) set DPO to 2 or 10 s, whereas Pustejovsky and Swan (2015) provide an illustration with DPO = 6 s. In the review performed by Pustejovsky and Runyon (2014), the most common DPOs were in the range of 1 to 120 s, with three of the 14 studies using greater values of the maximum DPOs, up to 500 s.

---

[1]Several possible distributions can be specified for the event durations and interim times, but we followed Pustejovsky and Runyon (2014) in using an exponential distribution.

(e) Average interim time (λ): this parameter was determined according to the previously defined average DPO and prevalence. Specifically, given that Pustejovsky and Runyon (2014) define prevalence, within the ARP framework, as being equal to $\pi = \mu/(\mu + \lambda)$, then $\lambda = (\mu - \mu\pi)/\pi$. Thus, for instance, for $\pi = 0.5$, the average interim time was equal to μ, whereas for $\pi = 0.3$ the average interim time ranged from 4.67 s (for $\mu = 2$ s) to 42 s (for $\mu = 18$ s). Ledford et al. (2015) did not set interim times explicitly, as they apparently followed the random onset model rather than the ARP model. Pustejovsky and Swan (2015) use 12 s, with the most common values ranging from 2 to 60 s according to the Pustejovsky and Runyon (2014) review.

(f) For discontinuous recording, the interval length (τ) has to be set. We used interval lengths ranging from 2 to 20 s. Ledford et al. (2015) used intervals of 2 and 20 s as well, whereas Rapp et al. (2011) used 10 s, and Pustejovsky and Swan (2015) 5 and 20 s. A relevant aspect highlighted by Ledford et al. (2015) and Pustejovsky and Swan (2015) is whether the interval is longer or shorter than the average DPO, which is related to the degree to which estimates of count and duration obtained from discontinuous recording misrepresent the measures from continuous recording. Specifically, Ledford et al. (2015) studied interval size relative to DPO ranging from 0.33 to 3.33. In the interactive graphs we developed we also included a calculation of interval size relative to DPO for the specific combination of conditions selected by the user. For the shortest interval ($\tau = 2$ s) and longest behavior ($\mu = 18$ s), the ratio τ/μ is 0.11, whereas for the longest interval ($\tau = 20$ s) and shortest behavior ($\mu = 2$ s), the ratio τ/μ is 10.

Figure 3 illustrates how the parameters can be selected in the web application and it also shows how the website presents the information about the ratio τ/μ, and about average interim time and incidence per minute for each of the values of prevalence.

Data Analysis

With the ARP model it is possible to assess the performance of discontinuous recording in two different ways (Pustejovsky and Runyon, 2014). On the one hand, it is possible to compare the measures from discontinuous recording to the ones that would be obtained in continuous recording. In this case, we would be assessing how well the observed behavior is represented, taking into account that MTS, PIR, and WIR entail time sampling within the observation session. This approach takes into consideration the fact that continuous recording does not contain intrasession sampling error (Suen and Ary, 1989). On the other hand, it is possible to compare both the measures from discontinuous recording and the measures from continuous recording to the parameters that generate the behavior stream. According to this latter approach, the behavior observed in a given session and measured via continuous recording is only a realization of the underlying process, as selecting the moments for the observation sessions also involves time sampling of the behavior of the organism studied. This approach takes into consideration the fact that continuous recording may contain intersession sampling

error (Suen and Ary, 1989). Both kinds of comparison are possible with the interactive graphs created.

The interactive graphs offer results for 1, 100, or 1000 samples. The results for 1 sample illustrate what could happen in any given study (in which the results from continuous recording need not match perfectly well the underlying process generating the behavior in a given observation session), whereas the results for 100 and especially for 1000 samples are more informative of the general performance of the discontinuous recording techniques as compared to continuous recording. When the results for 100 or 1000 samples are represented graphically, apart from the average value, we also provide information about the scatter: one or two standard deviations away from the mean, represented in orange and red, respectively.

The following terms are relevant for the results illustrated in the interactive graphs:

(a) Pseudofrequency: $PF = r + n_{01}$, where r is equal to 0 when the first interval in the observation session is not marked and it is equal to 1 when this interval is marked as denoting occurrence of the behavior, and n_{01} is the number of transitions from non-occurrence (unmarked interval) to occurrence (marked interval). According to Suen and Ary (1989), *PF* would be an unbiased estimator of frequency when the interval is shorter than the shortest behavior duration and shorter than one half the shortest interim time.

(b) Modified frequency (θ) = number of intervals for which the occurrence of the behavior of interest is marked when using MTS, PIR, or WIR. The modified frequency would be the most straightforward way of estimating frequency, although evidence has shown that it is imperfect. We have included this way of estimating frequency for MTS and PIR in order to enable exploring whether it is appropriate in any of the conditions tested. Moreover, the modified frequency is also in the basis of estimating prevalence; in general, it is assumed that prevalence is estimated as $\widehat{\pi} = \theta/n$, where n is the number of intervals into which the observation session is divided. However, for PIR and WIR corrections have been proposed (Suen and Ary, 1989): $\widehat{\pi}_{PIR} = (\theta - PF)/n$ and $\widehat{\pi}_{WIR} = (\theta + PF)/n$, respectively.

(c) For PIR we also applied a formula for estimating frequency that is not based solely on the modified frequency (Altmann and Wagner, 1970):

$$\widehat{f} = -\left(n \times \ln\left(1 - \frac{\theta}{n} \right) \right)$$

This formula is expected to function well when: (a) the behavior of interest is an event (i.e., it has a very short duration, practically equal to zero), and (b) the probability of occurrence of the behavior of interest is independent of the time that has passed since the last time it occurred, as the case for a Poisson distribution. In relation to point (b), in the ARP model "[a]ll interim times and all event durations are generated in a mutually independent manner, which means that the length of a given event is influenced neither by the
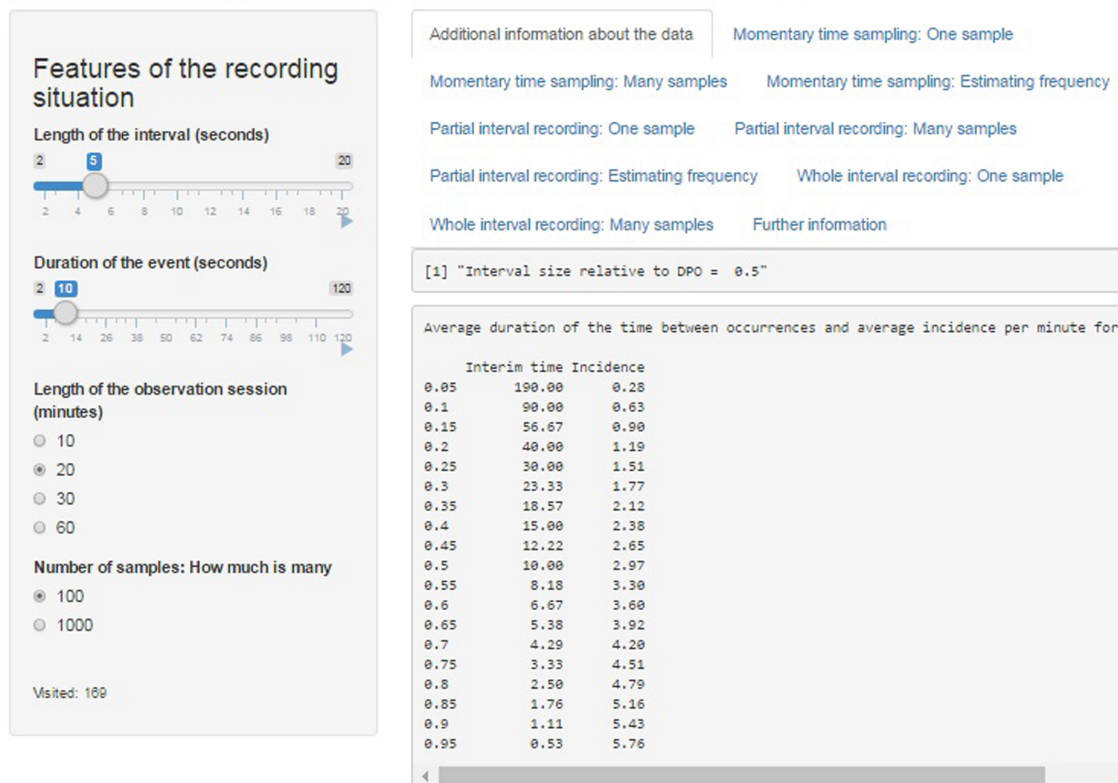
FIGURE 3 | Screenshot of the web application created using Shiny. The average interim time and the average incidence per minute can be seen for each value of prevalence, considering the average duration of the event.

length of previous events nor by how long it has been since the last event ended" (Pustejovsky and Runyon, 2014, p. 213).

Finally, the amount of error when estimating prevalence is quantified as relative bias, using the formula: $(\widehat{\pi} - \pi)/\pi$, where π is the value of the simulation parameter for prevalence and $\widehat{\pi}$ is the estimated obtained using MTS, PIR, or WIR. For PIR and WIR, relative bias is computed separately for estimating prevalence as $\widehat{\pi} = \theta/n$ or as $\widehat{\pi}_{PIR} = (\theta - PF)/n$ and $\widehat{\pi}_{WIR} = (\theta + PF)/n$.

Development of the Application

The illustrations are based on the ARP model and the *ARPobservation* package and have been prepared using Shiny applications[2], for two reasons. First, from the perspective of the interested reader, Shiny is freely available and user-friendly, given that the only actions required to obtain the graphical and numerical results are selecting options from the left-hand side menus and clicking the tabs in the upper part of the browser (see **Figure 3**). Second, from the perspective of the researcher

and developer, Shiny communicates easily with R[3], which is the free platform in which the ARP model is implemented. This communication is made efficient thanks to RStudio[4]. The interactive graphs and tables are available at http://jlosada. shinyapps.io/Prevalence.

OUTPUT OF THE APPLICATION

Obtaining the Results

When accessing http://jlosada.shinyapps.io/Prevalence the user can manipulate the options at the left of the web browser in order to specify several features defining the observation session: (a) length of the observation session; (b) length of the interval in seconds; (c) the average duration of the behavior of interest in seconds; and (d) the number of samples when presenting the results of more than one sample. When a selection is made (or with the default selection), information is provided in the initially active tab called "Additional information about the data." In the first row, the ratio of the interval length (τ) to average DPO (μ) is provided. Afterward, a table is presented containing the

---

[2]www.shinyapps.io

[3]http://cran.r-project.org
[4]https://www.rstudio.com/products/rstudio/

average interim time (λ) and the average incidence per minute for each of the values of prevalence (π) of the behavior of interest. A screenshot including this information is provided in **Figure 3**.

The remaining tabs offer two types of information. On the one hand, there are graphical representations of the estimated prevalence (on the ordinate) for each simulation parameter π on the abscissa (e.g., **Figures 4**, **5** for MTS and **Figure 6** for PIR). On the other hand, there are tabular representations of the estimated frequency (third column for MTS; third and fourth columns for PIR) compared to the average frequency as determined by continuous recording (second column), for each value of prevalence (e.g., **Figure 7** for PIR). The information is obtained by clicking on the tabs, with several seconds required for the corresponding simulations to take place and to provide the graphical or tabular output.

Using the Application for Pedagogic Purposes

An initial pedagogic purpose could be to illustrate the concept of sampling variability, clicking on any of the three tabs illustrating the results of one sample. When comparing the results of the recordings in a single observation session with the simulation parameters that defined the underlying process generating the behavioral stream, the graphs make obvious that not even continuous recording is absolutely perfect for estimating prevalence. This is due to the fact

the behavior observed in a given session is only a sample. The results for MTS and continuous recording are usually similar for short intervals and when the average DPO is longer than the interval used in MTS. **Figure 4** presents an example.

A second purpose could be to illustrate the degree to which there is overestimation or underestimation of prevalence, according to the interval length (τ) and average DPO (μ), while also considering the actual simulation parameter π. For that purpose, the play buttons for τ and μ can be used in order to provide a visual impression of the importance of these factors and how they interact. The play buttons are useful when presenting the results for one sample, because the use of many iterations requires time and the play buttons are not practical anymore. However, the graphical representations generated on the basis of 100 or 1000 iterations can be saved and compared afterward by putting them side by side.

In general, over many iterations, when the comparison is performed with the simulation parameters that defined the underlying process generating the behavioral stream, prevalence is estimated without bias when continuous recording and MTS are used. For MTS, more precise estimates of prevalence (i.e., narrower standard deviation bands, as represented on the interactive graphs) are obtained for: (a) shorter intervals (i.e., smaller τ), (b) behaviors with shorter duration μ, and (c) longer observation sessions. **Figure 5** presents an example.

For PIR prevalence is overestimated. However, when the correction proposed by Suen and Ary (1989) is applied,

FIGURE 4 | Screenshot of the web application created using Shiny. Prevalence of a behavior with average duration per occurrence of 8 s as estimated in a single observation session of 20 min, using continuous recording (green dots) and momentary time sampling [MTS] (empty triangles) based on a 5-s interval. The numerical values represent the relative bias of the estimation using MTS.

FIGURE 5 | Screenshot of the web application created using Shiny. Average prevalence of a behavior with average duration per occurrence of 8 s as estimated in 1000 observation sessions of 20 min, using continuous recording (green dots) and MTS (empty triangles) based on a 6-s interval. The dashed lines represent one and two standard deviations above and below the average estimate by MTS. The numerical values represent the relative bias of the estimation using MTS.

this overestimation is attenuated, although not removed, consistent with the findings of Rogosa and Ghandour (1991). Complementarily, for WIR prevalence is underestimated, but the correction leads to attenuating this overestimation. For both PIR and WIR, in terms of bias, the averages of estimates are closer to the simulation parameters for: (a) lower actual levels of prevalence ($\pi \leq 0.3$) than for higher ones, (b) shorter intervals in general (e.g., for $\tau = 2$ s PIR provides practically unbiased estimates of prevalence), (c) smaller τ/μ ratio, as reported by Ledford et al. (2015), and (d) longer observation sessions. More precise estimates of prevalence are obtained for actual prevalence close to 0 or 1, due to the bounds of the index, and also for the three previously mentioned situations. **Figure 6** shows an example for one of the favorable conditions for PIR, but for which the estimation of prevalence is also biased.

For PIR, regarding the estimation of frequency via the formula by Altmann and Wagner (1970), the results obtained indicate that in no condition (not even when $\mu = 2$ s) did the formula provide a good estimate of frequency, as computed via continuous recording. Actually, the results illustrated in the graphs are worse than the ones reported by Ledford et al. (2015), who used θ as an estimate of count and found that smaller counts were estimated better in longer intervals and larger counts were estimated better in shorter intervals. In few situations meeting these conditions the estimates of frequency using θ were within 10% of the actual count. **Figure 7** shows a snapshot of the table generated in the

website, illustrating the abovementioned findings about these two ways of estimating frequency when using PIR.

Using the Application for Applied Research Purposes

When the aim of the use of the Shiny application is to choose an appropriate interval for a given RAUT, there are several possible scenarios. First, if absolutely no prior information is available, the applied researcher would have to follow an approach similar to the one describe for the pedagogic use of the Shiny application.

Second, in some cases it is possible to have an empirically based expectation on the approximate prevalence of the behavior of interest. For instance, the estimated prevalence of on-task behavior for children with ADHD has been reported to be between 0.30 and 0.50 according to the moment of the observation session (Rapport et al., 2009), an average of 0.64 with a standard deviation of 0.06 (Junod et al., 2006), or as high as an average of 0.71 average with a standard deviation of 0.16 (Mahar et al., 2006). For such high values of expected prevalence, the even the $\widehat{\pi}_{PIR} = (\theta - PF)/n$ estimates of prevalence are always positively biased (e.g., see **Figure 6** and, specifically, the red crosses, denoting the estimates of prevalence, above the diagonal black line representing unbiased estimation, for prevalences greater than 0.2), but the overestimation is attenuated when

FIGURE 6 | Screenshot of the web application created using Shiny. Average prevalence of a behavior with average duration per occurrence of 12 s as estimated in 100 observation sessions of 60 min, using continuous recording (green dots) and partial interval recording (PIR) (empty triangles without the correction; red crosses with the correction) based on a 6-s interval. The dashed lines represent one and two standard deviations above and below the average of the corrected estimates by PIR. The numerical values represent the relative bias of the estimation using PIR: black values refer to using the modified frequency in the numerator, whereas red values refer to using modified frequency minus pseudofrequency in the numerator.

the average DPO is $\mu \geq 30$ and $\tau \leq 5$ (e.g., see **Figure 8** and, specifically, the red crosses on the diagonal black line for practically all values of prevalence). If the there is no evidence for assuming $\mu \geq 30$, on the one hand, and $\tau \leq 5$ is judged not to be practical, on the other hand, then PIR should not be considered as an adequate observation recording procedure. In such a case, it would be advisable to use MTS instead of PIR.

A third situation would entail having information about both the likely range of prevalence and the average DPO, although the latter has been claimed to be seldom reported (Ledford et al., 2015). If we use the information from Rapport et al. (2009) that the average duration of on task behavior for children diagnosed with ADHD and low attention is 2 min (120 s), an interval length of $\tau = 15$ (as actually used by Rapport et al., 2009) would be justified, as illustrated from **Figure 9** in which the estimates of prevalence (red crosses) are practically unbiased (i.e., close to the diagonal line).

Summary of the Results of the Application

Concerning the estimation of prevalence and frequency, the evidence of the performance of discontinuous recording procedures is very complex, due to the fact that this performance is affected by many interacting factors. This complexity makes difficult summarizing the results via a simple rule. For instance, Ayres and Gast (2010) suggest that WIR is more appropriate when the behavior of interest is of low frequency and long duration, whereas PIR is appropriate for behaviors of high frequency and short duration, given that the frequency of long duration behaviors may be overestimated. This statement can be verified from the interactive graphs. Moreover, more nuanced knowledge can be obtained, as it can be verified that the frequency of short duration behaviors is also overestimated, for certain combinations of interval length τ and average behavior duration μ with $\tau > \mu$, when the prevalence π is relatively low (below 0.45 for some combinations of τ and μ or below 0.75 for other combinations). Regarding MTS, Ayres and Gast (2010) state that it is appropriate for behaviors with high frequency and long durations and that this recording procedure has a tendency to underestimate frequency and overestimate duration. Using the interactive graphs it can be shown that prevalence is actually not overestimated, whereas the underestimation of frequency is only present when the length of the interval is greater than the average duration of the event ($\tau > \mu$); in contrast, frequency is overestimated when the length of the intervals is shorter than the average duration of the event ($\tau < \mu$) and the estimation is unbiased when behavior and interval are of the same length.

```
[1] "For each value of prevalence:"
[1] " "
      CR-count  Formula-estimate  Modif.Freq.-estimate
0.05    10.00            15.10                    13.63
0.1     19.65            30.00                    24.78
0.15    29.69            47.95                    35.71
0.2     39.32            67.35                    45.09
0.25    49.69            92.90                    54.40
0.3     59.74           114.31                    60.35
0.35    68.74           145.43                    66.41
0.4     80.01           181.01                    71.09
0.45    90.61              Inf                    74.44
0.5     99.90              Inf                    76.65
0.55   110.74              Inf                    78.22
0.6    118.16              Inf                    79.32
0.65   130.96              Inf                    79.72
0.7    139.82              Inf                    79.98
0.75   149.26              Inf                    79.99
0.8    161.34              Inf                    80.00
0.85   171.10              Inf                    80.00
0.9    179.99              Inf                    80.00
0.95   187.61              Inf                    80.00
```

FIGURE 7 | Screenshot of the web application created using Shiny. Frequency estimates of a behavior with average duration per occurrence of 6 s as estimated in 100 observation sessions of 20 min, using PIR) based on a 15-s interval (i.e., the rate of interval length to duration per occurrence is 2.5).

Although the aim of the interactive graphs was to provide nuanced information, taking into account the specific interval lengths, DPOs, and prevalences, it should be noted that **Table 2** includes a necessarily simplified summary of the performance of the time sampling methods for estimating prevalence and frequency. This summary suggest that MTS can be recommended to be used when the aim is to estimate prevalence (e.g., **Figure 5**), especially when interval is short and when the average DPO of the behavior is short. These results concurs with previous findings regarding the lack of systematic bias (Tyler, 1979; Harrop and Daniels, 1986); specifically, Rogosa and Ghandour (1991) note that MTS is useful for estimating prevalence, but not incidence or event duration.

In contrast, the results concur with previous findings about PIR overestimating of the frequency and prevalence of the categories (Tyler, 1979; Harrop and Daniels, 1986), which is why Rogosa and Ghandour (1991) state that PIR does not provide useful information on incidence, prevalence, or event duration. More specifically, the results from the interactive graphs suggest that PIR can only be used for estimating prevalence in case 2 $\tau < \mu$ and for $\pi \leq 0.3$ (e.g., **Figure 6**). For WIR, the requirement is even more stringent: 3 $\tau < \mu$.

This result is consistent with previous findings about the underestimation when using WIR being greater for longer intervals (Alvero et al., 2007). Thus, if the prevalence is not known beforehand and if the bout durations are relatively short, PIR and WIR should not be used when the objective is to estimate prevalence.

In terms of estimating frequency, this can be done without systematic error only when the average DPO is known and it is used for defining the interval length when using MTS. For PIR the requirements involve prevalence as well, which means that it is a less practical option. In summary, the choice of a time sampling method is an important one in order to avoid inaccurate descriptions of the degree to which the phenomena of interest are present or inaccurate comparisons, especially if different observational recording procedures are used for the different behaviors observed. For instance, Abikoff et al. (2002, p. 353) use MTS and WIR to obtain "behavioral rates" of children with ADHD and Junod et al. (2006) use MTS and PIR to estimate prevalence of several behaviors children with and without ADHD; in neither of the two cases is there any mention of average DPO or prevalence.

FIGURE 8 | Screenshot of the web application created using Shiny. Average prevalence of a behavior with average duration per occurrence of 35 s as estimated in 100 observation sessions of 60 min, using continuous recording (green dots) and PIR (empty triangles without the correction; red crosses with the correction) based on a 4-s interval. The dashed lines represent one and two standard deviations above and below the average of the corrected estimates by PIR. The numerical values represent the relative bias of the estimation using PIR: black values refer to using the modified frequency in the numerator, whereas red values refer to using modified frequency minus pseudofrequency in the numerator.

TABLE 2 | Performance of the observational recording procedures following a recording activated by units of time (RAUT) rule.

Feature	Momentary time sampling	Partial interval sampling	Whole interval sampling
General summary of performance for estimating prevalence	Unbiased estimation; more efficient for shorter τ and μ	Overestimation even with correction, but less severe when $\tau < \mu$ and when π is small	Underestimation, even with correction, but less severe when $\tau < \mu$ and when π is small
General summary of performance for estimating frequency	Estimation via modified frequency: (a) underestimation when $\tau > \mu$; (b) overestimation when $\tau < \mu$; (c) unbiased estimation when $\tau = \mu$	Estimation via modified frequency: overestimation, unless $\tau > \mu$, but depending on π. Estimation via the formula by Altmann and Wagner (1970): more severe overestimation.	Not included in the application, as the literature review does not provide support.

τ, interval length; μ, average duration per occurrence; π, prevalence.

DISCUSSION

Advantages and Limitations of the Application

The application constructed has several advantages. First, it is available online free of charge. Second, the application is user-friendly in the sense that no programming skills are required and the selection of the values of the factors defining the observational situation is made by clicking. Third, according to the review performed by Pustejovsky and Runyon (2014), the ARP model used for the simulation is a framework representing most of the simulation studies on observational data. Fourth, for obtaining the results of the simulation, it is not necessary to specify potentially unavailable information, such as the average incidence per minute. Accordingly, it is not strictly speaking necessary to know the average DPO beforehand, given that the user can select several likely values using the slider in the application. For the same purpose (i.e., not requiring specific knowledge about the expected prevalence), the graphical representations cover practically the whole range of possible prevalences. In that sense, it is not required to have information about the specific values of incidence, average DPO or prevalence
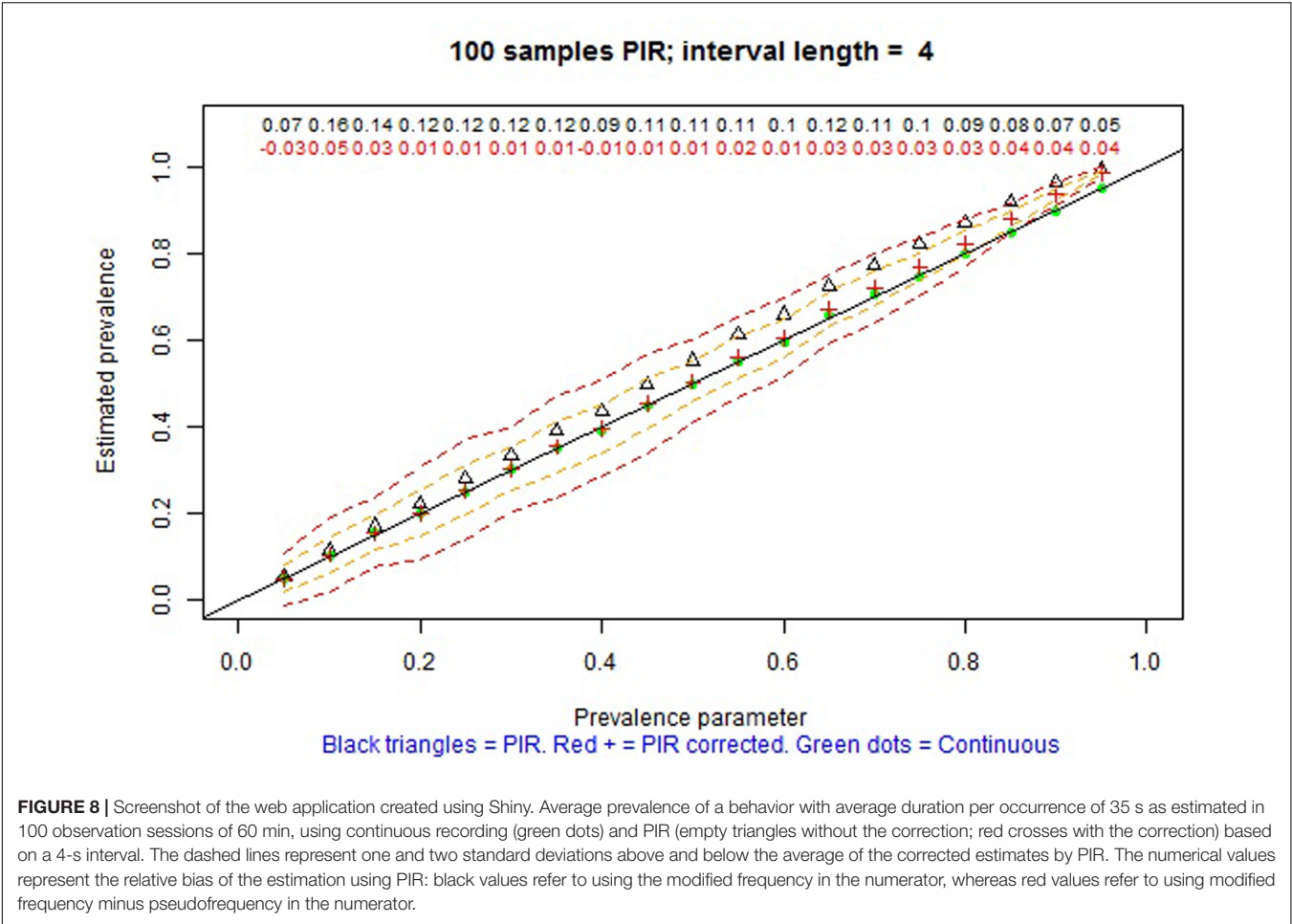
FIGURE 9 | Screenshot of the web application created using Shiny. Average prevalence of a behavior with average duration per occurrence of 120 s as estimated in 100 observation sessions of 60 min, using continuous recording (green dots) and PIR (empty triangles without the correction; red crosses with the correction) based on a 15-s interval. The dashed lines represent one and two standard deviations above and below the average of the corrected estimates by PIR. The numerical values represent the relative bias of the estimation using PIR: black values refer to using the modified frequency in the numerator, whereas red values refer to using modified frequency minus pseudofrequency in the numerator.

to get a general insight of the interval lengths that are justified to be used. Fifth, the variety of parameter values for defining the observation situation (i.e., observation session length, average DPO, prevalence of the behavior of interest, interval length and the average interim time, incidence, and ratio of interval length to average DPO) is greater than the one present in recent simulation studies.

Besides strengths, it is especially important to dedicate space to the limitations of the application, taking into account the use of simulation as a basis (e.g., von Oertzen and Brandmaier, 2013). Regarding the limitations of the application, an initial technical limitation refers to the fact that the simulations are performed when the user selects the values defining the observational situations rather than accessing information (e.g., stored in data matrices) of already performed simulations. Therefore, it is not possible to always obtain instantly the results when performing 100 or 1000 iterations. Our calculations suggest that for 1000 iterations for MTS approximately 5 s are needed, whereas for 100 iterations for PIR require between 10 and 15 s. Second, we can mention as limitations the assumptions of the ARP model mentioned previously (i.e., the event duration times corresponding to the same observation session are assumed to be identically distributed and there is a constant probability that an event is occurring at any given point in time during the observation session) and to the fact that we used only one distribution (the exponential) for modeling event durations and

interim times. Third, a limitation of the evidence provided in the Shiny application is related to the way in which the behavior stream is converted into strings of categories. Specifically, human error is not included in the simulation process and this represents a relevant future endeavor for modifying the *ARPobservation* package that is used as a basis of the simulations. Fourth, the graphical representations do not cover all possible combinations of average DPO and interval length. Therefore, as is the case for any simulation, the evidence cannot be considered as representing perfectly all real situations, but it can be used as an indication in absence of better simulation models or in absence of specific knowledge about interval lengths that have been proven to be useful for estimating the prevalence of given behaviors.

Implications for Teachers and Methodologists

In order to improve the way in which knowledge is transmitted or, more accurately, the way in which students construct knowledge (Driver et al., 1994), there are already efforts focused on statistical topics, including specialized journals such as *Understanding Statistics*. However, some topics specific to observational methodology need more attention. In that sense, from the perspective of the teacher or methodologist, the three types of competence (McLagan, 1997; Kaslow, 2004) are involved in constructing and using the interactive graphs

presented in the current text: (a) the fundament is the *attitude* to try to improve teaching methods; (b) specific *knowledge* is constructed by the teacher or methodologist in relation to the conditions (e.g., average duration per occurrence of the behavior, interval length, and ratio of the two) in which each of the discontinuous observation recording procedures perform best; and (c) methodological *skills* are developed by learning to use software specifically designed for simulating behavior in observation sessions and for using different recording procedures. Interactive graphs such as the ones presented here make possible a presentation of empirical findings that is both more detailed (i.e., covering a greater range of conditions) and more accurate (i.e., avoiding oversimplifications and representing the amount of bias present in the different conditions).

Implications for Students and Applied Researchers

The same three types of competence are also involved from the perspective of the student or applied researcher: (a) the fundament is the *attitude* or disposition to follow the best possible practices when choosing the recording procedure to use for observing overt behaviors; (b) *knowledge* or subject matter is constructed, in this case, on the topic referring to the strengths and limitations of different observation recording procedures (continuous recording, MTS, PIR, and WIR); and (c) methodological *skills* or abilities are expected to be developed by getting acquainted with the simulation procedure followed for studying the quality of the measures obtained in MTS, PIR, and WIR (i.e., extensive application to generated data with known characteristics or to actual behavioral data for which continuous recording has been carried out). In relation to the methodological abilities, it is crucial that students and applied researchers not only trust that the content taught by their teachers and textbooks is correct, but that they are aware that subject content is the result of research (e.g., via simulation) and that this research also presents certain limitations such as the ones mentioned in "Advantages and limitations of the application." In summary, getting to know how knowledge is obtained is expected to make students and applied researchers exercise their critical thinking skills (although comprehensive programs are required for developing such skills; Halpern, 1998) and the disposition to always look for more refined and more precise knowledge.

Limitations and Future Research

In terms of limitations, the present paper does not necessarily add new knowledge in terms of research findings. This is due to the fact that its purpose is mainly related to illustrating the complex relations of several factors influencing the accuracy of the estimates obtained via several observation recording procedures. Moreover, as previously mentioned, the factors included in the simulation do not include human error and one of its likely causes, fatigue. It could be logically argued that MTS entail smaller cognitive load (as attention is required only at the end of the interval), but fatigue is related to several additional factors such as the observer's familiarity with the behavior, the interval length, the number of categories to be recorded, the average DPO of the behaviors and the degree to which they are easily distinguished (Altmann, 1974). Such information has to be considered, jointly with the evidence on the estimation of prevalence and frequency when selecting a RAUT.

Future illustrations can focus on study of reliability and, more specifically, agreement between observers. Rapp et al. (2011) showed how the values of percentage of agreement are different according to the observation recording procedure, but such illustrations are also necessary for kappa, which is recommended for quantifying agreement (Suen and Ary, 1989). Specifically, the kappa value obtained for continuous recording on a second-by-second comparison (Bakeman and Gottman, 1986) can be compared to the kappa values obtained via MTS, PIR, and WIR for varying degrees of prevalence of the behavior of interest, given that this parameter has impact on the kappa values (Suen and Ary, 1989).

AUTHOR CONTRIBUTIONS

The initial idea was due to JL and RM and it was further developed jointly by both authors. The manuscript was written by JL (Introduction) and RM (Method, Results, and Discussion). Both authors participated in several revisions during the process of creating, discussing, and improving the manuscript. Both authors gave their consent that this final version submitted for publication and agreed in their co-responsibility regarding all aspects of the work, such as the accuracy of the data and the integrity of the research.

REFERENCES

Abikoff, H. B., Jensen, P. S., Arnold, L. E., Hoza, B., Hechtman, L., Pollack, S., et al. (2002). Observed classroom behavior of children with ADHD: relationship to gender and comorbidity. *J. Abnorm. Child Psychol.* 30, 349–359. doi: 10.1023/A: 1015713807297

Adamson, R. M., and Wachsmuth, S. T. (2014). A review of direct observation research within the past decade in the field of emotional and behavioral disorders. *Behav. Disord.* 39, 181–189.

Allport, F. H. (1942). Methods in the study of collective action phenomena. *J. Soc. Psychol.* 15, 165–185. doi: 10.1080/00224545.1942.9921526

Altmann, J. (1974). Observational study of behavior: sampling methods. *Behaviour* 49, 227–266. doi: 10.1163/156853974X00534

Altmann, S. A., and Wagner, S. S. (1970). Estimating rates of behavior from Hansen frequencies. *Primates* 11, 181–183. doi: 10.1007/BF01731143

Alvero, A. M., Rappaport, E., and Taylor, M. A. (2011). A further assessment of momentary time-sampling across extended interval lengths. *J. Organ. Behav. Manage.* 31, 117–129. doi: 10.1080/01608061.2011.569203

Alvero, A. M., Struss, K., and Rappaport, E. (2007). Measuring safety performance: a comparison of whole, partial, and momentary time-sampling recording methods. *J. Organ. Behav. Manage.* 27, 1–28. doi: 10.1300/J075v27n04_01

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. Washington, DC: Author. doi: 10.1176/appi.books. 9780890425596

Anguera, M. T., Blanco-Villaseñor, Á., and Losada, J. L. (2001). Diseños observacionales, cuestión clave en el proceso de la metodología observacional. [Observational designs, a critical question in the process of observational methodology]. *Methodol. Behav. Sci.* 3, 135–160.

Ardoin, S. P., and Martens, B. K. (2000). Testing the ability of children with attention deficit hyperactivity disorder to accurately report the effects of

medication on their behavior. *J. Appl. Behav. Anal.* 33, 593–610. doi: 10.1901/jaba.2000.33-593

Arrington, R. E. (1943). Time sampling studies of social behavior: a critical review of technique and results with research suggestions. *Psychol. Bull.* 40, 81–124. doi: 10.1037/h0053539

Ary, D., and Suen, H. K. (1986). Interval lengths required for unbiased frequency and duration estimates with partial, whole, and momentary time sampling. *Midwes. Educ. Res.* 8, 17–24.

Ayres, K., and Gast, D. L. (2010). "Dependent measures and measurement procedures," in *Single Subject Research Methodology in Behavioral Sciences*, ed. D. L. Gast and J. R. Ledford (New York, NY: Routledge), 129–165.

Bakeman, R., and Gottman, J. M. (1986). *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge: Cambridge University Press.

Barkley, R. A. (1990). *Attention Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment*. New York, NY: Guilford.

Barlow, D., Nock, M., and Hersen, M. (2009). *Single Case Experimental Designs: Strategies for Studying Behavior Change*, 3rd Edn. Boston, MA: Allyn and Bacon.

Beck, J. W., Scholer, A. A., and Schmidt, A. M. (2016). Workload, risks, and goal framing as antecedents of shortcut behaviors. *J. Bus. Psychol.* doi: 10.1007/s10869-016-9450-0

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Canton, A., and Hileno, R. (2016). Goal scoring in soccer: a polar coordinate analysis of motor skills used by Lionel Messi. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Cooper, J. O., Heron, T. E., and Heward, W. L. (2007). *Applied Behavior Analysis*, 2nd Edn. Upper Saddle River, NJ: Pearson.

Devine, S. L., Rapp, J. T., Testa, J. R., Henrickson, M. L., and Schnerch, G. (2011). Detecting changes in simulated events using partial-interval recording and momentary time sampling III: evaluating sensitivity as a function of session length. *Behav. Intervent.* 26, 103–124. doi: 10.1002/bin.328

Dishion, T. J., Mun, C. J., Tein, J. Y., Kim, H., Shaw, D. S., Gardner, F., et al. (2016). The validation of macro and micro observations of parent–child dynamics using the relationship affect coding system in early childhood. *Prev. Sci.* 18, 268–280. doi: 10.1007/s11121-016-0697-5

Dolado, R., Cooke, C., and Beltran, F. S. (2016). How many for lunch today? Seasonal fission-fusion dynamics as a feeding strategy in wild red-capped mangabeys (*Cercocebus torquatus*). *Folia Primatol.* 87, 197–212. doi: 10.1159/000449220

Driver, R., Asoko, H., Leach, J., Scott, P., and Mortimer, E. (1994). Constructing scientific knowledge in the classroom. *Educ. Res.* 23, 5–12. doi: 10.3102/0013189X023007005

Dunbar, R. I. M. (1976). Some aspects of research design and their implications in the observational study of behaviour. *Behaviour* 58, 78–98. doi: 10.1163/156853976X00244

Fassnacht, G. (1982). *Theory and Practice of Observing Behavior*. London: Academic Press.

Gardenier, N. C., MacDonald, R., and Green, G. (2004). Comparison of direct observational methods for measuring stereotypic behavior in children with autism spectrum disorders. *Res. Dev. Disabil.* 25, 99–118. doi: 10.1016/j.ridd.2003.05.004

Green, S. B., McCoy, J. F., Burns, K. P., and Smith, A. C. (1982). Accuracy of observational data with whole interval, partial interval, and momentary time-sampling recording techniques. *J. Behav. Assess.* 4, 103–118. doi: 10.1007/BF01321385

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: disposition, skills, structure training, and metacognitive monitoring. *Am. Psychol.* 53, 449–455. doi: 10.1037/0003-066X.53.4.449

Harrop, A., and Daniels, M. (1986). Methods of time sampling: a reappraisal of momentary time sampling and partial interval recording. *J. Appl. Behav. Anal.* 19, 73–77. doi: 10.1901/jaba.1986.19-73

Hutt, S. J., and Hutt, C. (1970). *Direct Observation and Measurement of Behavior*. Springfield, IL: Charles C Thomas.

Junod, R. E. V., DuPaul, G. J., Jitendra, A. K., Volpe, R. J., and Cleary, K. S. (2006). Classroom observations of students with and without ADHD: differences across types of engagement. *J. Sch. Psychol.* 44, 87–104. doi: 10.1016/j.jsp.2005.12.004

Kaslow, N. J. (2004). Competencies in professional psychology. *Am. Psychol.* 59, 774–781. doi: 10.1037/0003-066X.59.8.774

Langer, J. K., Lim, M. H., Fernandez, K. C., and Rodebaugh, T. L. (2016). Social anxiety disorder is associated with reduced eye contact during conversation primed for conflict. *Cogn. Ther. Res.* 41, 220–229. doi: 10.1007/s10608-016-9813-x

Ledford, J. R., Ayres, K. M., Lane, J. D., and Lam, M. F. (2015). Identifying issues and concerns with the use of interval-based systems in single case research using a pilot simulation study. *J. Spec. Educ.* 49, 104–117. doi: 10.1177/0022466915568975

Leger, D. W. (1977). An empirical evaluation of instantaneous and one-zero sampling of chimpanzee behavior. *Primates* 18, 387–393. doi: 10.1007/BF02383116

Mahar, M. T., Murphy, S. K., Rowe, D. A., Golden, J., Shields, A. T., and Raedeke, T. D. (2006). Effects of a classroom-based program on physical activity and on-task behavior. *Med. Sc. Sport. Exer.* 38, 2086–2094. doi: 10.1249/01.mss.0000235359.16685.a3

Mansell, J. (1985). Time sampling and measurement error: the effect of interval length and sampling pattern. *J. Behav. Ther. Exp. Psy.* 16, 245–251. doi: 10.1016/0005-7916(85)90070-9

McDowell, E. E. (1973). Comparison of time-sampling and continuous recording techniques for observing developmental changes in caretaker and infant behaviors. *J. Genet. Psychol.* 123, 99–105. doi: 10.1080/00221325.1973.10533192

McLagan, P. A. (1997). Competencies: the next generation. *Training Dev.* 51, 40–48.

Mudford, O. C., Beale, I. L., and Singh, N. N. (1990). The representativeness of observational samples of different durations. *J. Appl. Behav. Anal.* 23, 323–331. doi: 10.1901/jaba.1990.23-323

Mudford, O. C., Taylor, S. A., and Martin, N. T. (2009). Continuous recording and interobserver agreement algorithms reported in the Journal of Applied Behavior Analysis (1995-2005). *J. Appl. Behav. Anal.* 42, 165–169. doi: 10.1901/jaba.2009.42-165

Murphy, G., and Goodall, E. (1980). Measurement error in direct observations: a comparison of common recording methods. *Behav. Res. Ther.* 18, 147–150. doi: 10.1016/0005-7967(80)90109-6

Murphy, M. J., and Harrop, A. (1994). Observer error in the use of momentary time sampling and partial interval recording. *Br. J. Psychol.* 85, 169–179. doi: 10.1111/j.2044-8295.1994.tb02517.x

Pasquaretta, C., Klenschi, E., Pansanel, J., Battesti, M., Mery, F., and Sueur, C. (2016). Understanding dynamics of information transmission in *Drosophila melanogaster* using a statistical modeling framework for longitudinal network data (the RSiena package). *Front. Psychol.* 7:539. doi: 10.3389/fpsyg.2016.00539

Polanczyk, G. V., Willcutt, E. G., Salum, G. A., Kieling, C., and Rohde, L. A. (2014). ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *Int. J. Epidemiol.* 43, 434–442. doi: 10.1093/ije/dyt261

Powell, J., Martindale, B., Kulp, S., Martindale, A., and Bauman, R. (1977). Taking a closer look: time sampling and measurement error. *J. Appl. Behav. Anal.* 10, 325–332. doi: 10.1901/jaba.1977.10-325

Pustejovsky, J. E. (2015). Measurement-comparable effect sizes for single-case studies of free-operant behavior. *Psychol. Methods* 20, 342–359. doi: 10.1037/met0000019

Pustejovsky, J. E., and Runyon, C. (2014). Alternating renewal process models for behavioral observation: simulation methods, software, and validity illustrations. *Behav. Disord.* 39, 211–227. doi: 10.1177/019874291303900406

Pustejovsky, J. E., and Swan, D. M. (2015). Four methods for analyzing partial interval recording data, with application to single-case research. *Multivar. Behav. Res.* 50, 365–380. doi: 10.1080/00273171.2015.1014879

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/

Rapp, J. T., Carroll, R. A., Stangeland, L., Swanson, G., and Higgins, W. J. (2011). A comparison of reliability measures for continuous and discontinuous recording methods: inflated agreement scores with partial interval recording and momentary time sampling for duration events. *Behav. Modif.* 35, 389–402. doi: 10.1177/0145445511405512

Rapp, J. T., Colby-Dirksen, A. M., Michalski, D. N., Carroll, R. A., and Lindenberg, A. M. (2008). Detecting changes in simulated events using partial-interval

recording and momentary time sampling. *Behav. Intervent.* 23, 237–269. doi: 10.1002/bin.269

Rapport, M. D., Kofler, M. J., Alderson, R. M., Timko, T. M. Jr., and DuPaul, G. J. (2009). Variability of attention processes in ADHD: observations from the classroom. *J. Atten. Disord.* 12, 563–573. doi: 10.1177/1087054708322990

Repp, A. C., Roberts, D. M., Slack, D. J., Repp, C. F., and Berkler, M. S. (1976). A comparison of frequency, interval, and time-sampling methods of data collection. *J. Appl. Behav. Anal.* 9, 501–508. doi: 10.1901/jaba.1976.9-501

Rogosa, D., and Ghandour, G. (1991). Statistical models for behavioral observations. *J. Educ. Stat.* 16, 157–252. doi: 10.3102/10769986016003157

Sanson-Fisher, R. W., Poole, A. D., and Dunn, J. (1980). An empirical method for determining an appropriate interval length for recording behavior. *J. Appl. Behav. Anal.* 13, 493–500. doi: 10.1901/jaba.1980.13-493

Saudargas, R. A., and Zanolli, K. (1990). Momentary time sampling as an estimate of percentage time: a field validation. *J. Appl. Behav. Anal.* 23, 533–537. doi: 10.1901/jaba.1990.23-533

Simpson, M. J. A., and Simpson, A. E. (1977). One-zero and scan methods for sampling behaviour. *Anim. Behav.* 25, 726–731. doi: 10.1016/0003-3472(77)90122-1

Stahr, B., Cushing, D., Lane, K., and Fox, J. (2006). Efficacy of a function-based intervention in decreasing off-task behavior exhibited by a student with ADHD. *J. Posit. Behav. Interv.* 8, 201–211. doi: 10.1177/10983007060080040301

Suen, H. K., and Ary, D. (1989). *Analyzing Quantitative Behavioral Observation Data*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Taylor, M. A., Skourides, A., and Alvero, A. M. (2012). Observer error when measuring safety-related behavior: momentary time sampling versus whole-interval recording. *J. Organ. Behav. Manage.* 32, 307–319. doi: 10.1080/01608061.2012.729389

Test, D. W., and Heward, W. L. (1984). "Accuracy of momentary time sampling: a comparison of fixed-and variable-interval observation schedules," in *Focus on Behavior Analysis in Education*, eds W. L. Heward, T. E. Heron, D. Hill, and J. Trap-Porter (Columbus, OH: Merrill), 177–194.

Tyler, S. (1979). Time-sampling: a matter of convention. *Anim. Behav.* 27, 801–810. doi: 10.1016/0003-3472(79)90016-2

Vaughn, B. E., Santos, A. J., Monteiro, L., Shin, N., Daniel, J. R., Krzysik, L., et al. (2016). Social engagement and adaptive functioning during early childhood: identifying and distinguishing among subgroups differing with regard to social engagement. *Dev. Psychol.* 52, 1422–1434. doi: 10.1037/dev0000142

von Oertzen, T., and Brandmaier, A. M. (2013). Optimal study design with identical power: an application of power equivalence to latent growth curve models. *Psychol. Aging* 28, 414–428. doi: 10.1037/a0031844

Weick, K. E. (1968). "Systematic observational methods," in *Handbook of Social Psychology*, Vol. II, eds G. Lindzey and E. Aronson (Reading, MA: Addison-Wesley), 357–451.

Movement Notation Revisited: Syntax of the Common Morphokinetic Alphabet (CMA) System

Conrad Izquierdo[1]* and M. Teresa Anguera[2]*

[1] Faculty of Psychology, Autonomous University of Barcelona, Barcelona, Spain, [2] Faculty of Psychology, Institute of Neurosciences, University of Barcelona, Barcelona, Spain

Advances in the study of non-verbal behavior and communication have generated a need for movement transcription systems capable of incorporating continuous developments in visual and computer technology. Our research team has been working on the construction of a common morphokinetic alphabet (CMA) for the systematic observation of daily life activities. The project, which was launched several years ago, was designed to create a system for describing and analyzing body motion expression, physical activity, and physical appearance. In this paper, we describe an idiosyncratic application of Noam Chomsky's *phrase marker grammar* to the morphokinetic phrase, the objective being to establish the grammatical rules and basic order of the symbol string according to a relational tree formed by the breakdown of the syntactic components identified as structuring the visual description of movement. Criteria for using the CMA as a coding system and a free transcription system are proposed.

Keywords: movement behavior, observational methodology, field format, coding system, morphokinetic alphabet, grammar, movement phrase-structure, rules

INTRODUCTION

This article discusses the theoretical and methodological aspects of the behavior stream through consideration of the problem of movement notation (Hadar, 1994). As Freedman (1981) has pointed out, the analysis of movement behavior by means of narrative description (or natural language) was superseded by the postulated use of symbolic notation systems, and current efforts are focused on obtaining objective measures that require the development of reliable coding systems (Donaghy, 1988) which, given that language is a code of a notational system (Krajcsi and Szabó, 2012), fulfill linguistic conditions: "The term 'code' as it is used here refers to the final result of three parallel processes, simplifying the original material, organizing it so that the relationship among its elements can be clear, and restructuring the whole for easy transmission." (Dittmann, 1987, p. 39).

In the field of research into non-verbal communication any attempt to develop a movement alphabet must inevitably take into account Ray Birdwhistell's Kinesic Notational System (KNS) (Birdwhistell, 1952, 1970) and the Facial Action Coding System (FACS), which is limited to facial expressions (Ekman and Friesen, 1978), without forgetting that there is a wide range of other approaches to this problem (Barker and Collins, 1970; Kendon, 1981, 1997; Crivelli et al., 2016).

In terms of notational systems it is important to highlight the following: (1) the most important aspect is not the tokens but the ability of the system to provide an exhaustive representation of people's anatomical possibilities for movement; (2) how easy or hard a notational system is to use must be evaluated in terms of the relationship between clarity and precision; and (3) the acceptance and use of a notational system is determined by the consensus reached among researchers of the scientific community in question.

Hirsbrunner et al. (1987) reviewed the problem of movement transcription and argued that although the required efficiency of the coding language cannot be replaced by video technology for recording visual information (or audible information when analyzing a multimodal system of communication), neither is greater efficacy achieved by developing impressive notational schema to transform audiovisual information into data. Faced with such a situation the authors adopted the diagnosis and proposed solution of Frey and Pool (1976): "...current difficulties in movement description do not originate from the complexity of phenomena to be described, but from the investigators' failure to base their coding systems on the principle of time-series notation." (Hirsbrunner et al., 1987, p. 100).

The Bernese Time-Series Notation invokes a classic expression coined by Frey and Pool (1976), one which has proved highly powerful at resolving the continuity of the behavior stream (unifying speech and movement) by using nominal or categorical codes and seeking to detect space-time patterns. It starts by obtaining matrices of recorded data by transcribing movements, whether simple or complex, and thus generates a large amount of simultaneous data over time. Hirsbrunner et al. (1987) praise its possible application to the different parts of the body during movement and, after two decades of successive technological advances, the hurdle once posed by frame-by-frame transcription has now been totally overcome, thus ensuring high degrees of precision when obtaining data.

In this context, and having carefully reviewed the literature on the analysis of movement in the field of dance (Hutchinson-Guest, 1984) and physical appearance (Fink et al., 2015), as well as the most well-known and scientifically sound notational systems (Laban, 1926; Benesh and Benesh, 1956; Eshkol and Wachman, 1958; Eshkol, 1971; Bartenieff and Lewis, 1980; Farnell, 1996; Guest, 2011), we sought to develop a notational grammar of body movement which we call the Common Morphokinetic Alphabet (CMA).

The term "morphokinetic" is defined as a temporally demonstrable change in properties and spatial design of body motion form. By "common" we understand two things: (1) the notation system can be communicated and learned, as a balance is sought between clarity and precision and (2) the notation system shares the logic of meaning, physical identity/semantic content, which emerges from the writing of movement in the notational systems reviewed according to the choreographic model. Finally, the concept of "alphabet" denotes the conventional and discretional nature of the tokens and connotes the material condition sine qua non required to develop a notational system governed by grammatical rules.

In previous publications (Izquierdo and Anguera, 2001; Anguera and Izquierdo, 2006; Izquierdo, 2010) we have mapped out the different facets of this theoretical and methodological proposal that forms part of the movement observation process within the field of psychology (Chinellato et al., 2015; Castañer et al., 2016; Anguera et al., 2017). Now our aim is to present the CMA grammar and the criteria for its use in systematic observation studies. The following sections address the theoretical and methodological basis of the CMA notation, the grammatical formalization of the morphokinetic description, the general criteria for use in the coding format and as a free transcription system, and, finally, the possibilities offered by this notational system.

THEORETICAL-METHODOLOGICAL BASIS OF THE CMA

Advances in the study of non-verbal behavior and communication have led to the need for suitable systems for transcribing movement that are capable of incorporating continuous developments in visual and computer technology (Archer, 1991; Anguera, 2003; Blanco-Villaseñor et al., 2003; Portell et al., 2015a,b).

The range of possibilities offered by visual records and the physical analysis of behavior in the context of everyday human activity was clearly illustrated by the pioneering photographic work (in some cases, including magnificent images of reality) of cultural anthropologists, such as Mead and Bateson (1942) and Efron (1942), or clinical researchers, such as Scheflen (1964), who used images taken from film stills to indicate the path of gestures. Despite the promise of this early work, however, the relationship between visual records and the space-time analysis of movement in interactive and non-interactive situations has, as pointed out by Ekman (1964) faced a number of significant problems, some of which concern the KNS.

The warning raised in the context of interdisciplinary research into social behavior (Farnell, 1999) about the dangers of reducing kinesic descriptions to the anatomical functioning of the human body in order to achieve greater analytic rigor concerns the way of interpreting the application of kinesiology to kinesics more than it does the fact of basing the choreographic model of notation on anatomical and biomechanical knowledge (Shafir et al., 2016). In sum, "a comprehensive movement writing system has to resolve several technical difficulties. Human actions take place in three dimensions of space and one dimension of time and mobilize many parts of the body simultaneously. [. . .]. The task is complex, surely, but not insurmountable [. . .]" (Farnell, 1996, p. 868).

The CMA aims to code the visual form of body movement by describing it as a configuration sculpted in space-time. Each new configuration perceived by the observer implies a demonstrable change with respect to the immediately previous one. The change in configuration includes total or partial mobility of the body and relative stillness with respect to the following position, and, whenever necessary, the initial position can be maintained as a basic reference point for subsequent changes.

In terms of the spatial description of body movement a determining feature is that the body has a large number of degrees of freedom when executing movements. Bearing in mind this principle, CMA notation of spatial points is geared toward what is specific about the spatial design of a movement in accordance with the objective of the observation (Frey and Pool, 1976).

Given that body space is located in space/setting, CMA notation considers the movements through space that we can make with our body and the relationship between the use of space and overall body positions, that is, physical postures (standing, sitting, kneeling, lying down, etc.) and the postural movements produced by a form of established behavior (Mehrabian, 1969; Argyle, 1973; Poyatos, 1986).

In terms of timing, the morphokinetic description of a series of movements involves making a decision about the time interval to be used in order to obtain a good resolution of discontinuity, which results in the presence or absence of certain primary data that are considered in light of what is significant for the analysis (Page, 1996). Here a discretional criterion is used, which ranges from the frame-by-frame reading of the image to viewing at normal speed (Hirsbrunner et al., 1987).

A complete understanding of the temporal structure of movement phenomena involves the notation of the duration and temporal form (i.e., simultaneous or sequential) of movements. In addition to qualities concerning the speed, intensity and amplitude of changes, the use of signals derived from the physical appearance of the moving subjects and their socio-historical, cultural, and linguistic context are also transcribed in order to distinguish variations and individual differences in the kinesic form and style (Scheflen, 1972; Poyatos, 1986; Kendon, 1997).

From a methodological perspective the CMA has notable potential in that it is able to objectify behavioral units at the micro level due to the way it breaks up the stream of behavior (Condon and Ogston, 1970), and this gives it important analytic properties for subsequent empirical processing.

The first part of the analytic process consists of transforming the kinesic reality of human movement into units of behavior that are later turned into data with the aid of an observation instrument developed ad hoc; these data must be suitably managed before being analyzed, a task for which there are various approaches. Thus there are four stages that are necessary from a methodological point of view and that provide the CMA with its required consistency.

(1) The creation of molecular units (Schegloff, 2000) is a prior condition in the development of an observation instrument (Anguera and Izquierdo, 2006). In our view, demarcating the unit of behavior is clearly linked to the specific setting of objectives, and it must also be possible to demarcate, name, and define each unit.

(2) The decision taken regarding the demarcation of units opens the way to the development of an observation instrument: the field format. Originally, in the work of Weick (1968, 1985), this was a simple recording technique, but it has subsequently been revisited, developed (Anguera, 2003; Sánchez-Algarra and Anguera, 2013), and used widely in numerous fields, especially those involving

movement, such as sport (Anguera et al., 2003). The development of the instrument involves the following steps: (a) Establishment of the criteria or axes of the instrument, which are set in accordance with the study objectives (for example, in observing a person who is learning to swim these might be area of the swimming pool, entering the water, submersion, equilibriums, displacements, etc.). Some of these criteria may be broken down hierarchically into others. (b) Listing of behaviors/situations (this list is neither closed nor exhaustive, and is known as the catalog) corresponding to each one of the criteria, and noted according to the information provided by the exploratory stage of the study. For example, starting from the criterion entering the water the list of behaviors could be entering feet first with help, entering from a sitting position on the edge of the pool without help, entering head first without help, etc. (the etc., indicates precisely that further behaviors can be added as the list is not closed). (c) Assignment of a decimal coding system to each one of the listed behaviors/situations that are derived from each one of the criteria. This means that any of the behaviors or situations can be displayed in a hierarchical system of lower order. Depending on the complexity of the case in question or the desired range of molecularity, these coding systems may be double, triple, etc. For example, the codes of the criteria would be 1 (area), 2 (entering the water), 3 (submersion), etc. And from 2 we could derive 2_1 (entering feet first with help), 2_2 (entering from a sitting position on the edge of the pool without help), 2_3 (entering head first without help), 2_4 (entering by jumping feet first from the side of the pool without help), etc. However, from 2_2 we could also derive 2_2_1, 2_2_2, 2_2_3, and so on successively[1]. (d) Drawing up of a list of criteria configurations. The configuration is the basic unit in recording field formats, and consists of linking together the codes corresponding to simultaneous or concurrent behaviors, thus enabling an exhaustive recording of the behavior stream and greatly facilitating the subsequent analysis of data[2]. For example, in the event that four criteria have been proposed:

1_3 2_4 3_2_1 4_2
1_3 2_3 3_2_1 4_2
1_2 2_3 3_2_4 4_4
…

(3) Having a tailor-made field format offers enormous flexibility in terms of data gathering, but it must be properly managed if the data in question are to be positioned in a way that optimizes their subsequent analysis. Given that the field format configurations are chains of simultaneous codes (synchronous relationship), and the sequence of these criteria configurations is established over time (diachronically), the modification of a single code over time is sufficient to yield the recording in the next

---

[1] *Lower-case hyphen*: links the digits that form a new code comprised by two or more already-assigned codes.

[2] *Empty space in the code chain (rows)*: indicates the concurrence of different codes.

row. Furthermore, the passage of time can be measured in conventional units (seconds) or in frames, and it is even possible to consider a conventional interval of any chronometric unit for each row of the matrix. In other words, the code matrices obtained will have, at most, the same number of columns as there are field format criteria, while the number of rows will depend on the successive changeability of the observed situation. This is perfectly in keeping with the proposal of Coster (2005, p. 17) as regards the preparation of data corresponding to postural dynamics, and is consistent with the structure of time-series notation. In this regard, there is a correspondence with the gathering of homogeneous data through the "restrictive coding" suggested by Frey and Pool (1976). Coster, following the way in which music is written, refers to the *horizontal* dimension, indicative of diachrony, and the *vertical* dimension, corresponding to synchrony or concurrence of behaviors. This proposal was found to fit perfectly with that resulting from the data management obtained when recording by means of field formats.

(4) Once the recorded data have been suitably managed a decision must be made as to the most appropriate analytic technique, always bearing in mind the objectives proposed in each case and the corresponding design. The three favored options, in light of their analytic potential, are lag sequential analysis (Sackett, 1980; Bakeman and Gottman, 1986; Bakeman and Quera, 1995), used to detect, if present, patterns or regularities in the series of recorded behaviors; detection of T-Patterns (Magnusson, 1996, 2000; Magnusson et al., 2015), which has a wide range of applications (Anolli et al., 2005), including non-verbal communication (Haynal-Reymond et al., 2005) and facial expressions (Merten and Schwab, 2005); polar coordinate analysis (Perea et al., 2012; Castañer et al., 2016), and time-series analysis on the basis of categorical variables (Bakeman and Gottman, 1986; Albert, 2001). Graphical representations of the temporal structure of body movement during a given period of time are also of great interest (Frey et al., 1982).

GRAMMATICAL FRAMEWORK OF THE CMA

Chomsky referred to generative grammar as "a system of rules that in some explicit and well-defined way assigns structural descriptions to sentences" (Chomsky, 1965, p. 8). The function of these rules is to specify whether the minimum terminal units of syntactic function comprise well-formed strings (phrase markers).

From the methodological point of view, Chomsky (1956) bases his investigation of the syntax of a natural language on the detailed analysis of what traditional grammar has to say about a simple statement. To this end he analyzes the following example: *sincerity may frighten the boy*. After considering the example from different perspectives he distinguishes three levels of information which may be extracted from the sentence.

Each level implies referring to notions used in the syntactic and morphological analysis of the language; notions, such as "nominal phrase" and "verb", from the first information level, are clearly distinguished from functional grammatical notions (e.g., subject, predicate, direct object, etc.) on the second level. The lexical and grammatical elements appear on the third level. Chomsky aims to determine "how information of this sort can be formally presented in a structural description and how such structural descriptions can be generated by a system of explicit rules" (Chomsky, 1965, p. 64).

In terms of our grammatical framework for the morphokinetic alphabet, it is sufficient to consider these two questions in relation to the first level of information: the breakdown of the sentence [S] into successive series on the basis of nominal [NP] and verb [VP] syntagms. The phrase marker indicates three types of information: (1) category labels (i.e., NP, V, Det, etc.); (2) the hierarchical arrangement of these categories; and (3) the linear order of the terminal string (**Figure 1**). The linear order of the terminal string from the category symbol 'S', which represents "Sentence", is obtained by applying a sequence of rewriting rules (**Table 1**).

Movement-Phrase Structure

By analogy to Chomsky's procedure, the analysis of the structural components of a terminal string of the morphokinetic alphabet must answer three basic questions present within the movement notation systems reviewed (Izquierdo and Anguera, 2001; Izquierdo, 2010): "What has moved?" "What has changed?" and "How has it changed?" The information provided by Laban Notation, Benesh Movement Notation, and Eshkol-Wachman Notation in answer to these basic questions differs slightly as they have different reference frameworks and orthography. The Laban and Benesh systems have a richer vocabulary than the Eshkol-Wachman system when it comes to describing movement (form, space, time, and temporalization) and the qualitative aspects or "how one move" (Neagle et al., 2002; Guest, 2011).

Within the framework of the CMA, the first question requires us to name and identify the bodily form of movement: part of the body + figure. The second question must be answered by specifying spatial and temporal references with respect to position (overall physical posture, position on the floor, or any other aspect related to the maintenance of overall physical posture), orientation (position in the movement plane, direction, and height), and the duration of movements and their structure in time. The final question (How has it changed?) involves identifying the contextual factors which may affect the form of movement and classifying the specific mode of the motor action. The contextual factors, considered as invariant at least within the same observation session, include the situation where the activity takes place, the baseline body and psychosocial conditions of the person (or persons) in movement, the reference culture, and the acquired habituation in executing the movements (i.e., slow movers, lively movers, etc.). The qualities perceived for specific movements, that is, the impression we form of speed (e.g., slow/fast), intensity (e.g., gentle/strong), and amplitude (e.g., narrow/wide), as well as the use of physical appearance, including styles related to culture or social status that are not selected by the

FIGURE 1 | Triple structural information of the phrase marker: tree diagram.
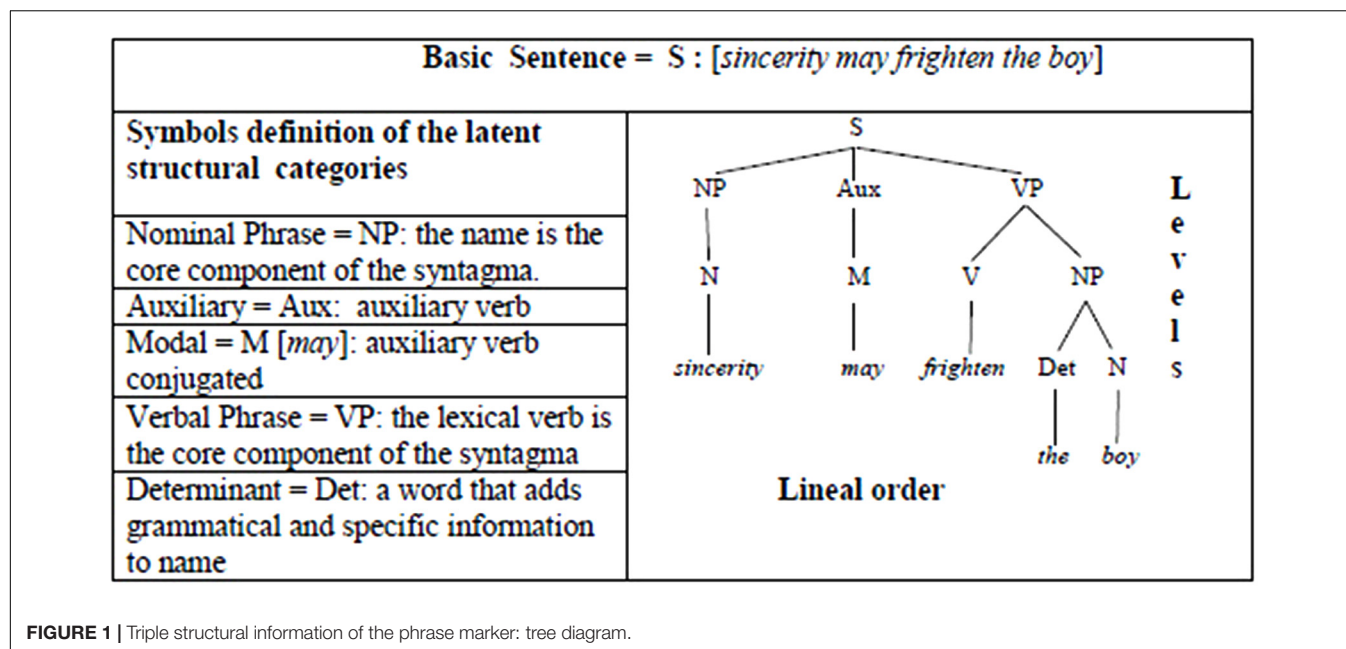
TABLE 1 | Branch generation of the terminal string for a given grammar.

Example

Given an initial string S	

Production Set P	Derivation
Rl. S→ AB	AB (Rl)
R2. A→ CD	CDB (R2)
R3. C→ c	cDB (R3)
R4. D→ d	cdB (R4)
R5. B→ b	cdb (R5)

Terminal string cdb

G = ({cdb}, {ABCD}, S, P)

Applying the sequence of production rules P the substitution of the element on the left of the rule by the one on the right is written on each derivation line. G is the grammar symbol, lower-case letters are the terminal symbols, and capital letters are the non-terminal symbols used by the generative grammar G.

situation, are the aspects that classify the idiosyncratic differences observed in the execution of specific movements.

Insofar as the aim of the morphokinetic alphabet is to symbolize elements of the visual image that have been recognized, named, and labeled using the words/concept that represent body movement (e.g., up/down of shoulder), then the structured organization of the symbols of a morphokinetic alphabet may use, as one option among other possible ones, the same way of representing structural information as phrase marker grammar. Therefore, the information recognized, named, and labeled leads us, on the one hand, to the structural categories of the morphokinetic expressions and, on the other, to the morphokinetic categories established in each case for the coding protocol. At the structural level, the nominal component is

referred to by the symbol NG (meaning "Nominal Identification Group") and the labeling component is referred to by the symbol DG (meaning "Differential Elements Group"). The breakdown of the components of these structural categories into hierarchical levels is shown in **Figure 2**.

Let us analyze a visual description with words, chosen from among the many possibilities to be found in common written texts or works of literature, and apply the proposed structural categories. The chosen text (from Poyatos, 1994, p. 140) is "*Every morning* [my father] *attended mass,* [all the time*] *with both knees on the floor, his hands together, pointing upwards at chest level, his hat on top of them*" (Alemán, GA, I, I)[3], and thus we obtain:

'*Every morning*' > DG: it is context [Det]: temporal reference;
'[my father]' > DG: it is context [Det]: personal reference with a social basis: family relationship;
'*attended mass*' > DG: it is context [Det]: religious activity: selects the repertoire of action;
'[all the time*]' > reader's inference > NG: it is time [T]: duration of the whole body figure;
'*with both knees*' = [kneeling] > NG: it is form [F]: body part + figure;
'[kneeling] *on the floor*' > NG: it is position [P]: location in physical space;
'*his hands together*' > NG: it is form [F]: part of the body + figure;
'*pointing upwards at chest level*' > NG: it is orientation [O]: vertical axis, sagittal plane;
'*his hat on top of them*' > NG: it is form [F]: supporting object;
'(...) *on top* (...)' > NG: it is orientation [O]: height scale;

---

[3] Authors' own translation. The original Spanish text is: "*Cada mañana* [mi padre] *oía su misa,* [todo el tiempo*] *sentadas ambas rodillas en el suelo, juntas las manos, levantadas del pecho arriba, el sombrero encima de ellas* (taken from *Guzmán de Alfarache* by Mateo Alemán: see Poyatos, 1994, p. 140).

FIGURE 2 | Syntactic components of the morphokinetic grammar K.

R1. K → NG DG
R2. DG → Det M
R3. NG → F S
R4. S → P O T
R5. F → *hands together*
R6. P → *kneeling*
R7. O → *pointing upwards chest*
R8. T → *all the time*
R9. Det → *every morning, attended mass*
R10. M → *hat top hands*

The "base mold" of grammar K is acceptable within the restrictive framework imposed by our interpretation of the structural components of the morphokinetic description. In this regard, the formalization of the systematization carried out here is characterized by the negligible abstraction of the categorical notions, and in concert a clear application effect on the grammatical ordering of the symbolized morphokinetic expressions. In some ways, the analogical attitude (as if) indicates that we have defined an intermediate space between the branching rules of phrase marker grammar and the rules of action.

'his hat on top of them' DG > it is modal [M]: it is form: familiar/strange emphasis.

This exercise is merely an initial approach to the adaptation of structural symbols to the morphokinetic information expressed in a word or group of words.

Rewriting Rules

Continuing with the *"phrase-structure"* analogy, let us consider a simple example of syncopated verbal-morphokinetic description (the order follows the conventional above-cited written text): "*every morning, attended mass, all the time, kneeling, hands together, pointing upwards chest, and hat on top hands*" (1). (The commas separate the word symbols; note that there are symbols composed of several words).

Representation of (1) using labeled square brackets (K is the initial symbol. *Vid* **Figure 2**):

$$^K\big[_{NG}\ \big[_F{}^{[hands\ together]}{}_F\ _S\big[_P{}^{[kneeling]}{}_P\ _O{}^{[pointing\ upwards\ chest]}{}_O\ _T{}^{[all]}{}_T\big]$$

$$_S\big]_{NG}\ _{DG}\ \big[_{Det}{}^{[every\ morning,\ attended\ mass]}{}_{Det}\ _M{}^{[hat\ on\ top\ hands]}{}_M\big]_{DG}\big]_K$$

Assuming that this formalizatiwic entities, the branch rewriting rules of the grammar K are:

TABLE 2 | Example 1: CMA codes.

Code 2_2: Entering from a sitting position on the edge of the pool without help. morphokinetic analysis of the shape of the state of the torso

F	S			Det	M
	P	O	T		
41/cox	1	2	7.00	1	4 2

Description
Standing up (1^P) + convex (411^F) + forward ($21f^O$) + duration (7^T) + first attempt (1^{Det}) + tense (4^M) + slow, deliberate waiting time (2^M)

Morphokinetic syntactic configuration Code 2_2 1-K: 1 411 21f 7 1 4 2

TABLE 3 | Example 2: CMA codes.

Code 2_2: Entering from a sitting position on the edge of the pool without help. Analysis of the sequence of positions of large head movements with or without speech

F	S			Det	M
	P	O	T		
111/lab	1	31	7.00	1	2 9

Description
Standing up (1^P) + labile head (111^F) + rotation speech (31°) + duration (7^T) + first attempt (1^{Det}) + slow, deliberate waiting time (2^M) + speech fear (9^M)

Morphokinetic syntactic configuration Code 2_2 2-K: 1 111 31 7 1 2 9

TABLE 4 | CMA selection of symbols (#) for free transcription.

Form	#	Spatial Points	#	Other Notations	#
Body parts		**Location**		**Body side**	
Joint/areas		At the center	000	Right/left	F′/F,
Neck	1	At center front	100	**Body participation**	
Shoulder	2	. . .	. . .	Bilateral movement	F*
Elbow	3	**Physical body posture**		Unilateral movement	F
Knee	4	Standing up	1	**Form timing movement**	
. . .	. . .	Sitting	2	Simultaneous compound	¨F
Head pole	11	. . .	. . .		
	12				
	. . .				
Arms pole	21	**Orientation**		Sequential compound	F̲
	. . .				
(e.g., index fingers)	(232)				
Legs pole	31	Vertical plane	11	**Height scale**	
	32	position	12	**body/vertical axis**	
	. . .		. . .	Higher. . .. . ..Lower	x5,...,x1
Trunk	41	Sagittal plane	21		
	42	position	22		
	. . .		. . .		
Figure		Horizontal plane	31	**Time-duration**	
		position	32		
			. . .		
Drawing and poetry	abbr./words	**Direction**		Non-recorded duration < 1 s	<
Technical terms	abbr./words/numbers	High/low V axis	h/w	≥1 s (decimal notation)	1.00
(e.g., labile head)	(111)	Right/left H axis	r/l	Minutes (u:v)	1:00
(e.g., convex torso). . .	(411). . .	Forward/backward S axis	f/k	**Determinants and attributes**	abbr./words/numbers
				(e.g., slow)	(2)
				(e.g., tense)	(4)
				(e.g., speech fear)	(9)
				. . .	
		Oblique directions	rf. . .	**Transcription number**	n-K

The readers should read these column by column.

CRITERIA FOR USE OF THE CMA

The proposed formal method for determining the hidden structure of "natural" morphokinetic expressions provides a syntax that orders the symbols of the morphokinetic alphabet: F ∩ S (P, O, T) ∩ Det ∩ M. As we have just seen, each element of the terminal string is a member of K in NG ∩ DG. For example, "*smoothly*" is a member of K in DG ∩ M.

The formalized syntax of the morphokinetic phrase serves as a guide not only when the movement image is observed live or through the viewing of photographs, film, or video but also when working with written texts. The grammar K channels the search for answers, and their writing, to the three basic questions: "What has moved?" "What has changed?" and "How has it changed?"

One way of optimizing the structural categories is to link them to the movement behavior criteria established in the field formats. The folder of each structural category can be displayed in as many sub-folders as necessary. Each folder contains complementary or alternative codes and, in addition, there are open options and specific catalogs (in accordance with the morphokinetic protocol created) so that the observer/analyst of movement selects, for each recording level, the codes that describe the image of the observed movement. This procedure can be carried out relatively easily using a database, such as Access.

In the example "learning to swim" (see above), one of the recording axes established in the field format is the criterion *entering the water* (code 2). Let us suppose that code 2_2, *entering from a sitting position on the edge of the pool without assistance*, requires a simplified morphokinetic description for some reason. In this case, the file of code 2_2 would contain the sub-files F, S, Det, and M, and the coding dimensions considered to be of interest, the codes, and the stipulated measurement specifications would all be displayed for each one of these sub-files. We propose two examples (**Tables 2**, **3**) of simplified morphokinetic coding [K]. See the list of symbols in **Table 4**.

Example 1: Analysis of the shape of the torso at the current moment in code 2_2 (**Table 2**).

Example 2: Analysis of the sequence of positions of large head movements with or without speech in code 2_2. See **Table 3**.

TABLE 5 | Transcription from Kendon example.

Identifiers	
1-K [transcription number]	
Action context [...^{DG}]	
Action sign [indicating to leave: Fext]	
Speech ["..., so maybe we ought + GESTURE"]	
Sequence actions order F	Movement behavior with CMA
1	F*232ext 11 ["two^{F*} extended^{Fext} index fingers^{F232} side by side^{V11}"]
2	F*21ext 21fdoor ["then extended^{Fext} both^{F*} arms^{F21} away from himself and upwards ^{O21h} in the direction of the door^{Detdoor}"]

Morphokinetic syntactic configuration 1-K: F* 232ext_21ext 21h door

When the aim is to prepare the simplified schema for the data collection work that will subsequently be carried out, the CMA functions as a free transcription system. In the context, it is necessary to economically transcribe the movement action for their analysis (**Table 4**). Free transcription also converts the kinesics present in written natural language into movement scores.

One example is the compound and sequential gesture described by Kendon (1987, p. 85) – "[action context and speech:...]. In this gesture he placed his two extended index fingers side by side and then extended both arms away from himself and upwards in the direction of the door". It is transcribed as follows (**Table 5**):

Finally, the use of general scripts (e.g., Labanotation) is compatible with our grammar. Any of these transcriptions can be converted to a decimal coding system or translated to the CMA vocabulary (alphanumeric and word symbols), achieving an accurate reproduction of the motor action, without loss of meaning.

CONCLUSION

The process of observing human movement depends on how morphokinetic changes are perceived and described. The CMA notation system simplifies, organizes, and restructures (Dittmann, 1987) the morphokinetic changes in the psychological space of the observer/analyst as distinct descriptive phrases or movement configurations. Changes in the body figure are demarcated by the variables of space and time, and the identification is completed through the inclusion of words that mark the context of activity and classify the movement's linguistic space.

Grammatical formalization is a way of forming acceptable symbol strings in accordance with the properties assigned to the syntactic component. The grammar has been developed here on the basis of phrase marker grammar. The simplest movement phrase, regardless of the size of the morphokinetic unit being considered, must be able to be analyzed as a basic expression: the visual form of the movement described provides information about, on the one hand, the perceived constraints between body, space, and time, and, on the other, the perceived connection between body, space, and time and the particular execution of the motor action.

Without doubt the most delicate question related to precision concerns the selection of movements and their description with respect to the reference framework adopted: body parts/space-time and other attributes.

Finally, the CMA may be useful for several basic reasons set out in this article: (1) it gives structure to the processes of identifying, writing, reading, rebuilding, reflecting upon, and analyzing raw data in the form of time function (video record); (2) it offers an open and flexible coding format that is compatible with the solutions offered by other notational systems for transcribing body movement; (3) it meets the frequent need to combine molar and molecular units in the same recording as if it were a zoom, in other words, without losing the unitary view of the whole body under consideration; (4) it allows the computerized management of visual notations; (5) it combines the principles of synchrony and diachrony of movement behavior, which enables advanced analytic techniques (time-series, sequential analysis, T-pattern analysis, etc.) to be applied to the matrix of reliable data; and (6) it performs an appreciable function as a transcription system in situations involving direct observation and when working with the kinesics of written texts.

The CMA is designed as a basic framework for developing specific coding schemes of body movement in a social context. In this regard, the next step will be to build a guide for recording and coding movement behaviors in each area of study. Further research is required on the applications of the CMA in order to assess its potential and scope. It is expected that new initiatives will provide additional evidence about the versatility of the system and assurances regarding the reliability of the trials carried out by different researchers who are committed to the grammatical principles of the morphokinetic observation on which the CMA is based: simplified coding, field format time recording, and syntactic rules of descriptions.

AUTHOR CONTRIBUTIONS

CI developed the project 'Common Morphokinetic Alphabet' (CMA) under supervision of MTA. Both authors reviewed the literature and discussed the references included in the proposal of a grammatical framework of body movement for observing and coding morphokinetic records in systematic observation studies. MTA incorporated the four stages that structure the process of systematic observation. CI designed the path followed in the conceptual analysis of the problem of movement notation, and made the draft of the manuscript. This manuscript is result of shared work.

FUNDING

REFERENCES

Albert, V. N. (2001). Using time-series analysis to evaluate the impact of policy initiatives in child welfare. *Eval. Program Plan* 24, 109–117. doi: 10.1016/S0149-7189(01)00002-7

Anguera, M. T. (2003). "Observational methods (General)," in *Encyclopedia of Psychological Assessment*, Vol. 2, ed. R. Fernández-Ballesteros (London: Sage), 632–637.

Anguera, M. T., Blanco, A., Losada, J. L., Ardá, T., Camerino, O., Castellano, J., et al. (2003). Match and player analysis in soccer: computer coding and analytic possibilities. *Int. J. Comp. Sci.* 2, 118–121.

Anguera, M. T., Camerino, O., Castañer, M., Sánchez-Algarra, P., and Onwuegbuzie, A. J. (2017). The specificity of observational studies in physical activity and sports sciences: moving forward in mixed methods research and proposals for achieving quantitative and qualitative symmetry. *Front. Psychol.* 8:2196. doi: 10.3389/fpsyg.2017.02196

Anguera, M. T., and Izquierdo, C. (2006). "Methodological approaches in human communication. from complexity of situation to data analysis," in *From Communication to Presence. Cognition, Emotions and Culture towards the Ultimate Communicative Experience*, eds G. Riva, M. T. Anguera, B. K. Wiederhold, and F. Mantovani (Amsterdam: IOS Press), 203–222.

Anolli, L., Duncan, S., Magnusson, M., and Riva, G. (2005). *The Hidden Structure of Social Interaction. From Genomics to Culture Patterns*. Amsterdam: IOS Press.

Archer, D. (1991). *A World of Gestures*. Berkeley: University of California Media Center.

Argyle, M. (1973). The syntax of bodily communication. *Linguist. Intern. Rev.* 112, 71–91.

Bakeman, R., and Gottman, J. M. (1986). *Observing Interaction. Introduction to Sequential Analysis*. Cambridge: Cambridge University Press.

Bakeman, R., and Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ*. Cambridge: Cambridge University Press.

Barker, L. L., and Collins, N. B. (1970). "Nonverbal and kinesic research," in *Methods of Research in Communication*, eds P. Emmert and W. D. Brooks (Boston, MA: Houghton-Mifflin), 343–372.

Bartenieff, I., and Lewis, D. (1980). *Body Movement: Coping with the Environment*. Amsterdam: Gordon and Breach.

Benesh, R., and Benesh, J. (1956). *An Introduction in Benesh Notation*. London: Adam & Charles Black.

Birdwhistell, R. (1952). *Introduction to Kinesics*. Louisville. KY: University of Louisville.

Birdwhistell, R. (1970). *Kinesics and Context. Essays on Body Motion Communication*. Philadelphia, PA: University of Pennsylvania.

Blanco-Villaseñor, A., Losada, J. L., and Anguera, M. T. (2003). Analytic techniques in observational designs in environment-behaviour relation. *Medio Ambient. Comport. Hum.* 4, 111–126.

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Cantón, A., and Hileno, R. (2016). Goal scoring in soccer: a polar coordinate analysis of motor skills used by lionel messi. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Chinellato, E., Castiello, U., and Sartori, L. (2015). Motor interference in interactive contexts. *Front. Psychol.* 6:791. doi: 10.3389/fpsyg.2015.00791

Chomsky, N. (1956). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge. MA: The MIT.

Condon, W. S., and Ogston, W. D. (1970). Method of micro-analysis of sound films of behaviour. *Behav. Res. Methods Instrum.* 2, 51–54. doi: 10.3758/BF03210994

Coster, M. (2005). *Postural Dynamics*. Available at: www.aqualide.com

Crivelli, C., Jarillo, S., and Fridlund, A. J. (2016). A multidisciplinary approach to research in small-scale societies: studying emotions and facial expressions in the field. *Front. Psychol.* 7:1073. doi: 10.3389/fpsyg.2016.01073

Dittmann, A. T. (1987). "The role of body movement in communication," in *Behaviour and Communication*, 2nd. Edn, eds A. Siegman and S. Feldstein (Hillsdale, NJ: Lawrence Erlbaum), 37–63.

Donaghy, W. (1988). "Nonverbal communication measurement," in *Measurement of Communication Behaviour*, eds P. H. Emmert and L. Barker (New York, NY: Longman), 296–332.

Efron, D. (1942). *Gesture and Environment*. New York, NY: Kings Crown.

Ekman, P. (1964). Body position, facial expression and verbal behaviour during interviews. *J. Abnorm. Soc. Psychol.* 68, 295–301. doi: 10.1037/h0040225

Ekman, P., and Friesen, W. V. (1978). *Facial Action Coding System: Full Training Materials*. Palo Alto, CA: Consulting Psychologists.

Eshkol, N. (1971). *Eshkol-Wachman Movement Notation. The Handbook*. Tel Aviv: The Movement Notation Society.

Eshkol, N., and Wachman, A. (1958). *Movement Notation*. Tel Aviv: Tel Aviv University.

Farnell, B. (1996). "Movement notation systems," in *The World's Writing Systems*, eds P. T. Daniels and W. Bright (New York, NY: Oxford University), 855–880.

Farnell, B. (1999). Moving bodies, acting selves. *Annu. Rev. Anthropol.* 28, 341–373. doi: 10.1146/annurev.anthro.28.1.341

Fink, B., Weege, B., Neave, N., Pham, M. N., and Shackelford, T. K. (2015). Integrating body movement into attractiveness research. *Front. Psychol.* 6:220. doi: 10.3389/fpsyg.2015.00220

Freedman, N. (1981). "Toward a mathematization of kinetic behaviour: a review of paul Bouissac's "La Mesure de Gestes"," in *Nonverbal Communication, Interaction, and Gesture*, ed. A. Kendon (The Hague: Mouton).

Frey, S., Hirsbrunner, H.-P., and Jorns, U. (1982). "Time-series notation: a coding principle for the unified assessment of speech and movement in communication research," in *Multimedial Communication: Semiotic Problem of its Notation*, Vol. 1, ed. E. W. B. Hess-Lüttich (Tübingen: Gunter Narr Verlag), 50–58.

Frey, S., and Pool, J. A. (1976). *A New Approach to the Analysis of Visible Behaviour*. Berne: Research Reports from the Department Psychology of the University of Berne.

Guest, A. H. (2011). *Labanotation: The System of Analyzing and Recording Movement*, 4th Edn. New York, NY: Taylor and Francis.

Hadar, U. (1994). Implications of methodological rigor in movement analysis for the study of human communication. *Behav. Brain Sci.* 17, 753–754. doi: 10.1017/S0140525X00037006

Haynal-Reymond, V., Jonsson, G. K., and Magnusson, M. (2005). "Nonverbal communication in doctor-suicidal patient interview," in *The Hidden Structure of Social Interaction. From Genomics to Culture Patterns*, eds L. Anolli, S. Duncan, M. Magnusson, and G. Riva (Amsterdam: IOS Press), 141–148.

Hirsbrunner, H.-P., Frey, S., and Crawford, R. (1987). "Movement in human interaction: description, parameter formation, and analysis," in *Nonverbal Behaviour and Communication,* 2nd Edn, eds A. W. Siegman and S. Feldstein (Hillsdale, NJ: Lawrence Erlbaum), 99–140.

Hutchinson-Guest, A. (1984). *Dance Notation: The Process of Recording Movement on Paper*. New York, NY: Dance Horizons.

Izquierdo, C. (2010). "Comunicación no verbal: el movimiento del cuerpo" [Nonverbal communication: body movement]," in *Investigar en Psicología de la Comunicación: Nuevas Perspectivas Conceptuales y Metodológicas*, eds C. Izquierdo and A. Perinat (Barcelona: Amentia), 143–176.

Izquierdo, C., and Anguera, M. T. (2001). "The role of the morphokinetic notational system in the observation of movement," in *Oralité et Gestalité. Interactions et Comportements Multimodaux Dans la Communication*, eds C. H. Cavé, I. Guaïtella, and S. Santi (Paris: L'Harmatta), 385–389.

Kendon, A. (1981). "Introduction: current issues in the study of "nonverbal communication"," in *Nonverbal Communication, Interaction, and Gesture*, ed. A. Kendon (The Hague: Mouton), 1–53. doi: 10.1515/9783110880021

Kendon, A. (1987). "On gestures: its complementary relationship with speech," in *Nonverbal Behaviour and Communication*, 2nd Edn, eds A. Siegman and S. Feldstein (Hillsdale, NJ: Lawrence Erlbaum), 65–97.

Kendon, A. (1997). Gesture. *Annu. Rev. Anthropol.* 26, 109–128. doi: 10.1146/annurev.anthro.26.1.109

Krajcsi, A., and Szabó, E. (2012). The role of number notation: sign-value notation number processing is easier than place-value. *Front. Psychol.* 3:463. doi: 10.3389/fpsyg.2012.00463

Laban, R. (1926). *Choreographie*. Löbitz: Eugen Diederichs.

Magnusson, M. S. (1996). Hidden real-time patterns in intra- and inter-individual behavior. *Eur. J. Psychol. Assess.* 12, 112–123. doi: 10.1027/1015-5759.12.2.112

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behav Res Methods Instrum. Comput.* 32, 93–110. doi: 10.3758/BF03200792

Magnusson, M. S., Burgoon, J. K., Casarrubea, M., and McNeill, D. (2015). *Discovering Hidden Temporal Patterns in Behaviour and Interactions: T-Pattern Detection and Analysis with THEME*. New York, NY: Springer.

Mead, M., and Bateson, G. (1942). *Balinese Character: A Photographic Analysis*. New York, NY: New York Academy of Sciences.

Mehrabian, A. (1969). Significance of posture and position in the communication of attitude and status relationships. *Psychol. Bull.* 71, 359–372. doi: 10.1037/h0027349

Merten, J., and Schwab, F. (2005). "Facial expression patterns in common and psychotherapeutic situations," in *The Hidden Structure of Social Interaction. From Genomics to Culture Patterns*, eds L. Anolli, S. Duncan, M. S. Magnusson, and G. Riva (Amsterdam: IOS Press), 149–158.

Neagle, R., Ng, K., and Ruddle, R. A. (2002). "Notational and 3D animation of dance movement," in *Proceedings of the International Computer Music Conference (ICMC 2002)*, Göteborg, 459–462.

Page, J. (1996). Images for understanding: movement notations and visual recordings. *Vis. Anthropol.* 8, 171–196. doi: 10.1080/08949468.1996.9966675

Perea, A., Castellano, J., Alday, L., and Hernández Mendo, A. (2012). Analysis of behaviour in sports through polar coordinate analysis with MATLAB. *Qual. Quant.* 46, 1249–1260. doi: 10.1007/s11135-011-9435-z

Portell, M., Anguera, M. T., Chacón, S., and Sanduvete, S. (2015a). Guidelines for reporting evaluations based on observational methodology (GREOM). *Psicothema* 27, 283–289.

Portell, M., Anguera, M. T., Hernández-Mendo, A., and Jonsson, G. K. (2015b). Quantifying biopsychosocial aspects in everyday contexts: an integrative methodological approach from the behavioural sciences. *Psychol. Res. Behav. Manage.* 8, 153–160. doi: 10.2147/PRBM.S82417

Poyatos, F. (1986). *New Perspectives in Nonverbal Communication: Studies in Cultural Anthropology, Social Psychology, Linguistics and Literature and Semiotics*. Oxford: Pergamon Press.

Poyatos, F. (1994). *Comunicación no verbal. I. Cultura, Leguaje y Conversación [Non-verbal Communication. I. Culture, Language and Conversation]*. Mexico: Istmo.

Sackett, G. P. (1980). "Lag sequential analysis as a data reduction technique in social interaction research," in *Exceptional Infant: Psychosocial Risks in Infant-Environment Transactions*, eds D. B. Sawin, R. C. Hawkins, L. O. Walker, and J. H. Penticuff (New York, NY: Brunner/Mazel), 300–340.

Sánchez-Algarra, P., and Anguera, M. T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: Impact of recording and coding predominating perspectives. *Qual. Quant.* 47, 1237–1257. doi: 10.1007/s11135-012-9764-6

Scheflen, A. (1964). The significance of posture in communication systems. *Psychiatry* 27, 316–331. doi: 10.1080/00332747.1964.11023403

Scheflen, A. (1972). *Body Language and Social Order. Communication as Behavioural Control*. Englewood Cliffs, NJ: Prentice-Hall.

Schegloff, E. (2000). On granularity. *Annu. Rev. Sociol.* 26, 715–720. doi: 10.1146/annurev.soc.26.1.715

Shafir, T., Tsacor, R. P., and Welch, K. B. (2016). Emotion regulation through movement unique sets of movement characteristics are associated with and enhance basic emotions. *Front. Psychol.* 6:2030. doi: 10.3389/fpsyg.2015.02030

Weick, K. E. (1968). "Systematic observational methods," in *Handbook of Social Psychology*, 2nd Edn, Vol. 2, eds G. Lindzey and E. Aronson (Reading, MA: Addison-Wesley), 357–451.

Weick, K. E. (1985). "Systematic observational methods," in *Handbook of Social Psychology*, 3rd Edn, Vol. 1, eds G. Lindzey and E. Aronson (Reading, MA: Addison-Wesley), 567–634.

Absolute and Relative Training Load and Its Relation to Fatigue in Football

Unai Zurutuza[1,2]*, Julen Castellano[1]*, Ibon Echeazarra[1] and David Casamichana[3]

[1] *Physical Education and Sport Department, Faculty of Education and Sport, University of the Basque Country, Vitoria-Gasteiz, Spain,* [2] *Physical Performance Department, Beasain, Spain,* [3] *Faculty of Physiotherapy and Speech Therapy, Gimbernat-Cantabria University School Associated with the University of Cantabria, Torrelavega, Spain*

The aim of the study was to assess the relationship of external and internal training load (TL) indicators with the objective and subjective fatigue experienced by 15 semi-professional football players, over eight complete weeks of the competition period in the 2015–2016 season, which covered microcycles from 34th to 41st. The maximum heart rate (HRmax) and maximum speed (Vmax) of all the players were previously measured in specific tests. The TL was monitored via questionnaires on rating of perceived exertion (RPE), pulsometers and GPS devices, registering the variables: total distance (TD), player load 2D (PL2D), TD at >80% of the V_{max} (TD80), TD in deceleration at < -2 m·sec-2 (TDD <-2), TD in acceleration >2 m·sec^{-2} (TDA >2), Edwards (ED), time spent at between 50 and 80% (50–80% HRmax), 80–90% (80–90% HRmax), and >90% of the HR_{max} (>90% HR_{max}), and RPE both respiratory/thoracic (RPEres) and leg/muscular (RPEmus). All the variables were analyzed taking into account both the absolute values accumulated over the week and the normalized values in relation to individual mean competition values. Neuromuscular fatigue was measured objectively using the countermovement jump test and subjectively via the *Total Quality Recovery* (TQR) scale questionnaire. Analytical correlation techniques were later applied within the general linear model. There is a correlation between the fatigue experienced by the player, assessed objectively and subjectively, and the load accumulated over the week, this being assessed in absolute and relative terms. Specifically, the load relative to competition correlated with the physical variables TD (−0.279), PL2D (−0.272), TDD < -2 (−0.294), TDA >2 (−0.309), and sRPEmus (−0.287). The variables related to heart rate produced a higher correlation with TQR. There is a correlation between objectively and subjectively assessed fatigue and the accumulated TL of a player over the week, with a higher sensitivity being shown when compared to the values related to the demands of competition. Monitoring load and assessing fatigue, we are closer to knowing what the prescription of an adequate dose of training should be in order for a player to be as fresh as possible and in top condition for a match. Normalizing training demands with respect to competition could be an appropriate strategy for individualizing player TL.

Keywords: team sports, training, physical load, physiological load, fatigue

INTRODUCTION

The main aim of training is to provide a stimulus which will optimize the player/team's performance during competition whilst minimizing the negative consequences of that training such as lack of freshness, fatigue, over-training, or injury (Gabbett et al., 2012). The load experienced by players in training and competition can provoke temporary metabolic, neuromuscular or mental fatigue (Campos and Toscano, 2014), reducing performance (Fessi et al., 2016) and increasing the possibility of injury to the player (Ehrmann et al., 2016). In fact, the inappropriate management of training loads (TDs) is emerging as one of the main risk factors in no contact injuries (Soligard et al., 2016).

However, appropriate doses of stimulus could improve performance and protect against possible injury (Gabbett et al., 2016). It is therefore vitally important for physical fitness and sports technicians to determine the optimum quantity of training required for the player to continue improving his/her fitness or to maintain it without putting at risk their freshness, and to reduce the probability of injury, with a view to the maximization of performance in competition.

By monitoring load (Akenhead et al., 2016), information can be obtained concerning the handling of its prescription to try to reduce, when appropriate, acute fatigue (thus improving freshness), so that performance does not decrease whilst avoiding placing the player at greater risk of injury (Gabbett, 2016). The search for an optimum relationship between load, fatigue-freshness and performance is no easy task, given that it concerns an individual process influenced by internal and external factors which are at times independent from the workload itself (Gabbett et al., 2012). In addition to knowing the external load placed on the players, it is necessary to discover how this affects each player (internal load) given that the same external load can have different repercussions in different players or even in the same player at different points in the season (Impellizzeri et al., 2005). Current scientific literature (Gaudino et al., 2013; Colby et al., 2014) presents different methods for controlling load levels (external and internal) in team sports players. That research used objective measurements such as GPS devices (Casamichana et al., 2013) or heart rate monitoring (HR) (Henderson et al., 2015), but also subjective measurements such as the rating of perceived exertion (RPE) (Los Arcos et al., 2014b).

As a consequence of imposed load, the performance of a player is temporarily reduced due to the fatigue which is generated. Fatigue is defined as any decrease in muscular performance associated with muscular activity (Nédélec et al., 2012a). There is currently (Gastin et al., 2013) widespread use of different methods (objective and/or subjective) to assess fatigue. To this effect, the procedure used repeatedly to objectively assess fatigue is a variety of vertical jumps, such as the countermovement jump test (CMJ) (McLean et al., 2010; Malone et al., 2015; Thorpe et al., 2015). Alternatively, subjective assessments of fatigue are done using questionnaires such as that of Hopper (Hooper and Mackinnon, 1995), variables associated to *Wellness* (Thorpe et al., 2016a), *Total Quality Recovery Scale* (TQR) (Kentta and Hassmen, 1998), *The Profile of Mood State* (McNair et al., 1971).

There is an increasing tendency to study (Thorpe et al., 2016b) the relationships between the TD borne by players and the fatigue which this produces, in an attempt to find the optimum load with which to increase physical fitness, allowing the soccer players to be fresh for the match whilst avoiding loads which by default or excess put him/her at risk of injury.

Accumulated load values are usually assessed in absolute terms (Gabbett and Ullah, 2012; Cross et al., 2016). To date, no research has normalized the TD to the mean values of the player in competition. This would allow the comparison of the TD demand placed on the player with the demands of competition, which have shown a high inter-individual variability (Schuth et al., 2016).

To that end, the aim of this research is to study the relationship between external and internal training load indicators (TL), in absolute values and relative to competition, with respect to the fatigue experienced by semi-professional football players measured using objective and subjective values. The results of this research could increase knowledge about how to manage the load imposed on players with a view to adjusting its prescription in order to optimize physical performance in competition.

MATERIALS AND METHODS

Subjects

A total of 15 semi-professional male football players (Defenders = 5, Midfielders = 8 and Forwards = 2, goalkeeper did not take part in the study) took part in the study (age = 25.2 ± 3.0 years; height = 177.8 ± 5.6 cm; weight = 76.9 ± 6.5 kg) percentage of body fat (Möhr and Johnsen, 1972) was 11.6 ± 2.7% from group IV of the third division in the Spanish League. The players did, on average, 3–4 weekly training sessions and played one official match every weekend. The Ethics Committee of research with humans (CEISH) of the University of the Basque Country (UPV/EHU) gave its institutional approval of the study. In accordance with the protocol, before taking part in the study, all the players involved signed an informed consent form. Both the participants and the team's technical body were kept informed at all times about the procedure and possible risks and benefits of the study.

Training Sessions and Competition Matches

All the training sessions and competition matches were monitored during the microcycles of the study. In total, 250 recordings were made in 20 training sessions (16.7 ± 3.6 per player) and 72 recordings from eight matches (4.9 ± 2.1 per player). The individual session or match recordings were grouped into microcycles, with a total of 69 weekly recordings (4.6 ± 1.3 per player). In order to calculate the mean value of the demands in competition, all the values were normalized to a 90 min match (mean match duration recorded per player ±SD).

Heart Rate

In all the training sessions and competition matches heart rate (HR) was recorded via a short range telemetry system (Polar

Team2 Pro System, Polar Electro Oy, Kempele, Finland). The reliability of the devices used in this study has been reported in previous studies (Macleod and Sunderland, 2012). To quantify the internal load from the HR the Edwards (1993) method was used. The Edwards method distributes the exertion of the HR in five different zones. Each zone has an established value (50–60% $HR_{max} = 1$, 60–70% $HR_{max} = 2$, 70–80% $HR_{max} = 3$, 80–90% $HR_{max} = 4$, 90–100% $HR_{max} = 5$) which are later added together.

To calculate the maximum HR for each player, a maximal progressive test was carried out on a treadmill with a HR monitor, beginning with a speed of 8 km/h^{-1} which was increased at a rate of 1 km/h^{-1} every minute until the point of physical exhaustion was reached (Graff, 2002). Furthermore, the minutes spent in each zone were taken into account in the following intensity ranges (Henderson et al., 2015): time spent between 50 and 80% of the maximum HR (50–80% HRmax), time spent between 80 and 90% of the maximum HR (80–90% HRmax); and time spent at more than 90% of the maximum HR (>90% HRmax).

Perceived Exertion Response

Once the training and/or match was finished, the players had to complete a subjective RPE. The RPE questionnaire used was a translation into Spanish of the Borg scale of 0–10 points modified by Foster (Foster et al., 2001), adapted to distinguish between perceived respiratory/thoracic exertion (RPEres) and the perceived exertion in legs/muscular (RPEmus) (Los Arcos et al., 2014b).

The players were able to respond with a plus symbol (for example 5+ means 5.5) next to the unit of assessment. The assessment was carried out 10 min after the end of the session or match (Ngo et al., 2011). Afterward, the value obtained in each of the scales was multiplied by the duration of the session or match (including warm-up and rests or pauses, but excluding cool down) to obtain the following variables (sRPEres and sRPEmus).

Physical Variables

The players' external load was monitored using GPS devices (Minimax S4, Catapult Innovations, Docklands, VIC, Australia, 2010) which function at a sampling frequency of 10 Hz and contain a 100 Hz triaxial accelerometer. The reliability and validity of the devices used in this study have been reported in previous studies (Castellano et al., 2011b; Gale-Ansodi et al., 2016). The mean (±SD) number of satellites during data collection was 12.5 (±0.6). The device was attached to the upper back of each player using a special harness. The GPS devices were activated 15 min before the start of each session or match, in accordance with the manufacturer's instructions. The data from the GPS devices was later downloaded to a PC to be analyzed using the Sprint v5.1.4 software package (Catapult Innovations, Docklands, VIC, Australia, 2010).

The following physical variables were studied: (a) TD, total distance (TD) in m; (b) TD80, distance covered at more than 80% of maximum speed (V_{max}) in m; (c) PL2D, player load 2D [in arbitrary units (AUs)]; (d) TD80%, percentage of distance covered at more than 80% of V_{max} (in %); (e) TDD < −2, TD in deceleration under −2 m/sec^{-2} (in m); and (f) TDA >2, TD in acceleration over 2 m/sec^{-2} (in m).

Assessment of Neuromuscular Fatigue

To assess neuromuscular fatigue, as in previous work (McLean et al., 2010), a test was carried out (vertical bipedal jump with countermovement and with hands on hips) using the previously validated app *My Jump v.1* (Balsalobre-Fernández et al., 2015). The protocol followed is similar to that of Malone et al. (2015). Prior to the test the players did a standard warm up including a 5 min low speed run with dynamic exercises and two 20 m progressions followed by three repeats of the jump.

The best value obtained in the 14 tests of the trial was used to calculate the maximum CMJ (CMJ_{max}) of each player. Furthermore, in each microcycle the level of absolute fatigue (FATabs) of each player was calculated by discovering the percentage of the value obtained by the player on the pre-match session day (always carried out 24 h before the next match) with respect to the CMJ_{max} value. The following formula was used to calculate the absolute fatigue value: CMJ_{pre}/CMJ_{max}. To calculate relative fatigue (FATrel) the formula CMJ_{pre}/CMJ_{post} was used. This second fatigue value was calculated in order to know specifically whether the load accumulated in the week prior to the one studied had any repercussion on the freshness or objective fatigue of the player. The coefficient of variation (CV) for each of the CMJ tests was of between 0.0 and 7.7%.

Assessment of Subjective Fatigue

The subjective questionnaire TQR scale (Kentta and Hassmen, 1998) was used as a subjective measurement to assess the fatigue suffered by the players. The questionnaire was given to the players 10 min before the start of training or pre-match warm-up. The players had to complete the TQR by answering the question "how recovered do you feel?" on a scale of 0–10, with 0 being rested and 10 extremely good recovery.

Procedure

This observational study was carried out during the competitive phase (March–April) of the 2015–2016 season during the microcycles from 34 to 41°. All training and matches were monitored via pulsometers and GPS. Two of the microcycles were excluded from the analysis as not all of the sessions were present. Before beginning the trial, the players underwent a maximal progressive resistance test on a treadmill (in laboratory) to calculate the maximum HR of each player and a 40 m speed test on the training ground whilst wearing the GPS devices. Furthermore, where higher values in peak speeds were detected, these were taken as the V_{max} of the player.

During the study period, performance in a CMJ test was recorded both on the first training day of the week (with a minimum of 48 h with respect to competition) and the last (24 h prior to the next competition). The test was always carried out indoors and the players previously familiarized with it. It was decided not to include the CMJ test for matches for two reasons – one was the lack of adequate facilities in away matches and the second was due to the difficulty in getting the players to carry out maximal tests on competition days, despite these having a low impact on fatigue. Furthermore, before beginning the first or the last training session of the

week or pre-match warm-up, the players completed the TQR questionnaire (TQRpost, TQRpre, and TQRcomp, respectively). Finally, after the training session or match they completed the RPE questionnaire.

For the correlation analysis between load and fatigue, both the absolute values accumulated through the week and the accumulated values normalized to individual competition were used. The competitive reference values were obtained from the competitions recorded during the same trial. For this, the mean values of each player in competition were used as reference.

Statistical Analysis

Starting from the relative values of the different physical, physiological and perception of exertion variables, correlation analysis techniques were implemented within the general linear model. The results are shown as mean and standard deviation (±SD). The Pearson correlation coefficient was calculated to determine whether there was a relationship, and if this was significant among the analyzed variables. To interpret the results, threshold values for the Pearson correlation coefficient used by Salaj and Marckovic (2011) were used: low ($r \leq 0.3$), moderate ($0.3 < r \leq 0.7$) to high ($r > 0.7$). The statistical analysis was conducting using SPSS v.23 (IBM, Corp., Chicago, IL, United States). Significance level was fixed at 0.05 ($p < 0.05$).

RESULTS

Table 1 shows the mean and ±SD values for each of the variables obtained by the players in the matches played during the trial. It also shows the mean and (±SD) values of the load accumulated by microcycle, normalized to the demand of competition. It can be seen that in all the analyzed variables, except that of TD80, the accumulated weekly load was higher than the mean load in competition (a value of 100% means that the demands of competition are repeated for this variable).

Table 2 shows the values obtained in the different CMJ tests carried out during the trial together with assessment of neuromuscular fatigue (FATabs and FATrel) and the subjective assessments of the state of fatigue (TQRpost, TQRpre, and TQRcomp). As it can see when TQR scale are closed to the matches the values of subjective fatigue are higher, that is, player finished the week with better wellness.

Table 3 shows the correlations between objective fatigue with the 13 load variables studied, both in absolute values and relative to the competitive demand for each player. The values show that the objectively measured fatigue (FATabs and FATrel) correlated only with some of the internal and external load variables, when the load was assessed in absolute and relative terms. Among the physical variables, only TDA > 2 is moderately correlated with the FATabs variable, while the other variables that are significantly correlated with FATabs present low correlations. There were higher correlations between relative values of the match demands than when were used absolute values. Subjective assessment of fatigue (TQRpre and TQRcomp) obtained correlations in a moderate range and were significantly high ($p < 0.01$) for the three HR variables.

DISCUSSION

The aim of this research was to study the relationship between the TD, from external and internal indicators, and the objective and subjective assessment of fatigue in semi-professional football players. This is the first piece of work which relates the load placed on semi-professional footballers in terms relative to the demands of competition with the fatigue accumulated in each microcycle of the competitive period, calculated using an objective CMJ test and also via the subjective perception of recovery quality (TQR).

The main results of the study can be summarized as follows: (1) normalizing training demands with respect to competition could be an appropriate strategy for individualizing player TD, (2) both the use of objective (from CMJ) and subjective (from TQR) fatigue indicators proved to be related to the load borne by players in the weekly microcycle.

The application of this procedure of individually monitoring TD and fatigue in players can be applied to load adjustment for each of the variables. The aim would be, on the one hand, to avoid players being fatigued on match day, and on the other hand, to increase the status of training among players, optimizing physical fitness and thus being able to give maximum performance on competition day. Previous studies (Impellizzeri et al., 2005; McLean et al., 2010; Nédélec et al., 2012b; Casamichana et al., 2013; Gastin et al., 2013; Gaudino et al., 2013; Colby et al., 2014; Los Arcos et al., 2014a; Henderson et al., 2015; Malone et al., 2015;

TABLE 1 | Mean and standard deviation values (±SD) for the profile of the external and internal demand on the players in competition and of the weekly training load (TD) in percentage values with respect to the individual demands of competition.

Variables (units)	Competition demand (units)		Weekly training demand (%)	
	Mean	**±SD**	**Mean**	**±SD**
TD (m)	9061.5	935.5	151.4%	40.4%
TD80 (m)	99.9	66.7	48.7%	39.6%
PL2D (AU)	534.7	83.2	177.7%	45.6%
TDD < −2 (m)	220.3	49.5	145.5%	43.8%
TDA >2 (m)	303.6	66.1	157.7%	42.3%
ED (AU)	345.4	31.5	144.5%	46.3%
>90% HRmax (min)	17.7	13.3	112.2%	122.6%
80–90% HRmax (min)	39.6	8.8	102.6%	38.6%
50–80% HRmax (min)	25.4	13.9	693.8%	478.0%
RPEres (AU)	6.5	1.2	177.9%	56.4%
RPEmus (AU)	6.8	1.3	167.2%	46.4%
sRPEres (AU)	581.9	103.9	165.1%	59.6%
sRPEmus (AU)	611.3	114.7	158.3%	48.3%

TD is total distance, PL2D is two dimension player load, TD80 is TD at more than 80% of the maximum speed (V_{max}), TDD < −2 is TD in deceleration below −2 m/sec^{-2}, TDA >2 is TD in acceleration above 2 m/sec^{-2}, ED is Edwards, 50–80% HRmax is the time spent between at 50–80% of the HR_{max}, 80–90% HRmax is the time spent between 80 and 90% of the HR_{max}, >90% HRmax is the time spent at more than 90% of the HR_{max}, RPEres is the perceived exertion response (respiratory/thoracic), RPEmus is the perceived exertion response (leg/muscular), sRPEres is the perceived exertion response (respiratory/thoracic) multiplied by the minutes in the session, and sRPEmus is the perceived exertion response (leg/muscular) multiplied by the minutes in the session.

Thorpe et al., 2015; Akenhead et al., 2016; Gabbett, 2016; Gabbett et al., 2016) have analyzed the load placed on players in training and matches, but none have compared the load accumulated by the players in a training microcycle normalized to the physical demands of competition, despite being a practice used by elite teams as a means of *training status* (Akenhead and Nassis, 2016).

The decision to take the physical demands of competition as an individual reference is due to the probability that similar doses of training (absolute load) do not suppose the same percentage in relation to that which competition demands of each player (% of the match). This is not only because of differences in imposed demands on the players depending on their position on the pitch (Di Salvo et al., 2007), but also due to inter-individual variations (Impellizzeri et al., 2005; Castellano and Blanco-Villaseñor, 2015) even among those playing in the same position. To consider only the demands of training in absolute values could lead to inappropriate decisions being taken in the prescription of TD, over-stimulating certain variables in some players whilst other players may not be sufficiently stimulated in relation to the values of some variables in competition.

This gives rise to a new hypothesis regarding the need to individualize the variables which can affect a player's fatigue or recovery, which will require further research. It is known that each player assimilates loads in a different way, due to his/her past and present characteristics (Impellizzeri et al., 2005), which provokes a particular state of fatigue which could be conditioned by the type of demand variable (e.g., those related to speed, acceleration/deceleration or metabolic system). That is why it is essential to individualize training as far as possible in order to strengthen collective training and thus optimize competition performance.

In order to normalize the weekly load, in this work it was decided to take as a reference the mean values of each player in competition, whilst being aware that these competition demands present a moderate-elevated variability (Castellano and Blanco-Villaseñor, 2015) in response to numerous situational variables such as place, current score or the quality of the teams (Castellano et al., 2011a). All of these could provoke a demand on the players at specific times which is higher than the estimated match average.

However, the quantity of load placed on the players should be conditioned by what they are able to assimilate, in order not to avoid over-training or increasing the probability of injury (Gabbett, 2016). To avoid unwanted negative effects from the load, it is necessary to study the player's accumulated fatigue. Special attention is currently being paid (McLean et al., 2010;

TABLE 2 | Mean and standard deviation values (±SD) of the objective and subjective fatigue variables.

	CMJpost (cm)	CMJpre (cm)	CMJmax (cm)	FATabs (%)	FATrel (%)	TQRpost (AU)	TQRpre (AU)	TQRcomp (AU)
Mean	34.8	34.7	37.4	0.93	0.99	7.34	7.44	7.91
±SD	3.6	3.6	3.4	0.04	0.05	1.80	1.26	1.25

CMJpost is countermovement jump carried out post-competition, CMJpre is countermovement jump carried out pre-competition, CMJmax is the maximum value of countermovement jump obtained in the trial, FATabs is absolute fatigue, FATrel is relative fatigue, TQRpost is the TQR questionnaire completed on the first day of training, TQRpre on the day before competition and TQRcomp done on match day.

TABLE 3 | Values of the Pearson correlations of the objective absolute (FATabs), relative (FATrel) and subjective fatigue variables on the day before (TQRpre) and on competition day (TQRcomp) in relation to the accumulated weekly values of the absolute load variables and those relative to competition.

Variables	FATabs		FATrel		TQRpre		TQRcomp	
	Rel	Abs	Rel	Abs	Rel	Abs	Rel	Abs
TD	−0.279*	−0.234	−0.233	−0.186	−0.046	−0.079	−0.190	−0.199
PL2D	−0.272*	−0.153	−0.192	−0.095	0.000	0.233	−0.144	0.015
TD80	−0.178	−0.221	−0.161	−0.277*	−0.034	0.016	0.095	0.055
TDD < −2	−0.294*	−0.278*	−0.251	−0.241	−0.029	−0.089	−0.135	−0.234
TDA >2	−0.309*	−0.283*	−0.227	−0.234	−0.036	−0.154	−0.134	−0.250
ED	−0.170	−0.138	−0.123	−0.129	0.195	0.215	0.075	0.109
>90% HRmax	0.074	0.042	0.079	−0.067	0.233	0.436**	0.193	0.288*
80–90% HRmax	0.031	−0.072	0.007	−0.033	0.350**	0.081	0.337*	0.164
50–80% HRmax	0.061	−0.302*	−0.090	−0.219	0.114	−0.186	0.146	−0.268*
RPEres	−0.039	−0.085	−0.045	0.096	0.022	−0.039	0.051	0.101
RPEmus	−0.186	−0.176	−0.086	−0.057	−0.021	−0.164	−0.052	−0.163
sRPEres	−0.166	−0.214	−0.141	−0.012	−0.037	−0.096	−0.032	0.021
sRPEmus	−0.287*	−0.283*	−0.174	−0.146	−0.086	−0.215	−0.120	−0.226

*TD is total distance, PL2D is player load 2D, TD80 is TD at more than 80% of maximum speed (V_{max}), TDD < −2 is TD in deceleration below −2 m/sec^{-2}, TDA >2 is TD in acceleration above 2 m/sec^{-2}, ED is Edwards, >90% HRmax is the time spent at more than 90% of the HR_{max}, 80–90% HRmax is the time spent between 80 and 90% of the HR_{max}, 50–80% HRmax is the time spent between 50 and 80% of the HR_{max}, RPEres is the perceived exertion response (respiratory/thoracic), RPEmus is the perceived exertion response (leg/muscular), sRPEres is the perceived exertion response (respiratory/thoracic) multiplied by the minutes in the session, and sRPEmus is the perceived exertion response (leg/muscular) multiplied by the minutes in the session. *p < 0.05 and **p < 0.01 (bilateral).*

Gabbett and Ullah, 2012) to the assessment of neuromuscular fatigue via a simple and objective vertical jump test (i.e., CMJ).

Assessment of neuromuscular fatigue from CMJ may not be sensitive when the aim is to compare acute fatigue from a football training session (difference in a test of jump over height reached in pre with respect to post-training) (Malone et al., 2015) perhaps because football training usually involves multidimensional demands (Gaudino et al., 2015). In our research, the neuromuscular fatigue measured with CMJ wassensitive to the different percentages of load borne by the players during the training week.

Along the same lines, Gathercole et al. (2015) in his study found significant correlations between different microcycles for the CMJ variable which measures neuromuscular fatigue. Although more research is needed, assessment of neuromuscular fatigue via CMJ, or other tests, could be a useful tool for adjusting optimum TD, by which the technical team could ensure that their players are fresh when they come to compete.

This innovative study has also analyzed variables connected to accelerations and decelerations in which correlations with fatigue have also been found. To be more specific, the variables related to the neuromuscular dimension (PL and accelerations and decelerations) showed a greater sensitivity (correlation) with this objective jump test.

The use of questionnaires such as the TQR has allowed us to discover the player's degree of subjective recovery at the end of the microcycle, which provides information on the fatigue generated in the player during the week. Despite there being practically no differences in the TQRpost with respect to the end of the week (TQRpre), it is worth pointing out that in the variables related to HR, and therefore the cardiovascular energy system, it was the TQR questionnaire which showed a higher sensitivity to the changes.

This suggests that it may be interesting to consider that just as different dimensions of TD are monitored, it could be relevant to have various tools available with which to assess the state of player fatigue or recovery, which would deal with different dimensions of the fatigue generated.

One of the limitations of the study was the relatively low number of recordings of training session load and match load. A higher number of recordings would have provided more information about each player and therefore established particular load-fatigue relationships for each one, in the search for the adequate dose and typology of TD for each player.

We should also highlight that using a higher number of players in the study, apart from providing information about their physical condition, would have shown how far players with different ability or fitness present a particular load-recovery relationship. This would thus allow attention to be paid to the capacity for bearing load and/or being fresher in competition or having better recovery after the match (Rabbani and Buchheit, 2016).

Finally, it should also be underlined that the reduced sample group did not allow the incorporation of variables with which to differentiate the players who played a match in the week prior to the studied microcycle. It is probable that different states of recovery at the beginning of the week could have conditioned some of the results of this work.

Practical Applications

With the information obtained from the monitoring of TD and assessment of fatigue (objective and subjective), we are closer to knowing what the prescription of an adequate dose of training should be in order for a player to be as fresh as possible and in top condition for a match.

CONCLUSION

The main conclusion of this study is that in those microcycles where the players accumulated a greater TD or high values in the load indicators normalized to those demanded in competition, the players showed a higher level of neuromuscular fatigue, measured with CMJ. However, the players were able to recover practically the same CMJ values (measured with the FATrel) as at the beginning of the week prior to competition. This research provides a better understanding of the load-fatigue relationship with respect to competition demands. Information about external objective load (distance, speed, and acceleration/deceleration), internal objective load (HR) and internal subjective load (RPE) on the one hand, and the objective (CMJ) and subjective (TQR) indicators for fatigue assessment on the other, can help trainers to better understand and adequately manage training status and player freshness throughout the training process. Finally, it would be necessary to know whether the load borne by the players in the weekly training process maintains or improves their fitness, and thus discover whether management of the load-fatigue binomial produces an improvement in the players' physical performance.

AUTHOR CONTRIBUTIONS

UZ: design, data getting, analysis, redaction; JC: design, analysis, redaction; IE: data getting, redaction; DC: design, redaction.

REFERENCES

Akenhead, R., Harley, J. A., and Tweddle, S. P. (2016). Examining the external training load of an English Premier League football team with special reference to acceleration. *J. Strength Cond. Res.* 30, 2424–2432. doi: 10.1519/JSC. 0000000000001343

Akenhead, R., and Nassis, G. P. (2016). Training load and player monitoring in high-level football: current practice and perceptions. *Int. J. Sports Physiol. Perform.* 11, 587–593. doi: 10.1123/ijspp.2015-0331

Balsalobre-Fernández, C., Glaister, M., and Lockey, R. A. (2015). The validity and reliability of an iPhone app for measuring vertical jump performance. *J. Sports Sci.* 33, 1574–1579. doi: 10.1080/02640414.2014.996184

Campos, M. A., and Toscano, F. J. (2014). Monitorización de la carga de entrenamiento, la condición física, la fatiga y el rendimiento durante el microciclo competitivo en fútbol. *Fútbolpf* 12, 23–36.

Casamichana, D., Castellano, J., Calleja-Gonzalez, J., San Roman, J., and Castagna, C. (2013). Relationship between indicators of training load in soccer players. *J. Strength Cond. Res.* 27, 369–374. doi: 10.1519/JSC.0b013e3182548af1

Castellano, J., and Blanco-Villaseñor, A. (2015). Análisis de la variabilidad de desplazamiento de futbolistas de élite durante una temporada competitiva a partir de un modelo mixto multivariable. *Cuad. Psicol. Deporte* 15, 161–168. doi: 10.4321/S1578-84232015000100016

Castellano, J., Blanco-Villaseñor, A., and Álvarez, D. (2011a). Contextual variables and time-motion analysis in soccer. *Int. J. Sports Med.* 32, 415–421. doi: 10. 1055/s-0031-1271771

Castellano, J., Casamichana, D., Calleja-González, J., San Román, J., and Ostojic, S. (2011b). Reliability and accuracy of 10 Hz GPS devices for short-distance exercise. *J. Sports Sci. Med.* 10, 233–234.

Colby, M. J., Dawson, B., Heasman, J., Rogalski, B., and Gabbett, T. J. (2014). Accelerometer and gps-derived running loads and injury risk in elite australian footballers. *J. Strength Cond. Res.* 28, 2244–2252. doi: 10.1519/JSC.0000000000000362

Cross, M. J., Williams, S., Trewartha, G., Kemp, S. P., and Stokes, K. A. (2016). The influence of in-season training loads on injury risk in professional rugby union. *Int. J. Sports Physiol. Perform.* 11, 350–355. doi: 10.1123/ijspp.2015-0187

Di Salvo, V., Baron, R., Tschan, H., Calderon, F. J., Bachl, N., and Pigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. *Sports Med.* 28, 222–227. doi: 10.1055/s-2006-924294

Edwards, S. (1993). "High performance training and racing," in *The Heart Rate Monitor Book*, 8th Edn, ed. S. Edwards (Sacramento, CA: Feet Fleet Press), 113–123.

Ehrmann, F. E., Duncan, C. S., Sindhusake, D., Franzsen, W. N., and Greene, D. A. (2016). GPS and injury prevention in professional soccer. *J. Strength Cond. Res.* 30, 360–367. doi: 10.1519/JSC.0000000000001093

Fessi, M. S., Zarrouk, N., Di Salvo, V., Filetti, C., Barker, A. R., and Moalla, W. (2016). Effects of tapering on physical match activities in professional soccer players. *J. Sports Sci.* 34, 2189–2194. doi: 10.1080/02640414.2016.1171891

Foster, C., Florhaug, J. A., Franklin, J., Gottschall, L., Hrovatin, L. A., Parker, S., et al. (2001). A new approach to monitoring exercise training. *J. Strength Cond. Res.* 15, 109–115. doi: 10.1519/00124278-200102000-00019

Gabbett, T. J. (2016). The training-injury prevention paradox: should athletes be training smarter and harder? *Br. J. Sports Med.* 50, 273–280. doi: 10.1136/bjsports-2015-095788

Gabbett, T. J., Hulin, B. T., Blanch, P., and Whiteley, R. (2016). High training workloads alone do not cause sports injuries: how you get there is the real issue. *Br. J. Sports Med.* 50, 444–445. doi: 10.1136/bjsports-2015-095567

Gabbett, T. J., Jenkins, D. G., and Abernethy, B. (2012). Physical demands of professional rugby league training and competition using microtechnology. *J. Sci. Med. Sport.* 15, 80–86. doi: 10.1016/j.jsams.2011.07.004

Gabbett, T. J., and Ullah, S. (2012). Relationship between running loads and soft-tissue injury in elite team sport athletes. *J. Strength Cond. Res.* 26, 953–960. doi: 10.1519/JSC.0b013e3182302023

Gale-Ansodi, C., Langarika-Rocafort, A., Usabiaga, O., and Castellano, J. (2016). New variables and new agreements between 10Hz global positioning system devices in tennis drills. *J. Sport. Eng. Tech.* doi: 10.1177/1754337115622867 [Epub ahead of print].

Gastin, P. B., Meyer, D., and Robinson, D. (2013). Perceptions of wellness to monitor adaptive responses to training and competition in elite

Australian football. *J. Strength Cond. Res.* 27, 2518–2526. doi: 10.1519/JSC.0b013e31827fd600

Gathercole, R., Sporer, B., and Stellingwerff, T. (2015). Countermovement jump performance with increased training loads in elite female rugby athletes. *Int. J. Sports Med.* 36, 722–728. doi: 10.1055/s-0035-1547262

Gaudino, P., Iaia, F. M., Alberti, G., Strudwick, A. J., Atkinson, G., and Gregson, W. (2013). Monitoring training in elite soccer players: systematic bias between running speed and metabolic power data. *Int. J. Sports Med.* 34, 963–968. doi: 10.1055/s-0033-1337943

Gaudino, P., Iaia, F. M., Strudwick, A. J., Hawkins, R. D., Alberti, G., Atkinson, G., et al. (2015). Factors influencing perception of effort (session-RPE) during elite soccer training. *Int. J. Sports Physiol. Perform.* 10, 860–864. doi: 10.1123/ijspp.2014-0518

Graff, K.-H. (2002). "Ergometrische untersuchungen," in *Sportärztliche Untersuchung und Beratung*, eds D. Clasing and I. Siegfried (Balingen: Spitta Verlag GmbH & Co. KG), 38–46.

Henderson, B., Cook, J., Kidgell, D. J., and Gastin, P. B. (2015). Game and training load differences in elite junior Australian football. *J. Sports Sci. Med.* 14, 494–500.

Hooper, S. L., and Mackinnon, L. T. (1995). Monitoring overtraining in athletes. *Sports Med.* 20, 321–327. doi: 10.2165/00007256-199520050-00003

Impellizzeri, F. M., Rampinini, E., and Marcora, S. M. (2005). Physiological assessment of aerobic training in soccer. *J. Sports Sci.* 23, 583–592. doi: 10.1080/02640410400021278

Kentta, G., and Hassmen, P. (1998). Overtraining and recovery. A conceptual model. *Sports Med.* 26, 1–16. doi: 10.2165/00007256-199826010-00001

Los Arcos, A., Martínez-Santos, R., Yanci, J., Martín, J., and Castagna, C. (2014a). Variability of objective and subjective intensities during ball drills in youth soccer players. *J. Strength Cond. Res.* 28, 752–757. doi: 10.1519/JSC.0b013e3182a47f0b

Los Arcos, A., Yanci, J., Mendiguchia, J., and Gorostiaga, E. M. (2014b). Rating of muscular and respiratory perceived exertion in professional soccer players. *J. Strength Cond. Res.* 28, 3280–3288. doi: 10.1519/JSC.0000000000000540

Macleod, H., and Sunderland, C. (2012). Reliability of the physiological and metabolic responses to a field hockey specific treadmill protocol for elite female players. *J. Sports. Med. Phys. Fitness* 52, 351–358.

Malone, J. J., Murtagh, C. F., Morgans, R., Burgess, D. J., Morton, J. P., and Drust, B. (2015). Countermovement jump performance is not affected during an in-season training microcycle in elite youth soccer players. *J. Strength Cond. Res.* 29, 752–757. doi: 10.1519/JSC.0000000000000701

McLean, B. D., Coutts, A. J., Kelly, V., McGuigan, M. R., and Cormack, S. J. (2010). Neuromuscular, endocrine, and perceptual fatigue responses during different length between-match microcycles in professional rugby league players. *Int. J. Sports Physiol. Perform.* 5, 367–383. doi: 10.1123/ijspp.5.3.367

McNair, P., Lorr, M., and Drappelman, L. (1971). *Manual for the Profile of Mood States*. San Diego, CA: Education and Industrial Testing Services.

Möhr, M., and Johnsen, D. (1972). Tables for evaluation of body weight of adult men and women by their optimal weight. *Z. Arztl. Fortbild. (Jena)* 66, 1052–1064.

Nédélec, M., McCall, A., Carling, C., Legall, F., Berthoin, S., and Dupont, G. (2012a). Recovery in soccer: part I - post-match fatigue and time course of recovery. *Sports Med.* 42, 997–1015. doi: 10.2165/11635270-000000000-00000

Nédélec, M., McCall, A., Carling, C., Legall, F., Berthoin, S., and Dupont, G. (2012b). Recovery in soccer: part II - recovery strategies. *Sports Med.* 43, 9–22. doi: 10.1007/s40279-012-0002-0

Ngo, J. K., Tsui, M. C., Smith, A. W., Carling, C., Chan, G. S., and Wong, D. P. (2011). The effects of man-marking on work intensity in small-sided soccer games. *J. Sports Sci. Med.* 11, 109–114.

Rabbani, A., and Buchheit, M. (2016). Ground travel-induced impairment in wellness is associated with fitness and travel distance in young soccer players. *Kinesiology* 48, 200–206.

Salaj, S., and Marckovic, G. (2011). Specificity of jumping, sprinting, and quick change of direction motor abilities. *J. Strength Cond. Res.* 25, 1249–1255. doi: 10.1519/JSC.0b013e3181da77df

Schuth, G., Carr, G., Barnes, C., Carling, C., and Bradley, P. S. (2016). Positional interchanges influence the physical and technical match performance variables of elite soccer players. *J. Sports Sci.* 34, 501–508. doi: 10.1080/02640414.2015.1127402

Soligard, T., Schwellnus, M., Alonso, J.-M., Bahr, R., Clarsen, B., Dijkstra, P., et al. (2016). How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *Br. J. Sports Med.* 50, 1030–1041. doi: 10.1136/bjsports-2016-096581

Thorpe, R. T., Strudwick, A. J., Buchheit, M., Atkinson, G., Drust, B., and Gregson, W. (2015). Monitoring fatigue during the in-season competitive phase in elite soccer players. *Int. J. Sports Physiol. Perform.* 10, 958–964. doi: 10.1123/ ijspp.2015-0004

Thorpe, R. T., Strudwick, A. J., Buchheit, M., Atkinson, G., Drust, B., and Gregson, W. (2016a). The influence of changes in acute training load on daily sensitivity of morning-measured fatigue variables in elite soccer players. *Int. J. Sports Physiol. Perform.* 5, 1–23. doi: 10.1123/ijspp.2016-0433

Thorpe, R. T., Strudwick, A. J., Buchheit, M., Atkinson, G., Drust, B., and Gregson, W. (2016b). Tracking morning fatigue status across in-season training

weeks in elite soccer players. *Int. J. Sports Physiol. Perform.* 11, 947–952. doi: 10.1123/ijspp.2015-0490

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

How Game Location Affects Soccer Performance: T-Pattern Analysis of Attack Actions in Home and Away Matches

Barbara Diana[1], Valentino Zurloni[1]*, Massimiliano Elia[1], Cesare M. Cavalera[2], Gudberg K. Jonsson[3] and M. Teresa Anguera[4]

[1] Human Sciences for Education Department, University of Milano-Bicocca, Milan, Italy, [2] Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy, [3] Human Behavior Laboratory, University of Iceland, Reykjavik, Iceland, [4] Faculty of Psychology, University of Barcelona, Barcelona, Spain

The influence of game location on performance has been widely examined in sport contexts. Concerning soccer, game-location affects positively the secondary and tertiary level of performance; however, there are fewer evidences about its effect on game structure (primary level of performance). This study aimed to detect the effect of game location on a primary level of performance in soccer. In particular, the objective was to reveal the hidden structures underlying the attack actions, in both home and away matches played by a top club (Serie A 2012/2013—First Leg). The methodological approach was based on systematic observation, supported by digital recordings and T-pattern analysis. Data were analyzed with THEME 6.0 software. A quantitative analysis, with nonparametric Mann–Whitney test and descriptive statistics, was carried out to test the hypotheses. A qualitative analysis on complex patterns was performed to get in-depth information on the game structure. This study showed that game tactics were significantly different, with home matches characterized by a more structured and varied game than away matches. In particular, a higher number of different patterns, with a higher level of complexity and including more unique behaviors was detected in home matches than in the away ones. No significant differences were found in the number of events coded per game between the two conditions. THEME software, and the corresponding T-pattern detection algorithm, enhance research opportunities by going further than frequency-based analyses, making this method an effective tool in supporting sport performance analysis and training.

Keywords: analysis of observational data, T-Patterns, sport performance analysis, game location, soccer

INTRODUCTION

The influence of game location on performance has been investigated in sport contexts for more than 30 years. Courneya and Carron (1992) defined the so-called "home advantage" as "the term used to describe the consistent finding that home teams in sport competitions win over 50% of the games played under a balanced home and away schedule" (p. 13). Research has pointed out how athletes and teams perform significantly better when competing at home (Allen and Jones, 2014; Sarmento et al., 2014). While it is true that home advantage does not play the same role in all sports

(Jones, 2013), there are no sports where athletes or teams perform better when playing away from their home venue (Courneya and Carron, 1992; Nevill and Holder, 1999; Jones et al., 2015).

Comprehensive models have been developed to guide the understanding of the home-advantage phenomenon. An outline on recent research (Allen and Jones, 2014) identified three models which take different positions on the explanation of the home advantage phenomenon: the standard model (Courneya and Carron, 1992; Carron et al., 2005), the territoriality model (Neave and Wolfson, 2003), and the "home disadvantage" model (Wallace et al., 2005).

Relevant to this work is the operationalization of the link between outcomes and performance, offered by Courneya and Carron (1992) within their standard model. We will refer to this model when speaking about three performance levels. The primary one represents the fundamental skill execution (e.g., free throw percentage in basketball, penalties per game in soccer); the secondary is the intermediate or scoring aspect of performance (e.g., points scored in basketball); and the tertiary corresponds to the traditional outcome measure (e.g., win—loss ratio).

Concerning soccer, game-location affects positively the secondary and, more often, the tertiary level of performance, representing an important factor in determining the result of a game (e.g., Pollard, 1986, 2006; Brown et al., 2002; Wolfson et al., 2005; Pollard and Gómez, 2013).

However, there are fewer evidences about the effect of game location on game structure (primary level of performance). Most of the studies analyzed professional British soccer teams. Sasaki et al. (1999) found that the 1st division team they studied from the 1996 to 1997 season performed a greater number of goal attempts, shots on target, shots blocked, shots wide, and successful crosses during home matches. Tucker et al. (2005) analyzed the matches of a professional team and found significant differences in the frequency of corner, crosses, dribbles, passes, shots (more in home matches), clearances, goal kicks, gains of possession, and losses of control (more in away matches). However, Taylor et al. (2008) showed that the outcome of most behaviors of the professional team they studied were not influenced by game location.

Other studies have tended to aggregate performance of different teams during analysis. Carmichael and Thomas (2005) revealed that in the Premier League home teams have significantly higher performance measures for attack indicators, such as shots and successful passes in the scoring zone, while away teams committed significantly more fouls and suffered more yellow and red cards. Lago-Peñas and Lago-Ballesteros (2011) analyzed 380 games of the Spanish professional soccer league, focusing on the effects of game location and team quality in determining technical and tactical performances, and showed that home team have significantly higher means for goal scored, total shots, shots on goal, attacking moves, box moves, crosses, offsides committed, assists, passes made, successful passes, dribbles made, successful dribbles, ball possession, and gains of possession, while visiting teams presented higher means for losses of possession and yellow cards. However, Seçkin and Pollard (2008) in their analysis of 301 matches during the season 2005–2006 in the Turkish Super League showed that the success

rates for shots, fouls and disciplinary cards do not differ between home and away teams.

The contradictory findings showed in these studies may be due to the fact that the authors examined the effects of match location on single indicators of performance rather than on patterns of play. Recent studies in sport contexts have underlined the importance of identifying visual undetectable structural regularities in order to better assess the complex reality they refer to (Anguera et al., 2013; Zurloni et al., 2014a; Cavalera et al., 2015).

With our study, we propose an innovative data analysis technique for the analysis of performance on a primary level. It involves the detection of temporal patterns (T-patterns), to reveal hidden yet stable structures that underlie the interactive situations during matches. Therefore, detecting hidden patterns could help the coach to better predict both the performer's behavior and the opponent's one thanks to an integrated system that allows for an increased depth of analysis (Zurloni et al., 2014a).

This methodology is based on systematic observation, allowing the analysis of all the relevant information about any aspect related to any interactions linked with primary factors. The same single behaviors, which appear with the same frequency, can combine with each other to form different patterns of play; t-pattern analysis allows to detect these patterns, surpassing the results obtained by exclusively considering single indexes of behavior (such as number of crosses, or yellow cards).

Temporal pattern analysis (Magnusson, 2000) has been applied to a great number of research experiments in very different fields. Patterns have been used to describe, interpret and understand phenomena such as deceptive communication (Anolli and Zurloni, 2009; Zurloni et al., 2013, 2016; Diana et al., 2015), animal and human behavior (Casarrubea et al., 2015a,b) a wide variety of observational and sports studies, such as analysis of soccer team play (Camerino et al., 2012; Castañer et al., 2016), deception detection in doping cases (Zurloni et al., 2014b), motor skill responses in body movement and dance (Castañer et al., 2009), effectiveness of offensive plays in basketball (Remmert, 2003; Fernández et al., 2009) and futsal (Sarmento et al., 2015) or tactics employed by runners (Aragón et al., 2015).

Only few studies have recently applied T-pattern methodology to soccer. In a preliminary investigation, Jonsson et al. (2010) have examined T-pattern in five Icelandic and nine international soccer matches and showed a correlation between the number of patterns identified in each match and the coaches' ratings of team performance. Moreover, data showed a more defined temporal structure in international matches than in national ones, suggesting that international soccer matches are characterized by the presence of a more structured game. In another study, thirteen national and seven international soccer matches were coded using T-pattern analysis, confirming that the players' behavior is more synchronized than the human eye can detect and suggesting that high levels of synchrony are correlated with a good evaluation of performance by professional coaches (Jonsson et al., 2003). Camerino et al. (2012) used T-pattern analysis in order to analyze five National League (Liga) matches and five Champions' League matches from the 2000 to 2001 season of

FC Barcelona. T-patterns detected revealed regularities in the playing styles of the observed team, including ball possession and ball position patterns during the attacking actions. Zurloni et al. (2014a) compared the T-patterns of attack actions in the won matches and in the lost ones played by a top club of the Italian National League Championship (Serie A) over the 2012–2013 season. The number of pattern occurrences and the number of different T-patterns detected was greater for lost matches and lower for the won matches, whereas the number of events coded was similar.

Rather than focusing on single indicators of performance, this paper aims to detect the effect of game location on the structure of play (primary level of performance) analyzing how single behaviors can combine to form patterns of play. Specifically, we will focus on attack actions, comparing T-patterns detected in home and away matches. Within the T-pattern approach, we define diversity of patterns as the number of unique behaviors included in patterns and pattern complexity as the synthesis of its length (the number of events that composes a pattern) and the number of levels (the hierarchical structure of a pattern). Given that, our hypotheses are:

H1: higher number of different patterns in home than in away matches;
H2: higher patterns' diversity in home than in away matches;
H3: higher complexity (more length and levels) of patterns in home than in away matches.

METHODS

We employed a systematic observation (Anguera, 1979; Anguera et al., 2011; Portell et al., 2015a) combined with recent technology, bringing great advantages in terms of recording quality, measurement of time, and capture of co-occurrences or diachrony (Borrie et al., 2002). We designed an *ad hoc* observation instrument, and observation was active, methodologically rigorous, non-participative, and characterized by total perceptivity. The systematic observation became more and more widespread within sport research (Lapresa et al., 2013a,b), because of its high flexibility and adaptability (Sánchez-Algarra and Anguera, 2013; Portell et al., 2015b).

The methods of analyzing performance in the game of soccer have evolved from the simple use of hand notation tracking of players' movements on scale plans of pitches to the current utilization of digital video recordings and computerized analyses.

Design

We consider three main criteria to give a taxonomic definition of the observational design, as applied to our study (Anguera et al., 2011). It was nomothetic (as opposed to idiographic, which refers to the number of subjects observed), since we observed different matches; punctual (as opposed to continuous, referring to the number of observations conducted on the same subject); multidimensional (as opposed to unidimensional, referring to the number of dimensions/criteria, which are in correspondence with observation instrument).

Using an N/P/M (nomothetic, punctual, multidimensional) design is the reason behind decisions made regarding the structure of the observation instrument, the type of data, data quality control, and data analysis.

Participants

We analyzed all games played by a top club during the first leg (19 matches) of the Italian National League Championship (Serie A), over the 2012–2013 season. **Table 1** reports half-time and full-time results for comparison (this data was also included as mixed criteria in the coding instrument—see Section Instruments).

This study has been approved by the Bioethic Committee of the University of Barcelona (Institutional Review Board IRB00003099).

Instruments

Coding Instrument.

LINCE, v. 1.2.1 is a freely available software program (Gabín et al., 2012) that can be loaded [www.observesport.com and/or www.menpas.com] with purpose-designed observation instruments for the systematic recording and coding of events. The time of occurrence and duration of events (in seconds or frames) are automatically registered. The program also incorporates a data quality control tool and allows datasets and results to be exported in different formats. LINCE software operates on fixed, mixed and changing criteria of the observation, therefore making it a consistent tool for the observational design. LINCE was used to record and code each of the 19 games, and also to check the quality of the data.

TABLE 1 | Half-time and full-time results for the observed matches.

Observed team home/away	Half-time result (observed team first)	Full-time result (observed team first)	Position in table (higher/lower than rival team)	Fixture
Away	1-1	1-2 (loss)	*	1
Home	2-1	3-1 (win)	Lower	2
Away	0-0	2-2 (tie)	Lower	3
Home	0-1	1-2 (loss)	Lower	4
Away	1-1	3-3 (tie)	Higher	5
Home	0-0	1-0 (win)	Higher	6
Away	1-1	2-3 (loss)	Lower	7
Home	1-0	1-0 (win)	Lower	8
Away	0-2	2-3 (loss)	Higher	9
Home	0-0	1-1 (tie)	Lower	10
Home	0-1	0-2 (loss)	Lower	11
Away	0-0	0-0 (tie)	Higher	12
Home	1-1	1-1 (tie)	Lower	13
Away	1-1	3-1 (win)	Higher	14
Away	1-1	2-2 (tie)	Higher	15
Home	1-1	2-2 (tie)	Lower	16
Away	0-0	0-1 (loss)	Lower	17
Home	1-0	3-0 (win)	Higher	18
Away	2-3	3-4 (loss)	Higher	19

*, N/A.

Observation Instrument

The observation instrument combines field format and category systems (**Figure 1**). The field format system comprised different dimensions/criteria, each of which formed the basis for an exhaustive and mutually exclusive category system. The fixed criteria are entered at the beginning of the match and are independent from the dynamic of play, while the mixed criteria apply every time there is a change in the score, number of players and between the first and the second half of the match. The changing criteria are coded throughout the whole match (i.e., passing, lateral position, shot, recovery). Each of these criteria gives rise to respective category systems that fulfill the conditions of exhaustiveness and mutual exclusivity (E/ME).

The dimensions/criteria considered in the present observation instrument correspond to the following criteria: lateral position, zone (**Figure 2**), lateral passing, zone passing, recovery and loss, ball out of play.

Procedure

In this study, we decided to focus on attack actions mainly because of shooting restrictions, since our video data were obtained from the TV recordings of the matches.

We defined an attack action as an action that brings the ball in the ultra-offensive zone and can end with: Goal (G), Non-Goal (NG), and Permanent Loss (PL) (see **Figure 1** for a complete description of events coded); this includes penalties, corner kicks, free kicks in ultra-offensive zone, and throw-ins in the ultra-offensive zone. Each attack action was coded according to two possibilities: (1) starting from the first pass that crosses the offensive line; (2) the same was coded when the ball got to the ultra-offensive zone thanks to a free kick or a throw-in conceded to the observed team without their active effort in bringing the ball to the zone. Therefore, actions starting with a pass were not interrupted in coding by temporary pauses (such as throw-ins, free kicks, or Temporary Losses, defined as a non-possession of the ball by the observed team for no longer than 2 passes completed by the rival team) as long as the interruptions let the action proceed [e.g., during the attack action the ball is taken by a defense player, who passes it on to a different player from his team who then loses the ball to the observed team (temporary loss case)]. Permanent Loss, Goals, and Non-Goals (shots on or off-target, or deviated/saved by rival team) work as interruption lines in the attack actions coding. These interruption lines, however, do not interfere with the subsequent pattern analysis (see Section Data Analysis), which considers the coded match as a whole and will take into account time passed between events, its main value being that of allowing for similar patterns to be found when time "stretches" are insignificantly different (on the basis of pattern search parameters).

Two observers used LINCE software (Gabín et al., 2012) to code the games selected. The same software calculated Cohen's kappa coefficient (Cohen, 1960) for all the criteria, by comparing two registered data files (~5% of all the collected data) related to the same match. The values ranged between 0.75 and 0.85, which provides a satisfactory guarantee of data quality. However, when particular disagreements were identified, the specific cases were discussed and agreed on by the two coders before moving on to the complete analysis.

Data Analysis

Datasets were analyzed with THEME 6.0 (http://patternvision.com/). This software detects the temporal structure of data sets, revealing repeated patterns (T-patterns) that regularly or irregularly occur within a period of observation. A T-pattern is essentially a combination of events where the events occur in the same order, with the consecutive time distances between consecutive pattern components remaining relatively invariant, regardless of the occurrence of any unrelated event in between them (Magnusson, 2005). THEME software allows the detection of repeated temporal patterns even when multiple unrelated events occur in between components of the patterns.

A quantitative analysis, with nonparametric and descriptive statistics, was carried out to test the hypotheses. A qualitative analysis on more complex patterns was performed to get in-depth information on the game structure expressed by the team. We chose to consider the more complex patterns because they represent the highest level of organization expressed by the team in the two conditions.

RESULTS

Of the 19 matches played, the observed team collected 30 points (3 points per victory, 1 per tie, and 0 for losses), with 9 wins, 7 losses, and 3 ties. Six wins out of 9 were obtained in home matches, highlighting a home advantage for the observed team at tertiary level of performance (Courneya and Carron, 1992).

Comparison between Home and Away Matches

We performed descriptive statistics as a preliminary analysis. Extreme values identified 5 matches as potential outliers. We decided to exclude them from next analysis. According to sample size and distribution, we used the non-parametrical Mann–Whitney test to assess differences between home and away matches. Here follow the results (**Table 2**) for each parameter considered in our hypotheses: a higher number of different patterns was detected in home matches ($Mdn = 127$) than in the away ones ($Mdn = 42$), $U = 1$, $Z = 3.003$, $p = 0.001$, $r = 0.80$. Home matches included more unique behaviors in their patterns ($Mdn = 31$) than away matches ($Mdn = 20$), $U = 2.5$, $Z = 2.83$, $p = 0.002$, $r = 0.76$. Home matches' patterns were more complex, with a higher number of levels ($Mdn = 1.9817$) than the away ones ($Mdn = 1,381$), $U = 2$, $Z = 2.875$, $p = 0.002$, $r = 0.77$, and a higher length ($Mdn = 3.2557$) than the away ones ($Mdn = 2.4048$), $U = 2$, $Z = 2.875$, $p = 0.002$, $r = 0.77$. No significant differences were found in the number of events coded per game between the home ($Mdn = 87$) and the away condition ($Mdn = 75$), $p = 0.259$.

A second analysis was performed to qualitatively compare home and away attack strategies, by combining the single datasets, from each of the two conditions, for the T-pattern

CRITERIA	CATEGORIES	CODE	DEFINITION
LATERAL POSITION			Lateral position of play
	Right	RI	
	Center	CE	
	Left	LE	
ZONE			Zone of play, defined according to five pitch areas or spatial strips
	Ultra-defensive	UD	
	Defensive	D	
	Central	C	
	Offensive	O	
	Ultra-offensive	UO	
LATERAL PASSING			Lateral direction of pass
	Right Passing	RP	
	Center Passing	CP	
	Left Passing	LP	
ZONE PASSING			Zone direction of pass
	Ultra-defensive Passing	UDP	
	Defensive Passing	DP	
	Central Passing	CP	
	Offensive Passing	OP	
	Ultra-offensive Passing	UOP	
RECOVERY & LOSS			
	Momentary Loss	ML	Ball is lost temporarily, the attack action continues. A ball is temporarily lost when there is a Recovery in the UO zone, or in the O zone after no more than 2 consecutive passes of the opposing team during loss time.
	Permanent Loss	PL	Ball is definitely lost, the attack action is over. It is considered permanent loss after 3 consecutive passes from the opposing team or if the ball is recovered in C, D, or UD zone.
	Recovery	R	Ball is recovered in UO zone, or in O zone no more than after 2 consecutive passes from the opposing team during loss time.
	Keeps Ball	KB	The player brings the ball from a zone to another or from a lateral position to another (or both at the same time). It can also happen if the player launches the ball forward and he himself gets to the ball in the next zone.
	Passive Ball	PB	Anything that is not attack action.
BALL OUT OF PLAY			
	Corner Kick	CK	
	Throw In	TI	
	Goal Kick	GK	
	Dropped Ball	DB	When the referee interrupts the game for reasons not due to the other conditions of BALL OUT OF PLAY (for example, medical staff is entering the area because a player is down but there is no foul).
	Goal	G	
	Not Goal	NG	Any shot not on target, or on target but saved by the keeper or a defender.
	Penalty Kick	PK	
	Free Kick	FK	

FIGURE 1 | Observation instrument.

detection. THEME detected 721 different T-patterns occurring in at least 80% of the home matches ($p = 0.005$). Two hundred and fifty-six out of 721 different T-patterns (36%) have four or more events (minimum threshold when speaking about complexity, since it means considering patterns that connect at least two different sub-patterns).

FIGURE 2 | Zones and lateral positions. Adapted from Camerino et al. (2012).
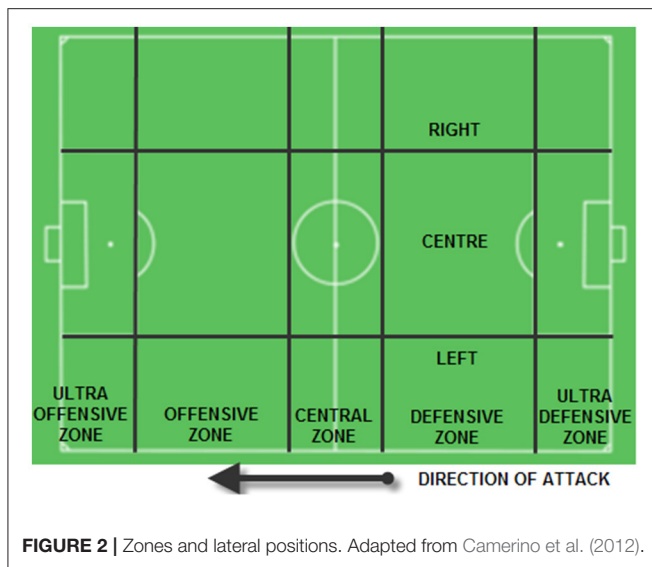
The most complex T-pattern (see **Figure 3**[1]) was detected 8 times in at least 80% of home matches, with a length of 10 events and 5 levels. It shows a ball keeping in left offensive area (o,le,kb), another ball keeping in left ultra-offensive area (uo,le,kb), a pass from the left ultra-offensive area to the central ultra-offensive area (uo,le,uop,cep), a permanent loss in this area (uo,ce,pl), a pass from the right ultra-offensive area to the central ultra-offensive area (uo,ri,uop,cep), a momentary loss in this area (uo,ce,ml), a corner kick from the right ultra-offensive area (uo,ri,ck), another pass from the right ultra-offensive area to the central ultra-offensive area (uo,ri,uop,cep), another momentary loss in this area (uo,ce,ml) and finally a permanent loss in this area (uo,ce,pl).

THEME detected 203 different T-patterns occurring in at least 80% of the away matches ($p = 0.005$). Twenty-nine out of 203 different T-patterns (14%) have four or more events.

The most complex T-pattern (see **Figure 4**) was detected 10 times, with a length of 5 events and 3 levels. It shows a pass from the central offensive area to the same area (o,ce,op,cep), followed by a momentary loss in the central ultra-offensive area (uo,ce,ml), a corner kick from the right ultra-offensive area (uo,ri,ck), a pass from the right ultra-offensive area to the central ultra-offensive area (uo,ri,uop,cep), and again a momentary loss in this area (uo,ce,ml).

---

[1]How to read the pattern tree graph: the left box of **Figures 3–6** shows the events occurring within the pattern, listed in the order in which they occur within the pattern. The first event in the pattern appears at the top and the last at the bottom. The lower right box shows the frequency of events within the pattern, each dot means that an event has been coded. The pattern diagram (the lines connecting the dots) shows the connection between events. The number of pattern diagrams illustrates how often the pattern occurs. Sub-patterns also occur when some of the events within the pattern occur without the whole of the pattern occurring. The upper box illustrates the real-time of the pattern. The lines show the connections between events, when they take place and how much time passes between each event.

Comparing Positive Attack Actions between Home and Away Matches

In order to explore the efficacy of the attack actions both in home and in away matches, we considered for a deeper analysis only those t-patterns that involved *Goals* or *NoGoals* (shots).

In home matches, THEME detected 82 common T-patterns including at least a Goal or a NoGoal event ($p = 0.005$). The most complex T-pattern (see **Figure 5**) was detected 12 times in at least 80% of home matches, with a length of 5 events and 3 levels. It shows a ball keeping in the left ultra-offensive area (uo,le,kb), followed by a pass from that zone to the central ultra-offensive area (uo,le,uop,cep), a momentary loss in this area (uo,ce,ml), a recovery in the central offensive area (o,ce,r), and finally a NoGoal (shot) from the same area (o,ce,ng).

THEME detected 20 common T-patterns in away matches with at least a Goal or a NoGoal ($p = 0.005$). The most complex T-pattern (see **Figure 6**) was detected 9 times in at least 80% of away matches, with a length of 4 events and 2 levels. It describes a momentary loss in the central offensive area (o,ce,ml), followed by a recovery in the same area (o,ce,r), a pass from the central offensive area to the same area (o,ce,op,cep), and a NoGoal from the same area (o,ce,ng).

DISCUSSION

Since the average number of events coded per game does not significantly differ between home and away matches, we can affirm that differences in patterns between the two conditions are not due to the number of events coded.

Mann–Whitney test results allow us to refuse the null hypothesis, hence confirming home and away matches to be significantly different in terms of patterns' number, diversity, and complexity. Descriptive statistics show that in home matches THEME detected a greater number of different patterns (H1 confirmed), characterized by a higher diversity (H2 confirmed) and complexity (H3 confirmed). Moreover, according to Cohen's classification (Cohen, 1988), results showed a large effect size, ranging from 0.7 to 0.8.

In terms of primary level of performance, home matches are characterized by a more structured (higher number of levels in patterns) and varied (longer patterns, each composed by different events) game; this could be due to the team having more confidence when applying rehearsed tactics, as well as new ones. In fact, away matches present a more stereotyped game, with simpler patterns; there seem to be more difficulties in structuring and changing the game tactics. The greater number of different behaviors, identified in home matches' patterns, confirms these data and provides the team with a wider range of opportunities not to be predictable, making it harder for the opponents to respond appropriately.

Qualitatively analyzing the more complex patterns of the two conditions, we noticed that they end similarly, with a corner kick and ball in the ultra-offensive area (positive events speaking about attack actions in soccer, because they can lead to a chance on goal). However, there is a difference if we focus on the events preceding this outcome. It seems that, in home matches, ball

	Home median	Away median	U	Z	p	r
Unique patterns	127	42	1	3.003	*0.001*	*0.80*
Unique behaviors	31	20	2.5	2.83	*0.002*	*0.76*
Patterns' levels	1.9817	1.381	2	2.875	*0.002*	*0.77*
Patterns' length	3.2557	2.4048	2	2.875	*0.002*	*0.77*
Number of events coded per game	87	75	NS	NS	*0.259*	NS

P- and r-values in italics are considered significant and related to a medium-large effect size.

FIGURE 3 | The most complex T-pattern from home matches. It occurred in 80% of the games analyzed. Events are (1) o,le,kb; (2) uo,le,kb; (3) uo,le,uop,cep; (4) uo,ce,pl; (5) uo,ri,uop,cep; (6) uo,ce,ml; (7) uo,ri,ck; (8) uo,ri,uop,cep; (9) uo,ce,ml; and (10) uo,ce,pl.
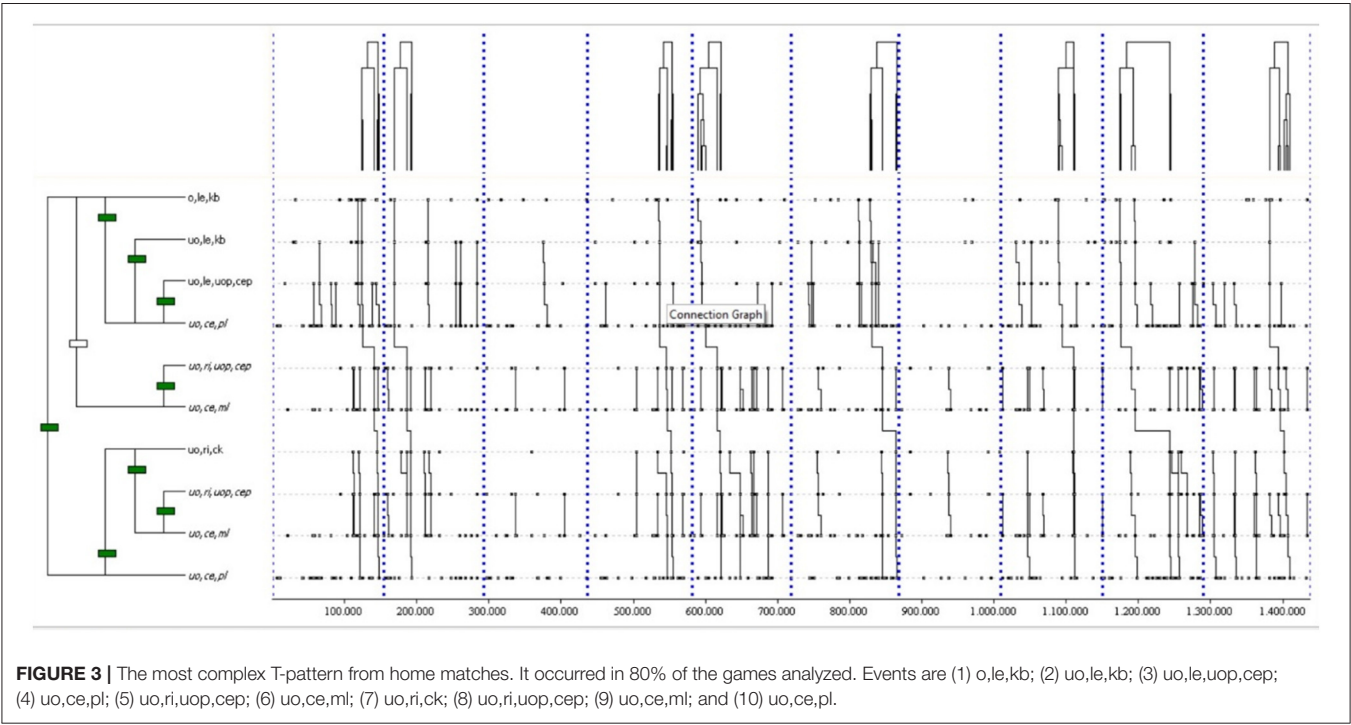
possession and the widening of the play on the lateral sides are an important part of the team's strategy. There are numerous and continuous changes of play with crosses from the sides directed to the central ultra-offensive area but there seem to be difficulties in exploiting those crosses, mistakes, and adversaries' defenses tend to prevail. Regarding away matches, the limited number of events composing the pattern makes it difficult to draw general conclusions about the team strategy, apart from highlighting a difficulty in the central penetration (from the offensive central to the ultra-offensive central zone), probably linked to a difficulty in widening the play. This simplicity, however, is an index of a difficulty for the team in creating and establishing a functional game strategy to getting important attack chances, such as the one described in the pattern. You could also say that this kind of occasions happened more randomly in away matches compared to what happened in the home condition.

By restricting the analysis to positive attack actions, we found that the number of patterns including goals and shots to goal is much higher in home rather than in away matches

(on a 4 to 1 ratio). This partially confirms previous findings. Sasaki et al. (1999), for example, found that the team they studied performed a greater number of goal attempts and shots on target during home matches. Different studies have shown significantly higher performance measures for attack indicators, such as shots to goal (Carmichael and Thomas, 2005; Tucker et al., 2005) and goals scored (Lago-Peñas and Lago-Ballesteros, 2011) when the teams they observed were performing at home. It seems that, in home matches, the shot to goal is more probably the result of a strategically repeated structured maneuver, compared to what happened in the away condition.

Qualitative analysis of the more complex patterns reveals that, like the previous case, they end similarly in the two conditions (a ball recovery in the offensive central zone and a shot from outside). However, while in home matches these events are preceded by a ball possession on the lateral sides and a penetration in the ultra-offensive zone, in the away matches the game develops in the central offensive zone, an area where it is more difficult for the team to create scoring chances.
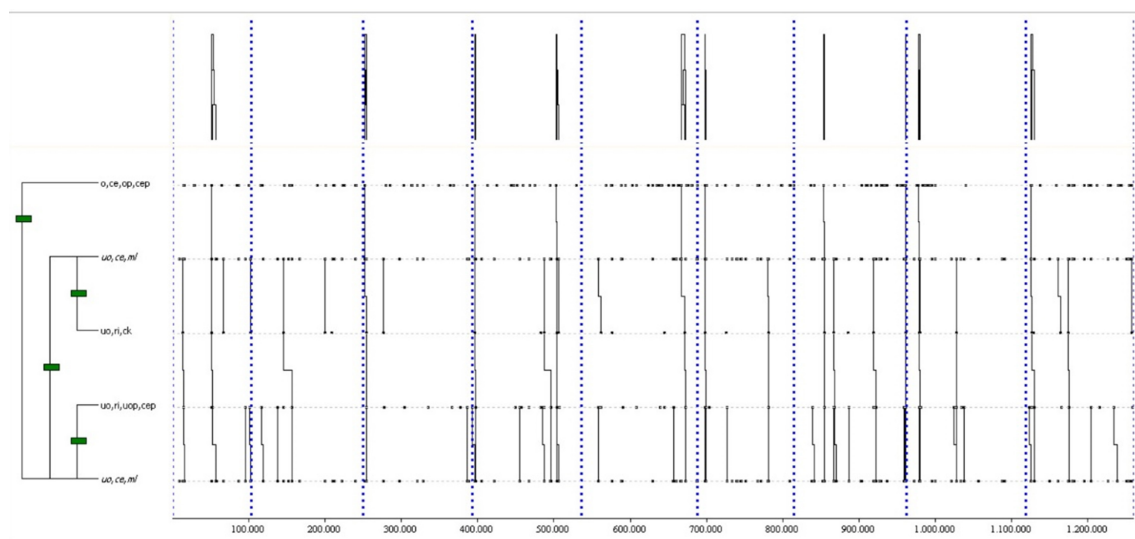
FIGURE 4 | The most complex T-pattern from away matches. It occurred in over 80% of the games analyzed. Events are (1) o,ce,op,cep; (2) uo,ce,ml; (3) uo,ri,ck; (4) uo,ri,uop,cep; and (5) uo,ce,ml.
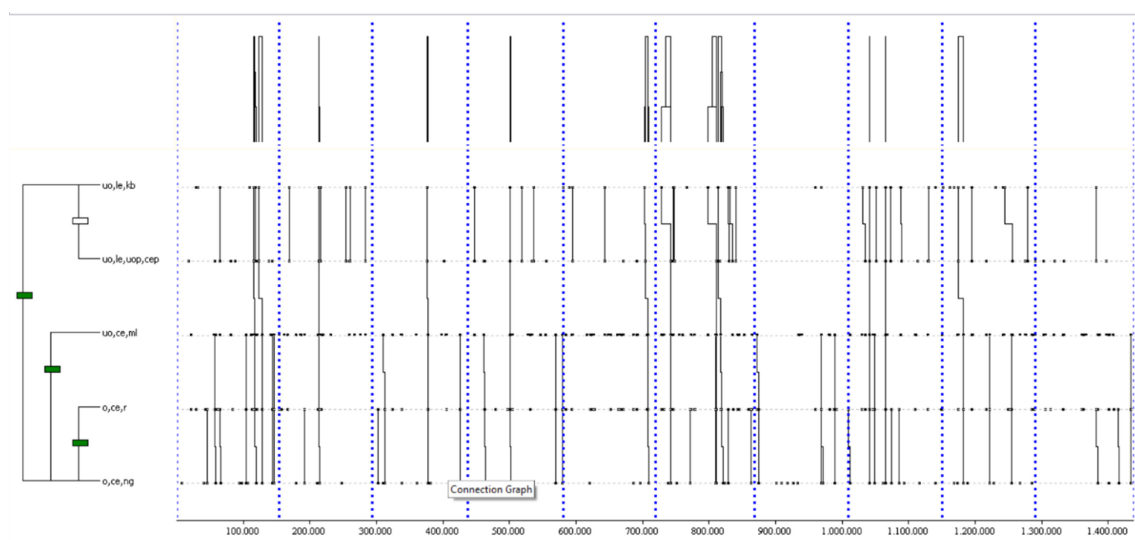
FIGURE 5 | One of the most complex T-patterns from home matches with a NoGoal. It occurred in 80% of the games analyzed. Events are (1) uo,le,kb; (2) uo,le,uop,cep; (3) uo,ce,ml; (4) o,ce,r; and (5) o,ce,ng.

CONCLUSION AND PRACTICAL APPLICATIONS

An important limitation of this study, similar to other studies previously discussed (e.g., Sasaki et al., 1999; Tucker et al., 2005; Taylor et al., 2008) is that we considered a single's team performance over a sustained period. Even if tactics and strategies are unique to individual teams and what is successful for one team may therefore not be for another (Tucker et al., 2005), aggregating performance of different teams during analysis may favor generalization of findings. Moreover, we did not examine the effects of match location on technical and tactical

performances as a function of team quality. Several studies have revealed that team quality affects the degree of home advantage obtained in sport (i.e., Schwartz and Barsky, 1977; Madrigal and James, 1999; Lago-Peñas, 2009; Lago-Peñas and Lago-Ballesteros, 2011).

However, in line with the findings of other studies that have applied T-pattern methodology to soccer (Jonsson et al., 2003, 2010; Camerino et al., 2012; Zurloni et al., 2014a), the results obtained and discussed above strengthen the belief that T-pattern analysis is an effective tool, producing different potential ways of supporting research in sport performance analysis and more specifically in game location effects.
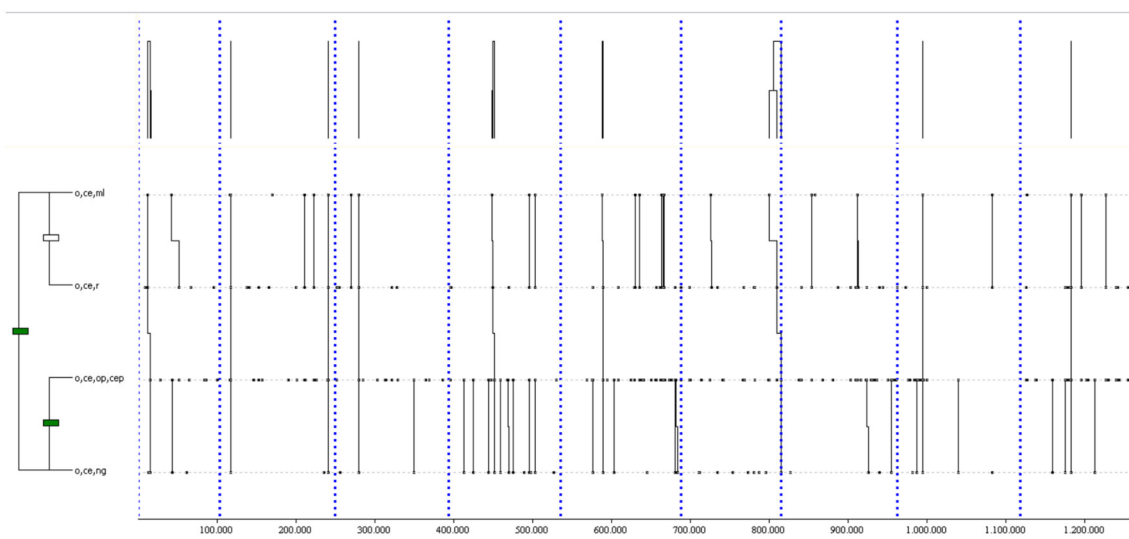
FIGURE 6 | The most complex T-pattern from away matches with a NoGoal. It occurred in over 80% of the games analyzed. Events are (1) o,ce,ml; (2) o,ce,r; (3) o,ce,op,cep; and (4) o,ce,ng.

In general, this study shows that game tactics are different between home and away matches. The home advantage detected at a tertiary level of performance is linked to a difference in the primary level, that t-pattern analysis translated in a more structured and varied game (more different patterns, longer and more complex), in which the ball possession in midfield and the numerous attempts to widen the play on the sides seem to represent an important part of the team's strategy. The situation is different for away matches; the "poor and stereotyped" structure of the detected patterns does not allow us to draw general indications on game attack strategies, apart from a difficulty in creating and establishing stable structures of play. It would be probably useful for the team to replicate the same "home strategy" of widening of the play, given the inefficacy of central penetration, which often leads to shots from outside.

These data need to be further developed and analyzed, for example by increasing the sample size, comparing tactics between different teams and extending the analysis to defensive play as well (*ad hoc* footage would be needed, since television footage usually follows the ball's movements).

These results also point toward the need to investigate the potential link between temporal structure detection and soccer managers' observations. In fact, the chance to create an observation instrument tailored on the needs of a specific team, exploiting the collaboration of the manager and his technical staff, would achieve even more significant results.

THEME and the corresponding T-pattern algorithm offer alternative means of analyzing team performance at a primary level, contributing to the research on the effects of game location, enriching the models already existing in literature and creating a different perspective that can be used to increase the team's performance at all levels. An important step will be to support this kind of analysis with the measurement of psycho-physiological factors, to understand which and how specific state and trait variables influence performance, aiming at a better monitoring and training of the team as a whole.

AUTHOR CONTRIBUTIONS

BD and VZ: Study designing, method development, data analysis, paper writing. ME: Data acquisition and coding, data analysis, paper writing. CC: Study designing, data acquisition and coding, paper writing. GJ: Method development, data analysis. MA: Method development, paper writing. All authors made suggestions and critical reviews to the initial draft and contributed to its improvement (until reaching the final manuscript), which was read and approved by all authors.

ACKNOWLEDGMENTS

REFERENCES

Allen, M. S., and Jones, M. V. (2014). The home advantage over the first 20 seasons of the english premier league: effects of shirt colour, team ability and time trends. *Int. J. Sport Exerc. Psychol.* 12, 10–18. doi: 10.1080/1612197X.2012.756230

Anguera, M. T. (1979). Observational typology. Quality & quantity. *Eur. Am. J. Methodol.* 13, 449–484.

Anguera, M. T., Blanco-Villaseñor, A., Hernández-Mendo, A., and Losada, J. L. (2011). Diseños observacionales: ajuste y aplicación en psicología del deporte. *Cuadernos de Psicol. del Deporte* 11, 63–76.

Anguera, M. T., Zurloni, V., Jonsson, G. K., Diana, B., and Elia, M. (2013). "T-Pattern analysis in soccer: detecting regularities in attack plays," in *18th Annual Congress of the European College in Sport Science (ECSS)* (Barcelona).

Anolli, L., and Zurloni, V. (2009). Cooperative and non cooperative lies in close and casual relationships. *Stud. Commun. Sci.* 9, 7–25.

Aragón, S., Lapresa, D., Arana, J., Anguera, M. T., and Garzón, B. (2015). Tactical behaviour of winning athletes in major championship 1500-m and 5000-m track finals. *Eur. J. Sport Sci.* 16, 279–286. doi: 10.1080/17461391.2015.1009494

Borrie, A., Jonsson, G. K., and Magnusson, M. S. (2002). Temporal pattern analysis and its applicability in sport: an explanation and exemplar data. *J. Sports Sci.* 20, 845–852. doi: 10.1080/026404102320675675

Brown T. D. Jr., Van Raalte, J. L., Brewer, B. W., Winter, C. R., Cornelius, A. E., and Andersen, M. B. (2002). World cup soccer home advantage. *J. Sport Behav.* 25, 134–144.

Camerino, O. F., Chaverri, J., Anguera, M. T., and Jonsson, G. K. (2012). Dynamics of the game in soccer: detection of T-patterns. *Eur. J. Sport Sci.* 12, 216–224. doi: 10.1080/17461391.2011.566362

Carmichael, F., and Thomas, D. (2005). Home-field effect and team performance: evidence from English premiership football. *J. Sports Econ.* 6, 264–281. doi: 10.1177/1527002504266154

Carron, A. V., Loughhead, T. M., and Bray, S. R. (2005). The home advantage in sport competitions: Courneya and Carron's (1992) conceptual framework a decade later. *J. Sports Sci.* 23, 395–407. doi: 10.1080/02640410400021542

Casarrubea, M., Davies, C., Faulisi, F., Pierucci, M., Colangeli, R., Partridge, L., et al. (2015a). Acute nicotine induces anxiety and disrupts temporal pattern organization of rat exploratory behavior in hole-board: a potential role for the lateral habenula. *Front. Cell. Neurosci.* 9:197. doi: 10.3389/fncel.2015.00197

Casarrubea, M., Jonsson, G. K., Faulisi, F., Sorbera, F., Di Giovanni, G., Benigno, A., et al. (2015b). T-pattern analysis for the study of temporal structure of animal and human behavior: a comprehensive review. *J. Neurosci. Methods* 239, 34–46. doi: 10.1016/j.jneumeth.2014.09.024

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Canton, A., and Hileno, R. (2016). Goal scoring in soccer: a polar coordinate analysis of motor skills used by lionel Messi. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Castañer, M., Torrents, C., Anguera, M. T., Dinušová, M., and Jonsson, G. K. (2009). Identifying and analyzing motor skill responses in body movement and dance. *Behav. Res. Methods* 41, 857–867. doi: 10.3758/BRM.41.3.857

Cavalera, C., Diana, B., Elia, M., Jonsson, G. K., Zurloni, V., and Anguera, M. T. (2015). T-pattern analysis in soccer games: relationship between time and attack actions. *Cuadernos de Psicologia del Deporte* 15, 41–50. doi: 10.4321/S1578-84232015000100004

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, Vol. 2.* Hillsdale, NJ: Lawrence Earlbaum Associates.

Courneya, K. S., and Carron, A. V. (1992). The home advantage in sport competitions: a literature review. *J. Sport Exerc. Psychol.* 14, 13–27. doi: 10.1123/jsep.14.1.13

Diana, B., Elia, M., Zurloni, V., Elia, A., Maisto, A., and Pelosi, S. (2015). "Multimodal deception detection," in *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection— WMDD'15* (New York, NY: ACM Press), 21–28.

Fernández, J., Camerino, O., Anguera, M. T., and Jonsson, G. (2009). Identifying and analyzing the construction and effectiveness of offensive plays in basketball by using systematic observation. *Behav. Res. Methods* 41, 719–730. doi: 10.3758/BRM.41.3.719

Gabín, B., Camerino, O., Anguera, M. T., and Castañer, M. (2012). Lince: multiplatform sport analysis software. *Proc. Soc. Behav. Sci.* 46, 4692–4694. doi: 10.1016/j.sbspro.2012.06.320

Jones, M. B. (2013). The home advantage in individual sports: an augmented review. *Psychol. Sport Exerc.* 14, 397–404. doi: 10.1016/j.psychsport.2013.01.002

Jones, M. R., West, D. J., Crewther, B. T., Cook, C. J., and Kilduff, L. P. (2015). Quantifying positional and temporal movement patterns in professional rugby union using global positioning system. *Eur. J. Sport Sci.* 15, 488–496. doi: 10.1080/17461391.2015.1010106

Jonsson, G. K., Anguera, M. T., Sánchez-Algarra, P., Olivera, C., Campanico, J., Castañer, M., et al. (2010). Application of T-pattern detection and analysis in sports research. *Open Sports Sci. J.* 3, 95–104. doi: 10.2174/1875399X010030100095

Jonsson, G. K., Bjarkadottir, S. H., Gislason, B., Borrie, A., and Magnusson, M. S. (2003). Detection of real-time patterns in sports: interactions in football. *L'Éthologie Appliquée Aujourd'Hui* 3, 37–46.

Lago-Peñas, C. (2009). The influence of match location, quality of opposition, and match status on possession strategies in professional association football. *J. Sports Sci.* 27, 1463–1469. doi: 10.1080/02640410903131681

Lago-Peñas, C., and Lago-Ballesteros, J. (2011). Game location and team quality effects on performance profiles in professional soccer. *J. Sports Sci. Med.* 10, 465–471.

Lapresa, D., Alvarez, L., Arana, J., Garzón, B., and Caballero, V. (2013a). Observational analysis of the offensive sequences that ended in a shot by the winning team of the 2010 UEFA Futsal Championship. *J. Sports Sci.* 31, 1731–1739. doi: 10.1080/02640414.2013.803584

Lapresa, D., Arana, J., Anguera, M. T., and Garzón, B. (2013b). Comparative analysis of the sequentiality using SDIS-GSEQ and THEME: a concrete example in soccer. *J. Sports Sci.* 31, 1687–1695. doi: 10.1080/02640414.2013.796061

Madrigal, R., and James, J. (1999). Team quality and the home advantage. *J. Sport Behav.* 22, 381–398.

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behav. Res. Methods Instrum. Comput.* 32, 93–110. doi: 10.3758/BF03200792

Magnusson, M. S. (2005). "Understanding social interaction: discovering hidden structure with model and algorithms," in *The Hidden Structure of Social Interaction. from Genomics to Culture Patterns*, eds L. Anolli, S. Duncan, M. S. Magnusson, and G. Riva (Amsterdam: IOS Press), 51–70.

Neave, N., and Wolfson, S. (2003). Testosterone, territoriality, and the home advantage. *Physiol. Behav.* 78, 269–275. doi: 10.1016/S0031-9384(02)00969-1

Nevill, A. M., and Holder, R. L. (1999). Home advantage in sport. *Sports Med.* 28, 221–236. doi: 10.2165/00007256-199928040-00001

Pollard, R. (1986). Home advantage in soccer: a retrospective analysis. *J. Sports Sci.* 4, 237–248. doi: 10.1080/02640418608732122

Pollard, R. (2006). Home advantage in soccer: variations in its magnitude and a literature review of the inter-related factors associated with its existence. *J. Sport Behav.* 29, 169–189.

Pollard, R., and Gómez, M. Á. (2013). Variations in home advantage in the national basketball leagues of Europe. *Rev. Psicol. Deporte* 22, 263–266.

Portell, M., Anguera, M. T., Chacón, S., and Sanduvete, S. (2015a). Guidelines for Reporting Evaluations based on Observational Methodology (GREOM). *Psicothema* 27, 283–289. doi: 10.7334/psicothema2014.276

Portell, M., Anguera, M. T., Hernández-Mendo, A., and Jonsson, G. K. (2015b). Quantifying biopsychosocial aspects in everyday contexts: an integrative methodological approach from the behavioral sciences. *Psychol. Res. Behav. Manag.* 8, 153–160. doi: 10.2147/PRBM.S82417

Remmert, H. (2003). Analysis of group-tactical offensive behavior in elite basketball on the basis of a process orientated model. *Eur. J. Sport Sci.* 3, 1–12. doi: 10.1080/17461390300073311

Sánchez-Algarra, P., and Anguera, M. T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: Impact of recording and coding predominating perspectives. *Qual. Quant. Int. J. Methodol.* 47, 1237–1257. doi: 10.1007/s11135-012-9764-6

Sarmento, H., Bradley, P., Anguera, M. T., Polido, T., Resende, R., and Campaniço, J. (2015). Quantifying the offensive sequences that result in goals in elite futsal matches. *J. Sports Sci.* 34, 621–629. doi: 10.1080/02640414.2015.1066024

Sarmento, H., Marcelino, R., Anguera, M. T., Campaniço, J., Matos, N., and Leitão, J. C. (2014). Match analysis in football: a systematic review. *J. Sports Sci.* 32, 1831–1843. doi: 10.1080/02640414.2014.898852

Sasaki, Y., Nevill, A., and Reilly, T. (1999). Home advantage: a case study of Ipswich Town football club during the 1996-97 season. *J. Sports Sci.* 17:831.

Schwartz, B., and Barsky, S. (1977). The home advantage. *Soc. Forces* 55, 641–661. doi: 10.2307/2577461

Seçkin, A., and Pollard, R. (2008). Home advantage in Turkish professional soccer. *Percept. Mot. Skills* 107, 51–54. doi: 10.2466/pms.107.1.51-54

Taylor, J. B., Mellalieu, S. D., James, N., and Sheraer, D. (2008). The influence of match location, qualify of opposition and match status on technical performance in professional association football. *J. Sports Sci.* 26, 885–895. doi: 10.1080/026404107018 36887

Tucker, W., Mellalieu, S. D., James, N., and Taylor, J. B. (2005). Game location effects in professional soccer. A case study. *Int. J. Perform. Anal. Sports* 5, 23–35.

Wallace, H. M., Baumeister, R. F., and Vohs, K. D. (2005). Audience support and choking under pressure: a home disadvantage? *J. Sports Sci.* 23, 429–438. doi: 10.1080/02640410400021666

Wolfson, S., Wakelin, D., and Lewis, M. (2005). Football supporters perceptions of their role in the home advantage. *J. Sports Sci.* 23, 365–374. doi: 10.1080/02640410400021567

Zurloni, V., Cavalera, C., Diana, B., Elia, M., and Jonsson, G. (2014a). Detecting regularities in soccer dynamics. *Rev. Psicol. Deporte* 23, 157–164.

Zurloni, V., Diana, B., Cavalera, C., Argenton, L., Elia, M., and Mantovani, F. (2014b). Deceptive behavior in doping related interviews: the case of Lance Armstrong. *Psychol. Sport Exerc.* 16, 191–200. doi: 10.1016/j.psychsport.2014.02.008

Zurloni, V., Diana, B., and Elia, M. (2013). "Imposing cognitive load to unmask prepared lies A temporal pattern detection approach," in *2013 10th IEEE International Conference and workshops on Automatic Face and Gesture Recognition (FG)* (Shanghai: IEEE), 1–6.

Zurloni, V., Diana, B., Elia, M., and Anolli, L. (2016). "Imposing cognitive load to detect prepared lies: a t-pattern approach," in *Discovering Hidden Temporal Patterns in Behavior and Interaction: T-Pattern Detection and Analysis with THEME*[TM], *Neuromethods, Vol. 111*, eds M. S. Magnusson, J. K. Burgoon, and M. Casarrubea (New York, NY: Springer Science+Business Media), 63–82.

Possession Zone as a Performance Indicator in Football. The Game of the Best Teams

Claudio A. Casal[1]*, Rubén Maneiro[2], Toni Ardá[3], Francisco J. Marí[4] and José L. Losada[4]

[1] Department of Science of Physical Activity and Sport, Catholic University of Valencia "San Vte Mártir", Valencia, Spain, [2] Department of Science of Physical Activity and Sport, Pontifical University of Salamanca, Salamanca, Spain, [3] Department of Physical and Sport Education, University of A Coruña, A Coruña, Spain, [4] Department of Methodology of Behavioral Sciences, University of Barcelona, Barcelona, Spain

Possession time in football has been widely discussed in research but few studies have analyzed the importance of the field area in which possession occurs. The objective of this study was to identify the existence of significant differences in the field zone of ball possession between successful and unsuccessful teams and to acknowledge if the match status modulates the possession model. To this end, 2,284 attacks were analyzed corresponding to the matches in the final phase of the UEFA Euro 2016 France, recording possession time and field zone in which possession occurred. Video recordings of matches were analyzed and coded post-event using notational analysis. We have found that successful offensive game patterns are different from unsuccessful ones. Specifically, field zone in which major possession occurs changes significantly between successful and unsuccessful teams ($x^2 = 15.72$, $p < 0.05$) and through Welch's T significant differences were detected in possession time between successful and unsuccessful teams ($H = 24.289$, $p < 0.001$). The former are characterized by longer possession times, preferably in the middle offensive zone, on the other hand, unsuccessful teams have shorter possession times and preferably on the middle defensive zone. Logistic regression also allowed us to identify that greater possession in the middle offensive zone is a good indicator of success in the offensive game, allowing us to predict a greater chance of victory in the match. Specifically, every time the teams achieve possession in the middle offensive zone, the chance of winning the match will increase 1.72 times and, the probability of winning the match making longer possessions in the middle offensive zone is 44.25%. Applying the Kruskal–Wallis test we have also been able to verify how match status modulates the teams possession time, specifically, when teams are winning they have longer possessions $x^2 = 92.628$, $p = 0.011$. Results obtained are expected to help gain more knowledge about successful offensive game models, as well as performance factors of the offensive phase, which will allow teams to optimize their training process and performance during the match.

Keywords: observational methodology, football, offensive phase, possession ball, performance indicators

INTRODUCTION

In sports games and specifically in football, encounter analysis through systematic observation is an effective and objective instrument to collect information and identify the most relevant events that occur in them, as revealed by Carling et al. (2009) when affirming that match analysis has taken a transcendental role in sports.

In many cases observation is the only way to study a phenomenon without distorting it, watching it as it occurs in game context since, according to Anguera (1993), it is a particular strategy of the scientific method that proposes the quantification of spontaneous behavior that occurs in unprepared situations, implying that to achieve results an orderly series of stages is required (problem definition, design, data collection, data analysis, and results interpretation). This is the only scientific methodology that allows data to be collected directly from playing participants in competitions, without eliciting the response from the direct apprehension of perceptible information, preferably helping us through recording, which is the usual method to access information (Anguera and Mendo, 2013).

Observational methodology is a scientific procedure that allows the detection of behaviors perceiving them in their usual context, proceed with systematic recording and analysis, both qualitative and quantitative and mixed methods (Anguera et al., 2014), using a suitable instrument, enabling the detection of different types of relations and evaluating them. This will require the selection and use of the most appropriate analysis tool depending on the data collected nature (qualitative, quantitative or mixed) and the intended results (descriptive, comparative, or predictive). Observational methodology, proposes certain procedural structures –observational designs– through a set of criteria which are the natural bases of observational studies. In each study, once the objectives have been defined, the observational designs established then guide the entire process, influencing the preparation of the observation instruments, the recording and its metrics, the observational sample, data quality control and to a large extent the choice of the most appropriate analysis techniques. They also have a significant repercussion on the interpretation of the results.

In football, unlike other team sports of cooperation-opposition and simultaneous participation, due to its complex nature (Davids et al., 2005; Araújo et al., 2006; Perl, 2006), high uncertainty and multifactoriality (Gréhaigne, 2001; McGarry et al., 2002; Lames and McGarry, 2007), the search is not easy, which means that identifying factors that affect success is of particular interest (O'Donoghue, 2010). Performance in this sport has a multidimensional setting and can be grouped into two broad areas of study. On the one hand, we would find analytical factors related to conditional aspects and, on the other hand, competition factors that would require an analysis in its natural context. Within the latter, tactical-strategic aspects allow to better reflect the nature of the game and to better understand its development.

In recent years this type of work has proliferated (Hughes and Franks, 2005; Lago, 2009) aiming to detect successful play patterns through the analysis of different game situations and different variables. Some of these studies focus their interest in studying the offensive phase (Ensum et al., 2000; James et al., 2004; Jones et al., 2004; Hughes and Franks, 2005; Lago and Martín, 2007; Acar et al., 2009; Lago-Ballesteros et al., 2012; Collet, 2013; Casal et al., 2015; Ric et al., 2016), others the defensive phase (Barreira et al., 2013; Vogelbein et al., 2014; Andujar, 2015; Mohammad et al., 2016; Ric et al., 2016) and others in the analysis of situational variables (Borrás and Sainz de Baranda, 2005; Tucker et al., 2005; Taylor et al., 2008; Lago, 2009, 2012; Lago-Peñas and Dellal, 2010; Lago-Peñas and Lago-Ballesteros, 2011; De Oliveira, 2012; Sainz de Baranda and López-Riquelme, 2012; Sánchez Flores et al., 2012; Ardá et al., 2014; Casal et al., 2014, 2015).

Another element of great interest in football's performance analysis is the identifying and understanding differences between game patterns developed by successful and unsuccessful teams (Hughes and Bartlett, 2002). To acquire objective information that permits assessing team performance (Carling et al., 2005) and differentiate playstyle from successful teams and unsuccessful ones (Mckenzie and Cushion, 2012).

One of the most studied indicators in football research has been possession (Bate, 1988; Dawson et al., 2000; Garganta, 2000; Hadley et al., 2000; Carmichael et al., 2001; Hughes and Bartlett, 2002; Hughes, 2003; McGarry and Franks, 2003). This is because it can lead a team to take the initiative of the offensive game, though it doesn't necessarily mean to win the match. In recent years this variable has acquired greater significance due to the success of teams like F.C. Barcelona and the Spanish national team who have maintained hegemony in European and world football using a playstyle based on possession and taking the lead through keeping the ball.

This fact is reinforced by some studies that claim that greater possession implies greater team success. Hook and Hughes (2001) reported that successful teams in the UEFA Champions League, FIFA World Cup, and UEFA Euro achieved longer possession time than the unsuccessful teams. Bloomfield et al. (2005a) showed that the top three teams in the 2003–2004 English Premier League (Arsenal, Chelsea, and Manchester United) achieved longer possession time than their opponents. James et al. (2004) detected significant differences in possession between successful and unsuccessful teams from the English Premier League. Carling et al. (2005) obtained the same results in a study from the same league but in the 1996–1997 season. Grant et al. (1999) analyzed the 1998 FIFA World Cup and Hook and Hughes (2001) 2000 UEFA Champions League, both studies reaching the same conclusion, that possession is linked to team success. Casal et al. (2015) analyzed the 2008 UEFA Euro, concluding that a longer offensive phase predicts greater success, and studies Grant et al. (1999), Hook and Hughes (2001), James et al. (2004), Bloomfield et al. (2005b), Carling et al. (2005), Hughes and Franks (2005), Collet (2013), Casal et al. (2015) also corroborate the relationship between greater possession and team success.

But it seems presumptuous to claim that longer possession time ensures greater success, as the results of different studies are inconclusive and reality shows how teams with low possession

time are also successful, as demonstrated by studies like those of Bate (1988) which indicates that teams are more likely to achieve goals having the ball near the goal zone and not the longer you keep the ball on your own possession, even though these two variable are often related. Stanhope (2001) also indicates that possession did not represent the successful teams of the 1994 FIFA World Cup, although it seems that the game strategies used by the successful teams have evolved over the years into a more possession based playstyle. Studies Lago and Martín (2007), Lago (2009), and Lago-Peñas and Dellal (2010) indicate that in the Spanish League greater possession is a feature observed in teams that are either losing or tying the game. Collet (2013) concludes that the effect of possession time in matches of the domestic league was negative, in the UEFA Champions League had no effect and in National team tournaments was not significant, leading to think that the influence of possession on success will depend on team capacity. Moreover, we must emphasize that in season (2015/16), according to data collected on FIFA's official website, the top teams in the major European leagues (Bundesliga in Germany, France Ligue 1 in France, Spanish Liga, A Series in Italy and the Premier League in England) have possession times over 50% with the exception of Leicester, leader of the Premier League, which has 42% of possession during matches. Possession time or offensive phase duration could also be explained by the playstyle selected or some situational variables. Some studies have shown that possession is influenced by the match status (Sasaki et al., 1999; James et al., 2004; Jones et al., 2004; Bloomfield et al., 2005a; Lago and Martín, 2007; Taylor et al., 2008). Studies Lago and Martín (2007) and Lago (2009) found that losing teams had longer possession times in the offensive zone rather than the defensive zone.

Another variable that modulates possession time is the match location, and some studies show that home teams have longer possession times than away teams (James et al., 2002, 2004; Jones et al., 2004; Lago-Peñas and Dellal, 2010). The quality of rival team also varies possession time, being greater when facing rivals with low capacity level (Jones et al., 2004; Bloomfield et al., 2005a; Tucker et al., 2005; Lago and Martín, 2007; Lago, 2009). A transcendental aspect when possession is analyzed as a performance indicator is to discern the quality of it, as Collet (2013) advises. It will therefore be important to not only quantify the time a team retains possession during the offensive phase, but also to identify the zone in which it is carried out as keeping the ball in fruitless offensive zones (away from the goal) might not guarantee the success of the offensive phase, although it may be a recommended strategy to defend possession in circumstances that reccomend it. In this study, an analysis of ball possession of the 2016 UEFA Euro France was realized, the main objective being to identify the possible relationship between possession time and the zone in which it develops with team success, reflected in the results of the match. That is, we want to know whether the successful and unsuccessful teams are characterized by more or less possession in certain zones of the field, showing a different offensive game. The main contribution that this study provides to the scientific field is conducting a quantitative and qualitative analysis of ball possession, as it not only means to quantify the time of team possession but also to identify the area

where this occurs in order to determine the quality of the same. On the other hand, performing a multivariate analysis to identify the influence of possession time and area on the outcome of the match, and identifying a model that enables us to predict team success based on these variables.

The hypothesis of this study is that team level modulates the type of ball possession, both quantitatively (possession time) and qualitatively (area of possession).

MATERIALS AND METHODS

Participants

To control some of the situational variables that can potentially affect tactical and strategic team behavior, such as quality or level of opposing teams and the match location (Kormelink and Seeverens, 1999; Carling et al., 2005), 12 matches corresponding to the round of eighth-finals, quarterfinals, semifinals and final of the 2016 UEFA Euro France have been selected in which 2.284 ball possessions occured. Switzerland, Poland, Croatia, Portugal, Wales, Northern Ireland, Hungary, Belgium, Germany, Slovakia, Italy, Spain, France, Eire, England, and Iceland were the teams analyzed. Three games (Switzerland vs. Poland; Poland vs. Portugal and Germany vs. Italy) have been excluded from the analysis since the match outcome was a draw having in account regular time and extensions, which makes impossible to label the teams as successful or unsuccessful. This sample ensures that all matches are played on neutral ground, the teams have a similar level and, by eliminating the games of the group phase, we also make sure that the teams look for the victory in their matches, since defeat will mean elimination. In the group phase matches, it may happen that some team is more interested in drawing or losing any of their matches, to avoid a particular opponent in the following phases, this would lead to incorrect results in the study.

Instruments

Four national coaches and experts in football research designed an ad hoc observation instrument combining a field format and category system (Anguera and Mendo, 2013) was created (**Table 1**). Variables designed for the study are time (time that teams have ball possession in each field zone, in minutes); possession zone (spatial division of the field in defensive half and offensive half); match outcome (determined based on the number of goals scored and conceded at the end of the match); match status (match result at the time of registering each possession); match half; move outcome.

Procedure

In order to carry out the study, a direct, non-participatory, systematic, and natural observational methodology was used (Anguera et al., 2011).

Matches were recorded from TV emitted images and were registered and analyzed post-event. Because the video recordings were public, confidentiality was not an issue and authorization was not required from the players observed or their representatives. Furthermore, the information cannot be

Criterion	Categories
Time	Possession time in each zone
Possession zone	Middle defensive zone
	Middle offensive zone
Match outcome	Win
	Draw
	Loss
Match status	Winning
	Drawing
	Losing
Match half	First half
	Second half
Move outcome	Goal
	Shot
	Own corner kick
	Opponent's corner kick
	Own throw-in
	Opponent's throw-in
	Own's foul
	Opponent's foul
	Lost possession

considered either personal or intimate, as the research consisted solely of naturalistic observations in public places, and it was not anticipated that the recordings would be used in a manner that could cause personal harm (The American Psychological Association's [APA's], 2010). No experimental analysis involving human studies is performed in the study.

Basic Concepts

Basic concepts used in this study are, firstly, the definition of ball possession. We have adopted the definitions of two previous studies (Castellano, 2000; Casal, 2011), determined that a team starts a possession, while it is in play or when a player gets the ball while it is in possession of the other team must meet at least one of the following criteria:

(1) The player who receives the ball must touch it at least two times.
(2) The player intercepts the ball and a partner continues possession.

If the ball is stationary, a team starts a possession, when the ball has been put into play after a reglementary interruption had been decreed and consequently the match stopped. The analysis unit was composed for the entire offensive phase of the team, since ball possession started until it was lost or the match was interrupted.

Space arrangement used harnesses the subdivision performed by field regulation, dividing the field into two parts by a vertical line (central line). The zone of the field comprised between the central line and the bottom line of the goal of a team has been called middle defensive zone and the other half, bounded by the central line to the bottom line of the opposing goal has been called middle offensive zone.

Criteria used for the division of the teams into two groups, successful and unsuccessful, has been the outcome of the match (Lago-Peñas et al., 2010), excluding penalties. This way, all the teams that won their matches during reglementary time or extensions were classified as successful and teams who lost their matches as unsuccessful.

Data Quality Control

To try to ensure data reliability, all matches were registered and analyzed by four observers, all of them national soccer coaches with more than 10 years of experience in the field of training, teaching, and research in football through observational methodology. In addition, the following training process was carried out: First, eight observing sessions were conducted on teaching the observers following the Losada and Manolov (2014) criteria and applying the criterion of consensual agreement (Anguera, 1990) among observers, so that recording was only done when agreement was produced. To ensure inter-reliability consistency of the data (Berk, 1979; Mitchell, 1979) the Kappa coefficient was calculated for each criterion (Table 2), it revealed a strong agreement between observers, which means high reliability, taking Fleiss (1981) as a reference, who establishes a classification for the Kappa values where it characterizes as regular values found between 0.40 and 0.60, good between 0.60 to 0.75, and excellent above 0.75. Moreover, the procedure was repeated after 2 weeks (to exclude any learning effects) to check intraobserver reliability (Mitchell, 1979).

Statistical Analysis

Variables analyzed were Match Status, Half Match, Possession Zone, Move Outcome in relation to Possession Time. In the case of possession time and match status result was significant, and was complemented by a Kruskal–Wallis *post hoc* test to know among which categories the differences existed. Half match proved to be non-significant while possession zone has a significant result.

In the case of possession time and move outcome, several play options are analyzed, applying the Kruskal–Wallis test to see if differences were found between them (Tenga and Sigmundstad, 2011).

A comparative analysis of possession zone between successful and unsuccessful teams (match outcome) was also carried out, with significant differences between both groups of teams. The size effect was calculated in terms of Cramer's and Chupov that showed low intensity between the two variables. We also found differences between possession time and successful and unsuccessful teams, using Welch's T. To know the size of the effect, a point-biserial correlation was applied (Nakagawa and Cuthill, 2007), indicating that a relationship exists, but with a low intensity.

Finally, a logistic regression model was performed, to know the influence that possession time and possession zone (predictor variables) have on match outcome (variable explained). The model's degree of adjustment was verified (Ato and López, 1996; Hair et al., 1999), and once verified the success probability estimation was calculated, depending on the values of predictor variables.

TABLE 2 | Observers inter-reliability by criterion.

Criteria	Ob$_1$-Ob$_2$	Ob$_1$-Ob$_3$	Ob$_1$-Ob$_4$	Ob$_2$-Ob$_3$	Ob$_2$-Ob$_4$	Ob$_3$-Ob$_4$
Time	0.9	1	0.89	0.95	1	0.92
Possession zone	1	0.93	1	0.98	0.97	1
Match outcome	1	1	1	1	1	1
Match status	1	1	1	1	1	1
Match half	1	1	1	1	1	1
Move outcome	1	1	1	1	1	1
K_{total}	0.98	0.99	0.98	0.98	0.99	0.98

To perform statistical analysis the R program (v.3.2.0) was used, libraries used were epiDisplay, pscl, BaylorEdPsych, and Modeva. Significance level for each performance indicator was set at 5%, as usual in comparable scientific studies (Taylor et al., 2005).

RESULTS

Agreeing with Allen (2003) definition, the most common way of describing a set of interrelated data is to calculate the mean value and a dispersion measure around this mean value. We started presenting the related values between "match status" and "possession time," which shows that the average possession time in a winning team is 20.3 m with a deviation of ±16.0 m ($N = 667$) during the match. In case of a draw, shows a mean value of 18.2 m with a standard deviation of ±16.8 m ($N = 912$). Finally, in the case of losing the mean length of possession is 13.7 m, with a deviation of ±12.3 m ($N = 705$). The relationship between the three categories of the variable "match status" indicate that there are significant differences between them (p-overall < 0.01) (**Table 3**). The standard error is important, because records have a large dispersion, most are outliers.

Average "possession time" (**Figure 1**) is smaller with the result "losing." For a "draw" result the average increases slightly, and finally presents the greatest value for the result "winning."

In order to know among which categories the differences occur, comparisons are proposed two to two, with a Kruskal–Wallis *post hoc* test. The Kruskal–Wallis test shows a chi-square statistic value of 92,628, with p-value = 0.011, indicating differences between categories. In the *post hoc* contrast, significant differences are found in the WINNING–LOSING and DRAWING–LOSING pairs (**Table 4**).

Analyzing the relationship between the variables "match half" and "possession time," we obtained a mean value of 17.6 m with a

FIGURE 1 | Match status and possession time diagram box.

TABLE 4 | Categories differences based on pairwise comparisons.

	Obs.dif	Critical.dif	Difference
Winning–Drawing	26.02823	40.77737	False
Winning–Losing	75.40521	43.22467	True
Drawing–Losing	49.37698	40.12126	True

standard deviation of ±15.3 m ($N = 1,190$ plays) in the first half, while in the second half the average length of possession is 17.2 m with a deviation of ±15.8 m ($N = 1,094$). These differences were not statistically significant ($p = 0.73$) (**Table 5**).

In **Figure 2** a slight reduction in possession time is observed in the second half of the match.

Variables "possession zone" and "possession time" have a mean value of 16.0 m, with a standard deviation of 13.5 m ($N = 1,053$) in the middle defensive zone. The middle offensive zone has a mean value of 18.6 m with a standard deviation of ±17.0 m

TABLE 3 | Relationship of possession time and match status.

	Winning	Drawing	Losing	p_overall
	$N = 667$	$N = 912$	$N = 705$	
Possession time	20.3 ± 16.0	18.2 ± 16.8	13.7 ± 12.3	<0.01

TABLE 5 | Relationship possession time – half match.

	First half	Second half	p_overall
	$N = 1190$	$N = 1094$	
Possession time	17.6 m ± 15.3 m	17.2 m ± 15.8 m	0.73

FIGURE 2 | Half match and possession time diagram box.

FIGURE 3 | Possession zone (field zone) and possession time diagram box.

FIGURE 4 | Move outcome and possession time diagram box.
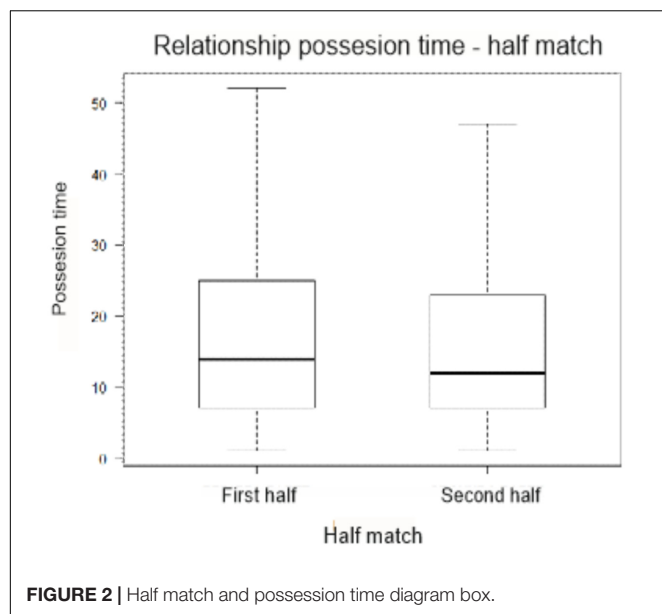
(N = 1,231). Differences are significant (p = 0.04-overall), (**Table 6**).

It is seen in mean values of "possession time" which is slightly lower for the middle defensive zone compared to the offensive zone. We observed that distance in the middle offensive zone in the interquartile range is greater, plus a greater dispersion of the observations (**Figure 3**). This means that the hold time is increased in this area.

In the box diagram (**Figure 4**), like the previous box diagrams (**Figure 3**), outliers have been deleted on their behalf to have a better view of the distributions of each category of move outcome based on possession time. There are differences in interquartile ranges, as well as the last values of the upper whiskers, while the difference between the values of the initial whiskers are not significant. This indicates that distributions are biased.

To find significant differences between these variables the Kruskal–Wallis test was applied, with a value of 68.062, and p-value = 0.3408 all means being equal.

Possession time is a success indicator identified in several works, the objective is to determine possession time and field zone in which possession occurs since having possession in the middle defensive zone does not necessarily mean more success since the ball is away from the rival goal.

Table 7 provides an overview of use frequency of the various zones of the field and mean durations of possession in different zones, depending on successful and unsuccessful teams.

Possession zone changes significantly between groups, x^2 = 15.72, p < 0.05. Specifically, successful teams occupied
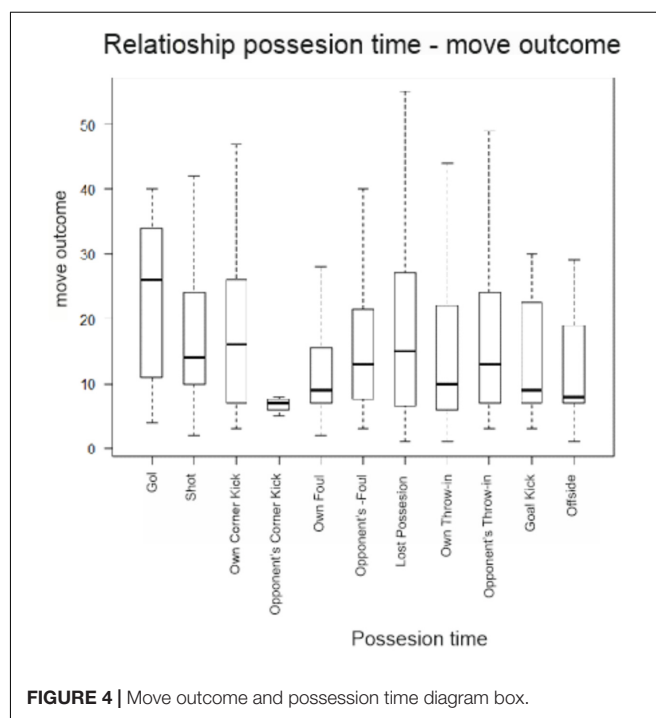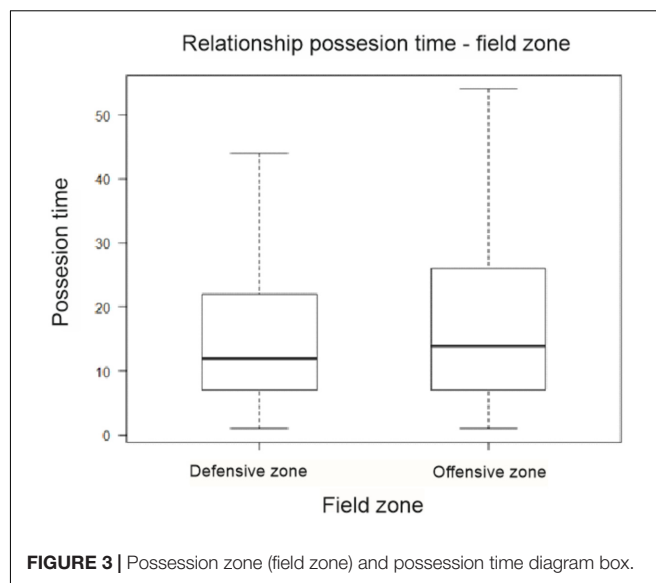
more frequently the middle offensive zone than the unsuccessful (712 times against 588, respectively). On the other hand, the unsuccessful teams occupied a greater number of times the middle defensive zone than the successful teams (578 vs. 406).

Observation indicates that successful teams spend more time in the middle offensive zone, and is accompanied by a greater possession time (20.23 s) as contrary to the middle defensive zone (15.82 s). While unsuccessful teams spend more time inside the middle defensive zone, with a longer possession time (14.23 s) than in the middle offensive zone (12.74 s).

TABLE 6 | Descriptive possession time – possession zone.

	Middle defensive N = 1.053	Middle offensive N = 1.231	p_overall
Possession time	16.0 m ± 13.5 m	18.6 m ± 17.0 m	0.04

TABLE 7 | Relationship between group, zone, and possession time.

Group	Field zone	Fr.	$\bar{x}$ PT*
Successful	Middle defensive	406	15.82
	Middle offensive	712	20.23
Unsuccessful	Middle defensive	578	14.23
	Middle offensive	588	12.74

Possession time.
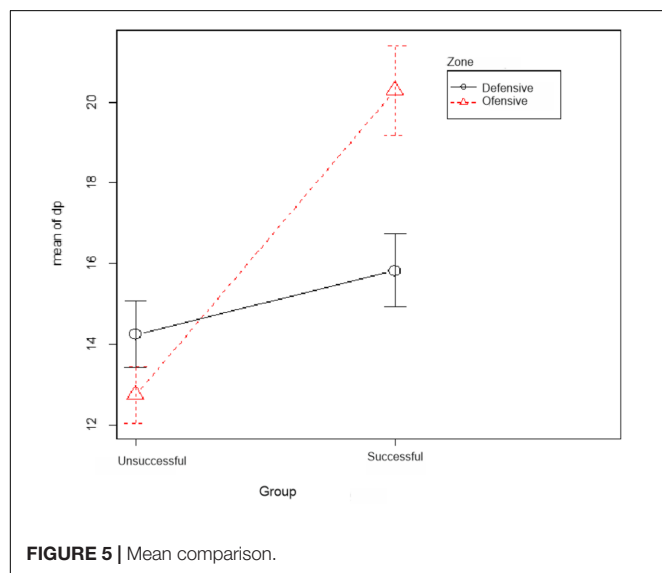
FIGURE 5 | Mean comparison.

Intensity determined by association coefficients Cramer's V 0.13, and Chuprov coefficient T^2 0.13 used to measure symmetrical association between variables showed low intensity relationship between variables.

Significant differences were also found in possession time between successful and unsuccessful teams, $H = 24.289$, $p < 0.001$. To study the relationship between possession time and match outcome Welch's T was used.

In this case, statistic $t = 5.408$, $p < 0.001$ with a confidence interval of 95% between -7.072 and -3.305, and 18.67 in successful and 13.48 in unsuccessful teams.

Observing the three variables (**Figure 5**) shows that in the successful teams attack patterns, teams stay longer in the middle offensive zone with a longer possession time than unsuccessful teams, while unsuccessful teams stay longer in the middle defensive zone with longer times in possession.

Size effect was measured applying a r_{bp}, formula quoted by Nakagawa and Cuthill (2007) with a value of 0.32. Positive coefficient indicates that high scores on possession time implies greater success for the team, although with small intensity.

To determine the influence of possession time and possession zone in successful and unsuccessful teams a logistic regression model was used (**Table 8**).

Successful/unsuccessful = possession zone + possession time

Team success increases 1.72 times when playing on the middle offensive zone against the middle defensive zone in the one

variable model. The two variable model shows an increase of 1.67. Possession time didn't show significant differences between successful and unsuccessful teams.

Probability of being successful in explanatory variable terms, with X_1 being field zone and X_2 possession time is:

$$P\left(Exitosos\right) = \frac{\exp^{(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2)}}{1 + \exp^{(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2)}} = 0.4425 \qquad (1)$$

44.25% is the probability of a team being successful.

Some authors (Hair et al., 1999) recommend using several methods to evaluate the model's goodness of fit. A value of Hosmer–Lemeshow 0.797 indicates goodness of fit (**Figure 6**).

Evaluation model based on pseudo-coefficients R^2, CoxSnell = 0.048, adjusted Nagelkerke = 0.065, Adj. McFadden = 0.036 and Tjur = 0, indicate low prediction. In classification terms, the model has 47.90% of sensitivity and 69.91% of specificity. UAC = 0.62 value with ratio of 1.23 shows an average ability to classify.

DISCUSSION

The main objective of this study was to determine whether possession time and possession zone are performance indicators that distinguish the successful football elite teams from the unsuccessful. What differentiates this study from its precedents is that it has tried to control some of the situational variables identified by previous studies as an influence to ball possession. Specifically, all analyzed matches were played at neutral grounds and team level was similar having into account that the teams were the best European national teams. The match status, the other situational variable identified as influential in possession time, was also analyzed to observe level of influence. In order to generalize results we have not studied only one team but analyzed several national teams in the same competition.

Study results have allowed us to detect significant differences between possession time and match status. Longest possession time occurs when teams are winning, these results are similar to those reported by Bloomfield et al. (2005b) and Taylor et al. (2008) and contradict those found by Jones et al. (2004), Lago and Martín (2007), Lago (2009), and Lago-Peñas and Dellal (2010) who indicate that teams losing or drawing have longer periods of possession. Multiple factors may explain these differences, such as playstyles adopted by the teams during competitions, since behavior may be different depending on whether it is a national competition or an international tournament, with different national teams. It has been shown that the main differences are found between the result of winning–losing and losing–drawing, while possession time does not change when teams are either winning or drawing. These findings indicate a team tendency to not change either their playstyle nor their game pattern according to the match status, using the same strategy despite the score, and teams characterized by an attack pattern of long possessions shall not change to make counterattacks or direct attacks when they get ahead on the scoreboard, but try to keep the lead through ball possession, and teams with short attacks will not change to make long attacks

	Estimate	SE	Z	Pr > |z|	Crude OR (95% CI)	Adj OR (95% CI)
(Intercept)	−0.77	0.13	5.74	$9.19e − 09$***		
[T.Offensive]	0.51	0.13	3.67	0.01***	1.72 (1.31, 2.25)	1.67 (1.27, 2.19)
Dp	0.03	0.01	5.00	$5.73e − 07$ **	1.03 (1.02, 1.04)	1.03 (1.02, 1.04)

**$p < 0.05$; ***$p < 0.001$.
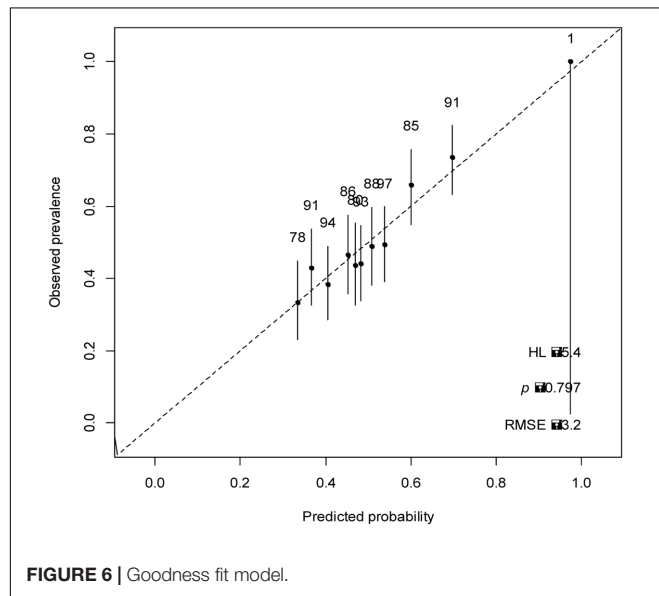
FIGURE 6 | Goodness fit model.

when changing a favorable marker to an adverse one. However, to consolidate these data, it would be interesting that future studies considered not only the encounter's partial result, but also score difference, since possibly a team that is losing by the difference of two or more goals and has little time left for the end of the match, regardless of their style of play, will make short possessions to try to increase the frequency of finalizations. On the other hand, teams that are winning by the difference of two or more goals, lacking a short time to the end of the encounter, although their style of offensive game is characterized by short possessions, surely will try to increase the possession time not allowing the opposing team to create chances of finalization.

Results also suggest that possession time is slightly higher in the first part of the match. This can be explained due to the fact that in the second half of the match there is a greater accumulation of fatigue and therefore the player will have a lower technical and tactical level, which will cause a greater number of errors in the technical executions and the tactical decisions, consequently producing a greater number of ball losses and possession changes. The offensive game actions ending with a goal or shot are those with a longer possession time. Data is consistent with results found in the study of Casal et al. (2015) who suggest that long possessions offer a greater chance of successful outcomes. Regarding the area of the field, results show that most of the time possessions are located in the middle offensive zone, this data is consistent with results obtained by Collet (2013) which indicated the need for effective possessions,

meaning these possessions should be located in dangerous places for the opponent's team, for example, near the opponent's goal. Effectively, for possession to be effective, it must occur as close as possible to the opposing goal, trying to disrupt the opposing team's defense and create a chance of finalization. Ball possession happening far from the rival team's goal and without intention to progress is totally ineffective.

Having into account team quality the bivariate analysis has allowed to draw several evidence on the possession type of different teams. Specifically, we detected significant differences in spatial occupation frequency that teams perform, successful teams occupy a greater number of times the middle offensive zone and for longer times, on the contrary, unsuccessful teams occupy most often the middle defensive zone. Data is consistent with results obtained by Bate (1988) indicating that probability of scoring a goal depends on the number of times a team gets close to the opposite goal having possession, this being an indicator of successful teams. It seems obvious that the most advantageous possession zones are those close to the goal zones of the opposing team and that maintenance of possession in zones far from the goal don't guarantee offensive success.

Results reinforce the established by various studies (Andujar, 2015; Casal et al., 2015) showing that modern playstyle has evolved into a positional game in which possession is the fundamental argument in collective game. Playstyle has changed from a model that produced success, identified by shooting to the opposite goal, thanks to the turnovers in the middle offensive zone and short possessions, to one in which once possession has started the attacking phase becomes elaborate and parsimonious.

Significant differences were found in possession time between different group teams. Successful teams has longer possession times than unsuccessful teams. Results agree with other studies (Jones et al., 2004; Bloomfield et al., 2005a; Hughes and Franks, 2005; Lago and Martín, 2007; Lago-Peñas and Dellal, 2010) indicating that greater possession characterizes successful teams. This fact is reflected in the playstyle the best teams of both European domestic leagues and national teams, both European and worldwide are using today. The F.C. Barcelona (2015–2016 Spanish Liga and 2015 UEFA Champions League champions), Spain's national team (2008 and 2012 UEFA Euro 2010 FIFA World Cup champions) and German's national team (2014 FIFA World Cup champions) are characterized by an offensive playstyle that consists of taking the lead through possession, using as an overall tactical offensive model, the combination attack.

Multivariate analysis tried to describe the relationship between possession time in each zone with team successfulness. Results showed that successful teams differ significantly from

unsuccessful teams in this regard. Specifically, successful teams occupied more frequently the middle offensive zone and remained longer in the same keeping possession. On the contrary, unsuccessful teams occupied more times and middle defensive zone staying longer times than successful teams. Results reinforce those obtained with bivariate analysis which agreed with the importance of being near the goal and having long possessions to ensure team success.

Logistic regression analysis allowed us to determine possession time and zone influence on the outcome of the match, as well as identifying a model that allows us to predict team success in terms of these variables. This model indicates that each time that a possession is carried out in the middle offensive zone, the chances of winning will increase 1.72 times and the probability of success having longer possession times on the middle offensive zone will be 44.25%. These data are the main potentiality of the present work since no previous investigations have been found that carry out this type of analysis, studying the relationship between possession time, possession zone and team's successfulness, in order to identify a game pattern with greater success.

One limitation of the study has been that only national teams matches have been analyzed, so results cannot be extrapolated to other kinds of meetings, because as indicated by studies James et al. (2004), Tucker et al. (2005), Bloomfield et al. (2005a), Lago and Martín (2007), Lago (2009), and Collet (2013) the type of competition and, in particular, quality of the rival team, influences the type of possession that will be carried out in the meeting. On the other hand, we also believe that the fact of having as sample teams of an even competitive level (being the best in Europe) is work's fortitude and, if significant differences were found between the teams, all of a similar competitive level, it is feasible that between different level teams differences will be even greater. We are also aware of the existence of other extraneous variables that can influence the results as the playstyle in different competitions (Rienzi et al., 2000), arbitration decisions, weather conditions or the state of the field, but it would be impossible controlling all of these variables, so this study has tried to show the influence of some of them.

Results obtained are expected to help giving more knowledge about successful offensive game models, as well as performance factors of the offensive phase, which will allow teams to optimize their training process and performance during the match. In the field of research contributions could prove useful in future studies of possessions, taking into consideration not only possession time but also the area in which it occurs and team quality.

CONCLUSION

This study allows us to identify, characterize and differentiate different attack patterns between successful and unsuccessful teams, based on possession time and zone in which it occurs. Results show that significant differences between the two groups are found. Data establishes that successful teams are characterized by an offensive game pattern with greater possession and more presence in the middle offensive zone. On the other hand, unsuccessful teams have shown an offensive game pattern with lesser possession time. In addition, longer possession time in the middle offensive zone, predicts greater chance of victory in the match.

Current football's empirical observation and analysis leads to the identification of a possession playstyle generalized commitment, it seems that coaches and teams have opted for this model, but what makes teams have higher success rates than others? Probably the answer is related to the individual effectiveness of the actors (players) in the collective framework. We can never forget that individualities build the collective game and, therefore, individual quality of the players is a key factor of performance, which will mark the collective success of the teams.

AUTHOR CONTRIBUTIONS

CAC developed the project, review the literature and wrote the manuscript. JLL was responsible for performed analysis, the method section and revised the content critically. RM and TA collected and analyzed the data and supervised the drafting of the manuscript. FJM translated the manuscript. All authors approved the final, submitted version of the manuscript.

ACKNOWLEDGMENTS

REFERENCES

Acar, M. F., Yapicioglu, B., Arikan, N., Yalcin, S., Ates, N., and Ergun, M. (2009). "Analysis of goals scored in the 2006 world cup," in *The Proceeding of the Sixth World Congress on Science and Football, Science and Football, VI*, eds T. Reilly and F. Korkusuz (London: Routledge), 233–242.

Allen, R. J. (2003). The error of expected loss minimization. *Law Probab. Risk* 2, 1–7. doi: 10.1093/lpr/2.1.1

Andujar, M. A. (2015). *La Transición Defensiva en el Fútbol de Élite. Análisis de la Copa Mundial de la FIFA Sudáfrica 2010*. A Coruña: Universidad de A Coruña.

Anguera, M. T. (1990). "Metodología observacional," in *Metodología de la Investigación en Ciencias del Comportamiento*, eds J. Arnau, M. T. Anguera, and J. Gómez (Murcia: Secretariado de Publicaciones de la Universidad de Murcia), 125–236.

Anguera, M. T. (1993). "Proceso de categorización," in *Metodología Observacional en la Investigación Psicológica*, Vol. I, ed. M. T.

Anguera (Barcelona: Promociones Publicaciones Universitarias), 115–167.

Anguera, M. T., Blanco-Villaseñor, A., Hernández-Mendo, A., and Losada J. L. (2011). Diseños observacionales: ajuste y aplicación en psicología del deporte. *Cuad. Psicol. Deporte* 11, 63–76.

Anguera, M. T., Camerino, O., Castañer, M., and Sánchez, P. (2014). Mixed methods in research into physical activity and sport. *Rev. Psicol. Deporte* 23, 123–130.

Anguera, M. T., and Mendo, A. (2013). La metodología observacional en el ámbito del deporte. *Rev. Ciencias Deporte* 9, 135–161.

Araújo, D., Davids, K. Y., and Hristovski, R. (2006). The ecological dynamics of decision making in sport. *Psychol. Sport Exerc.* 7, 653–676. doi: 10.1016/j.psychsport.2006.07.002

Ardá, T., Maneiro, R., Rial, A., Losada, J. L., and Casal, C. A. (2014). Análisis de la eficacia de los saques de esquina en la copa del mundo de fútbol 2010. Un intento de identificación de variables explicativas. *Rev. Psicol. Dep.* 23, 165–172.

Ato, M., and López, J. (1996). *Análisis Estadístico para Datos Categóricos*. Madrid: Editorial Síntesis.

Barreira, D., Garganta, J., Guimaraes, P., Machado, J. C., and Anguera, M. T. (2013). Ball recovery patterns as a performance indicator in elite soccer. *J. Sport Eng. Technol.* 228, 61–72. doi: 10.1177/1754337113493083

Bate, R. (1988). "Football chance: tactics and strategy," in *Science and Football*, eds T. Reilly, A. Lees, and W. J. Murphy (London: E. & F.N. Spon), 293–301.

Berk, R. (1979). Generalizability of behavioral observations: a clarification of interobserver agreement and interobserver reliability. *Am. J. Ment. Defic.* 83, 460–472.

Bloomfield, J. R., Polman, R. C. J., and O'Donoghue, P. G. (2005a). Effects of score-line on intensity of play in midfield and forward players in the FA Premier League. *J. Sport Sci.* 23, 191–192.

Bloomfield, J. R., Polman, R. C. J., and O'Donoghue, P. G. (2005b). Effects of score-line on team strategies in FA Premier League Soccer. *J. Sport Sci.* 23, 192–193.

Borrás, D., and Sainz de Baranda, P. (2005). Análisis de córner en función del momento del partido en el mundial de Corea y Japón 2002. *Cult. Cien. Deporte* 2, 87–93.

Carling, C., Reilly, T., and Williams, A. M. (2009). *Performance Assessment for Field Sports*. London: Routledge.

Carling, C., Williams, A. M., and Reilly, T. (2005). *Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance*. Abingdon: Routledge. doi: 10.4324/9780203448625

Carmichael, F., Thomas, D., and Ward, R. (2001). Production and efficiency in association football. *J. Sports Econ.* 2, 228–243. doi: 10.1177/152700250100200303

Casal, C. A. (2011). *Cómo Mejorar la Fase Ofensiva en el Fútbol: Las Transiciones ofensivas*. Alemania: Editorial Académica Española.

Casal, C. A., Losada, J. L., and Ardá, T. (2015). Análisis de los factores de rendimiento de las transiciones ofensivas en el fútbol de alto nivel. *Rev. Psicol. Deporte* 24, 103–110.

Casal, C. A., Maneiro, R., Ardá, T., Losada, J. L., and Rial, A. (2014). Effectiveness of indirect free kicks in elite soccer. *Int. J. Perform. Anal. Sport* 14, 744–750.

Castellano, J. (2000). *Observación y Análisis de la Acción de Juego en el Fútbol*. Vitoria-Gasteiz: Universidad del País Vasco.

Collet, C. (2013). The possession game? A comparative analysis of ball retention and team success in European and international football, 2007–2010. *J. Sport. Sci.* 31, 123–136. doi: 10.1080/02640414.2012.7272455

Davids, K., Araújo, D., and Shuttleworth, R. (2005). "Applications of dynamical systems theory to football," in *Science and Football V*, eds T. Reilly, J. Cabri, and D. Araujo (London: Routledge), 537–550.

Dawson, P., Dobson, S., and Gerrard, B. (2000). Stochastic frontiers and the temporal structure of managerial efficiency in English soccer. *J. Sports Econ.* 1, 24–32. doi: 10.1177/152700250000100402

De Oliveira, P. (2012). A Influencia das Variáveis Situacionais no Comportamento Posicional de uma equipa Profissional de Futebol no Instante de Recuperação da Posse da Bola. Um Estudo de Caso. Pontevedra: Universidade de Vigo.

Ensum, J., Williams, M., and Grant, A. (2000). An analysis of attacking set plays in Euro 2000. *Insight* 4, 36–40.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions*. New York, NY: John Wiley and Sons.

Garganta, J. (2000). Análisis del juego del fútbol. El recorrido evolutivo de las concepciones, métodos e instrumentos. *Rev. Entrenamiento Deportivo* 14, 6–13.

Grant, A. G., Williams, A. M., and Reilly, T. (1999). An analysis of the successful and unsuccessful teams in the 1998 World Cup. *J. Sport Sci.* 17, 827.

Gréhaigne, J. F. (2001). *La Organización del Juego en el Fútbol*. Barcelona: Editorial Inde SA.

Hadley, L., Poitras, M., Ruggiero, J., and Knowles, S. (2000). Performance evaluation of national football league teams. *Managerial and Decision Economics* 21, 45–56. doi: 10.1002/1099-1468(200003)21:2<63::AID-MDE964>3.0.CO;2-O

Hair, J., Anderson, R., Tatham, R., and Black, W. (1999). *Análisis Multivariante*. Madrid: Prentice Hall.

Hook, C., and Hughes, M. (2001). "Patterns of play leading to shots in 'euro 2000,'" in *Center for Performance Analysis*, ed. Pass.com (Cardif: UWIC).

Hughes, M., and Franks, I. (2005). Analysis of passing sequences, shots and goals in soccer. *J. Sport. Sci.* 23, 509–514. doi: 10.1080/02640410410001716779

Hughes, M. D. (2003). "Notational analysis," in *Science and Football*, eds T. Reilly and M. Williams (London: Routledge), 245–264.

Hughes, M. D., and Bartlett, R. M. (2002). The use of performance indicators in performance analysis. *J. Sport Sci.* 20, 739–754. doi: 10.1080/026404102320675602

James, N., Jones, P. D., and Mellalieu, S. (2004). Possession as a performance indicator in soccer. *Int. J. Perform. Anal. Sport* 4, 98–102.

James, N., Mellalieu, S., and Hollely, C. (2002). Analysis of strategies in soccer as a function of European and domestic competition. *Int. J. Perform. Anal. Sport* 2, 85–103.

Jones, P. D., James, N., and Mellalieu, S. (2004). Possession as a performance indicator in soccer. *Int. J. Perform. Anal. Sport* 4, 98–102.

Kormelink, H., and Seeverens, T. (1999). *Match Analysis and Game Preparation*. Spring City, PA: Reedswain.

Lago, C. (2009). The influence of match location, quality of opposition, and match status on possession strategies in professional association football. *J. Sport Sci.* 27, 1463–1469. doi: 10.1080/02640410903131681

Lago, C. (2012). The role of situational variables in analysing physical performance in soccer. *J. Hum. Kinet.* 35, 89–95. doi: 10.2478/v10078-012-0082-9

Lago, C., and Martín, R. (2007). Determinants of possession of the ball in soccer. *J. Sport Sci.* 25, 969–974. doi: 10.1080/02640410600944626

Lago-Ballesteros, J., Lago, C., Rey, E., Casáis, L., and Domínguez, E. (2012). El éxito ofensivo en el fútbol de élite. Influencia de los modelos tácticos empleados y de las variables situacionales. Motricidad. *Eur. J. Hum. Mov.* 28, 145–170.

Lago-Peñas, C., and Dellal, A. (2010). Ball possession strategies in elite soccer according to the evolution of the match-score: the influence of situational variables. *J. Hum. Kinet.* 25, 93–100.

Lago-Peñas, C., and Lago-Ballesteros, J. (2011). Game location and team quality effects on performance profiles in professional soccer. *J. Syst. Sci. Med.* 10, 465–471.

Lago-Peñas, C., Lago-Ballesteros, J., Dellal, A., and Gómez, M. (2010). Game-related statistics that discriminated winning, drawing and losing teams from the Spanish soccer league. *J. Sports Sci. Med.* 9, 288–293.

Lames, M., and McGarry, T. (2007). On the search for reliable performance indicators in game sports. *Int. J. Perform. Anal. Sport* 7, 62–79.

Losada, J. L., and Manolov, R. (2014). The process of basic training, applied training, maintaining the performance of an observer. *Qual. Quant.* 49:339. doi: 10.1007/s11135-014-9989-7

McGarry, T., Anderson, D. I., Wallace, S. A., Hughes, M. D., and Franks, I. M. (2002). Sport competition as a dynamical self-organizing system. *J. Sport Sci.* 20, 771–781. doi: 10.1080/026404102320675620

McGarry, T., and Franks, I. (2003). "The science of match analysis," in *Science and Football V*, eds T. Reilly, J. Cabri, and D. Araujo (Londres: Routledge), 265–275.

Mckenzie, R., and Cushion, C. (2012). Performance analysis in football: A critical review and implications for future research. *J. Sport. Sci.* 12, 1–38.

Mitchell, S. (1979). Interobserver agreement, reliability and generalizability of data collected in observational studies. *Psychol. Butl.* 86, 376–390. doi: 10.1037/0033-2909.86.2.376

Mohammad, M., Kourosh, D., and Farhad, A. (2016). Ball recovery consistency as a performance indicator in elite soccer. *Rev. Bras. Cineantropometr. Hum.* 18, 72–81. doi: 10.5007/1980-0037.2016v18n1p72

Nakagawa, S., and Cuthill, I. (2007). Effect size, confidence interval and statistical significance: a practical guide for biolo-gists. *Biol. Rev.* 82, 591–605. doi: 10.1111/j.1469-185X.2007.00027.x

O'Donoghue, P. (2010). *Research Methods in Performance Analysis in Sport*. London: Routledge.

Perl, J. (2006). Qualitative analysis of team interaction in games by means of the load-performance-metamodel PerPot. *Int. J. Perform. Anal. Sports* 6, 34–51.

Ric, A., Torrents, C., Gonçalves, B., Sampaio, J. E., and Hristoski, R. (2016). Soft-assembled multilevel dynamics of tactical behaviors in soccer. *Front. Psychol.* 7:1513. doi: 10.3389/fpsyg.2016.01513

Rienzi, E., Drust, B., Reilly, T., Carter, J., and Martin, A. (2000). Investigation of antrohropometric and woed-rate profles of elite South American international soccer players. *J. Sport Med. Fitness* 40, 162–169.

Sainz de Baranda, P., and López-Riquelme, D. (2012). Analysis of corner kicks in relation to match status in the 2006 World Cup. *Eur. J. Sport Sci.* 12, 121–129. doi: 10.1080/17461391.2010.551418

Sánchez Flores, J., García Manso, J. M., Martín, J. M., Ramos, E., Arriaza, E., and Da Silva, M. E. (2012). Análisis y evaluación del lanzamiento de esquina (córner) en el fútbol de alto nivel. *Rev. Andaluza Med. Deporte* 5, 140–146. doi: 10.1016/S1888-7546(12)70022-9

Sasaki, Y., Nevill, A., and Reilly, T. (1999). Home advantage: a case study of Ipswich Town football club during the 1996–97 season. *J. Sport Sci.* 17, 831.

Stanhope, J. (2001). "An investigation into possession with respect to time in the Soccer World Cup 1994," in *Notational Analysis of Sport III*, ed. M. D. Hughes (Cardiff: UWIC), 155–162.

Taylor, J. B., James, N., and Mellalieu, S. D. (2005). "Notational analysis of corner kicks in english premier league soccer," in *Science and Football V*, eds T. Reilly, J. Cabri, and D. Araujo (Londres: Routledge), 225–230.

Taylor, J. B., Mellalieu, S., James, N., and Shearer, D. A. (2008). The influence of match location, quality of opposition, and match status on technical performance in professional association football. *J. Sport Sci.* 26, 885–895. doi: 10.1080/02640410701836887

Tenga, A., and Sigmundstad, E. (2011). Characteristics of goalscoring possessions in open play: comparing the top, in between and bottom teams from professional soccer league. *Int. J. Perform. Anal. Sport* 11, 545–552.

The American Psychological Association's [APA's] (2010). *Ethical Principles of Psychologists and Code of Conduct*. Washington, DC: The American Psychological Association's.

Tucker, W., Mellalieu, S., James, N., and Taylor, J. B. (2005). Game location effects in professional soccer: a case study. *Int. J. Perform. Anal. Sport* 5, 23–35.

Vogelbein, M., Noops, S., and Hökelmann, A. (2014). Defensive transition in soccer are prompt possession regains a measure of success? A quantitative analisys of German Fuβball-Bundesliga 2010/2011. *J. Sport Sci.* 32, 1076–1083. doi: 10.1080/02640414.2013.879671

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer EP and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Mastery in Goal Scoring, T-Pattern Detection, and Polar Coordinate Analysis of Motor Skills Used by Lionel Messi and Cristiano Ronaldo

Marta Castañer[1], Daniel Barreira[2], Oleguer Camerino[1]*, M. Teresa Anguera[3], Tiago Fernandes[2] and Raúl Hileno[1]

[1] National Institute of Physical Education of Catalonia, Observation Laboratory in Physical Activity and Sports, University of Lleida, Lleida, Spain, [2] Faculty of Sport, Centre of Research, Training, Innovation and Intervention in Sport, University of Porto, Porto, Portugal, [3] Faculty of Psychology, University of Barcelona, Barcelona, Spain

Research in soccer has traditionally given more weight to players' technical and tactical skills, but few studies have analyzed the motor skills that underpin specific motor actions. The objective of this study was to investigate the style of play of the world's top soccer players, Cristiano Ronaldo and Lionel Messi, and how they use their motor skills in attacking actions that result in a goal. We used and improved the easy-to-use observation instrument (OSMOS-soccer player) with 9 criteria, each one expanded to build 50 categories. Associations between these categories were investigated by T-pattern detection and polar coordinate analysis. T-pattern analysis detects temporal structures of complex behavioral sequences composed of simpler or directly distinguishable events within specified observation periods (time point series). Polar coordinate analysis involves the application of a complex procedure to provide a vector map of interrelated behaviors obtained from prospective and retrospective sequential analysis. The T-patterns showed that for both players the combined criteria were mainly between the different aspects of motor skills, namely the use of lower limbs, contact with the ball using the outside of the foot, locomotion, body orientation with respect to the opponent goal line, and the criteria of technical actions and the right midfield. Polar coordinate analysis detected significant associations between the same criteria included in the T-patterns as well as the criteria of turning the body, numerical equality with no pressure, and relative numerical superiority.

Keywords: soccer, goal scoring, motor skills, pattern detection, polar coordinate analysis

INTRODUCTION

Soccer performance research is broadly developed and implemented (Ali, 2011), contributing to a rapid and continuous enhancement of players' performance in the last few years (Lago-Ballesteros et al., 2012). Given the complexities and dynamic nature of soccer, observation and measurement processes throughout the design of match analysis systems have made it possible to collect data by embracing technical, behavioral, physical, and tactical factors (Carling et al., 2005).

Due to the high complexity of soccer games, it is known that general research in soccer presents some flaws, such as: (i) lack of context; (ii) missed operational definitions (MacKenzie and Cushion, 2013); and (iii) inability of parameters such as official match statistics and physiological and

performance data to provide information for a comprehensive evaluation of the soccer players (Perić et al., 2013). The idiosyncrasies of dynamic systems theory make it possible to overcome these limitations (Glazier and Robins, 2013), entailing that mathematical models of analysis must incorporate a wider range of organismic, environmental, and task constraints (Glazier and Davids, 2009). Specifically, dynamical systems theory plays an important role with its multidisciplinary theoretical framework for sports performance analysis by (i) facilitating the linkage of behaviors to outcomes due to its more process-oriented than product-oriented focus, and (ii) stabilizing the same principles and concepts governing patterns in intra- and inter-individual levels of sports performance (Glazier and Robins, 2013).

There is a general belief that talented people display superior performance in a wide range of activities, such as superior athletic ability and mental abilities (Feltovich et al., 2006). Notwithstanding, to understand sport expertise, multi-scale and multi-disciplinary theoretical descriptions are needed (Araújo et al., 2010). In the domain of team play analysis, McGarry et al. (2002) mention that the main soccer research focuses on tactical and technical factors. Technical analysis includes the testing of key sport skills, including the mechanical aspects of technique, and is concerned with the way the skill is performed in terms of kinetic and kinematic detail of the movement involved (O'Donoghue, 2010). In the perspective of Ali (2011), it becomes useless if the player does not perform the right action at the right time, i.e., when a tactical approach to the players' behavior does not exist. Behavioral specific and representative information is continuously apprehended from the environment by dynamical movement systems, to structure and to adapt functional patterns of play. This sensibility to contextual information regulates the motor system number of biomechanical degrees of freedom, however, more critical than attending each behavior separately, is to form and

to develop functional synergies that arise between parts of the body used to achieve movement goals (Davids et al., 2006).

However, with the application of styles of play that incorporate and encourage individual actions and skills, which improve overall game strategies and outcomes (Carling et al., 2008), relevant individual behaviors in soccer, such as goal-scoring, need to be analyzed with regard to motor skills (Castañer et al., 2016b). Although, goal-scoring, the ultimate objective of attacking effectiveness in competition settings, has been extensively used in match performance research (Tenga et al., 2010; Lago-Ballesteros et al., 2012), the objectivity of this research remains insufficient with regard to the motor skills that support goal-scoring patterns (Castañer et al., 2016a).

Indeed, in elite soccer, the use of motor skills has largely been studied from a subjective perspective (Duch et al., 2010), but mastery of these skills (Castañer et al., 2009, 2016a; Wallace and Norton, 2014) is directly linked to motor versatility (Bishop et al., 2013) and consequently to the ability to execute complex intentional actions (Memmert et al., 2013). Motor versatility in both individual and team sports requires the integration of multiple skills (Bishop et al., 2013); it is a particularly important quality in attackers such as strikers and wingers and is closely linked to motor anticipation (Murgia et al., 2014). In fact, the ability to efficiently and effectively execute skilled movement patterns is the most important aspect of soccer performance and players must apply cognitive, perceptual and motor skills to rapidly changing situations (Ali, 2011). These multiple skills are essential to execute soccer moves such as ball control, dribbling, and shots. Motor skills involve axial movements in the form of turns and pivots, spatial orientation of the player's body in relation to the side lines and goal line, and the use of one limb or another (laterality). These movements not only underpin all soccer moves but also contribute to the uniqueness of each player (Castañer et al., 2016a). In addition, most of these movements are interlinked. Laterality (Teixeira et al., 2011; Bishop et al., 2013), for example, refers not only to left-right preference but also to how a player orients his body spatially (Bishop et al., 2013; Loffing et al., 2015). Previous research (Castañer et al., 2016a) has demonstrated that Lionel Messi—a left-footed player whose has achieved some of his best results playing on the right wing—is a good example of laterality. Cristiano Ronaldo does not have the singular characteristic of being left-footed in goal-scoring, but he is also an example of motor skills versatility. This is the main aspect of our research: studying the motor skills that configure the uniqueness of a striker.

Cristiano Ronaldo and Lionel Messi are considered to be the best soccer players who have ever existed. Since 2008, no other player has won the FIFA best player award: Messi has won 5 times and Ronaldo 4 times. In a comparison of Ronaldo's and Messi's goal-scoring in *La Liga* since the 2009–2010 season, Shergold (2016) found that Messi scored 270 goals in 252 matches, playing 21,218 min and taking 953 shots, and Ronaldo scored 270 goals in 247 matches, playing 21,206 min, and taking 1,318 shots. These data show that Messi has a shot conversion rate of 28.77%, compared with 20.03% for Ronaldo. Nevertheless, both players show unusual accuracy as well as uniqueness in motor skills.

TABLE 1 | Goals scored by Lionel Messi and Cristiano Ronaldo considered.

Competition	Season	Lionel Messi	Cristiano Ronaldo	Total
Champions League	2013–2014	5	13	18
	2014–2015	10	7	17
	2015–2016	5	12	17
	Total	20	32	52
La Liga	2013–2014	15	17	32
	2014–2015	26	26	52
	2015–2016	11	22	33
	Total	52	65	117
Copa del Rey	2013–2014	4	0	4
	2014–2015	3	1	4
	2015–2016	4	0	4
	Total	11	1	12
	Total	83	98	181

TABLE 2 | OSMOS-soccer player (Observation system for motor skills in soccer).

Criterion	Category	Code	Description
1. Body part	Left foot	LF	Player touches the ball with left foot
	Right foot	RF	Player touches the ball with right foot
	Left leg	LL	Player touches the ball with left leg (not including foot)
	Right leg	RL	Player touches the ball with right leg (not including foot)
	Chest	CH	Player touches the ball with chest
	Back	BA	Player touches the ball with back
	Head	HD	Player touches the ball with head
2. Foot contact zone	Tip	TI	Player touches the ball with tip of foot
	Outside	OU	Player touches the ball with outside of foot
	Inside	ID	Player touches the ball with inside of foot
	Heel	HL	Player touches the ball with heel
	Sole	SO	Player touches the ball with sole
	Instep	IT	Player touches the ball with instep
	Non-observable	NON	No clear contact zone between player and ball
3. Body orientation with respect to rival goal line	Facing goal	FG	Player's chest facing rival goal line
	Facing right	OR	Player's chest facing right side line
	Back to goal	BT	Player's back facing rival goal line
	Facing left	OL	Player's chest facing left side line
4. Stability	Right turn	RT	Player makes a full or half turn to the right (vertical axis)
	Left turn	LT	Player makes a full or half turn to the left (vertical axis)
	Right foot pivot	RFP	Player pivots to the right on right foot
	Left foot pivot	LFP	Player pivots to the left on left foot
	Jump	JUM	Elevation of the body
5. Locomotion	One	ONE	Player takes one step without touching the ball
	Two	TWO	Player takes two steps without touching the ball
	Three	THR	Player takes three steps without touching the ball
	Four	FOU	Player takes four steps without touching the ball
	Five	FIV	Player takes five steps without touching the ball
	More	MOR	Player takes more than five steps without touching the ball
6. Technical actions	Control	CT	Player gains control of the ball following diverse actions
	Dribbling	CD	Player dribbles the ball
	Shot	SH	Player shoots
	Feint (shot)	SHF	Player pretends to shoot.
	Feint (pass)	PAF	Player pretends to pass
	Feint (change of dir.)	DIF	Player tricks a defender by changing direction
	Volley	VO	Player makes contact with the ball before it touches the ground
7. Centre of the game	Relative numerical inferiority	PR	Attacking team has one or two influent players less than the opponent in the centre of the game
	Absolute numerical inferiority	PA	Attacking team has at least less three or more influent players in relation with the opponent in the centre of the game
	Numerical equality with pressure	PE	Attacking team has the same number of players than the opponent in the centre of the game. The ball carrier has his back oriented to the opponent's goal and an opponent is marking from behind
	Numerical equality with no pressure	NPE	Attacking team has the same number of players than the opponent in the centre of the game. The ball carrier has his chest oriented to the opponent's goal, with conditions to progress into the pitch offensive zones

(Continued)

TABLE 2 | Continued

Criterion	Category	Code	Description
	Relative numerical superiority	NPR	Attacking has one or two influent players more than the opponent in the centre of the game
	Absolute numerical superiority	NPA	Attacking team has three or more influent players than the opponent in the centre of the game
8. Side	Right wing	RW	Part of the pitch between the right side line and the right midfield
	Right midfield	CR	Part of the pitch between the left midfield and the right side line
	Left midfield	CL	Part of the pitch between the right midfield and the left side line
	Left wing	LW	Part of the pitch between the left side line and the left midfield
9. Zones	Ultraoffensive 1	UOO	Between the goal line and the front of the goal area
	Ultraoffensive 2	UOT	Between the front of the goal area and the penalty box
	Offensive	OFF	Between the front of the penalty box and the half-way line (excl. centre circle)
	Central	CEN	Centre circle

TABLE 3 | Reliability: sample and values per player and for both players.

	Lionel Messi			Cristiano Ronaldo			Total		
	n	%[a]	Kappa	n	%[a]	Kappa	n	%[a]	Kappa
Inter-observer	25	30	0.95	30	31	0.94	55	30	0.95
Intra-observer	15	18	0.98	15	15	0.97	30	17	0.98

n = number of goals.

[a]Percentage of goals with regard to the final sample (see **Table 1**) for both players.

For instance, Jafari and Smith (2016) hypothesized that Lionel Messi has acquired higher motor skills than most other players, and that this frees up much cognitive capacity. And Hong et al. (2012) describe the "knuckling shot" as one of the characteristics of Ronaldo.

We believe that the above-mentioned attributes, which describe two singular styles of playing soccer, have not been analyzed from an objective, scientific perspective. This sort of analysis is challenging because soccer is a complex game that requires a wide repertoire of individual skills used for the benefit of the team and characterized by constant interactions among technical, tactical, psychological, and physical factors. There are various methods for identifying an expert, for example the retrospective method. Using this method, one can determine who is an expert by looking at how well an outcome or product is received (Chi, 2006). Here we followed Hodges et al. (2006), who assume that tasks are what elucidate the underlying mechanisms that afford consistent expert performance.

Thus, the overall objective of this study was to perform an objective analysis of Lionel Messi's and Cristiano Ronaldo's use of motor skills prior to scoring a goal using two complementary methods: T-pattern analysis and polar coordinate analysis. The methodological aim was to detect temporal structures of behavior underlying the two players' styles by means of T-pattern analysis and, complementarily, to obtain an idea of the behavior in its entirety using polar coordinate analysis, whose powerful data reduction feature facilitates the interpretation of data by means of a vectorial representation of the associations detected between behaviors.

METHODS

Given that our study fulfilled the requisite, established by Anguera (2003), of having perceivable and regular behaviors in a natural setting, we employed systematic observation (Anguera, 1979). The choice of methodology is also justified by the implementation of an *ad-hoc* observation instrument to record, analyze, and interpret the behaviors exhibited by Messi and Ronaldo in the goals analyzed.

Observational methodology offers eight types of observational designs (Blanco-Villaseñor et al., 2003; Sánchez-Algarra and Anguera, 2013; Anguera and Hernández-Mendo, 2014; Portell et al., 2015) that offer different possibilities in terms of the number of participants, the continuity of the recording and the number of criteria observed. These designs have been widely applied in the analysis of individual and team sports (Jonsson et al., 2006; Fernández et al., 2009; Camerino et al., 2012a,b; Lapresa et al., 2013; Castañer et al., 2016a; Tarragó et al., 2016); in the analysis of motor skills in physical activity and sport (Castañer et al., 2009, 2016b) and in mixed methods research in sports (Anguera et al., 2014). We decided to use the N/S/M design, where N refers to nomothetic (focusing on two players), S refers to intersessional follow-up (analyzing specific motor skills and contextual aspects recorded from the beginning to the end of different sequences of numerous matches), and M refers to multidimensional (addressing multiple criteria and responses in the *ad-hoc* observation instrument designed).

Two particularly fitting techniques for the analysis of such complexity are temporal pattern (T-pattern) detection (Casarrubea et al., 2015; Magnusson et al., 2016) and polar coordinate analysis (Sackett, 1980). T-pattern detection has been successfully used in numerous studies to reveal hidden patterns underlying different soccer actions (Anguera and Jonsson, 2003;
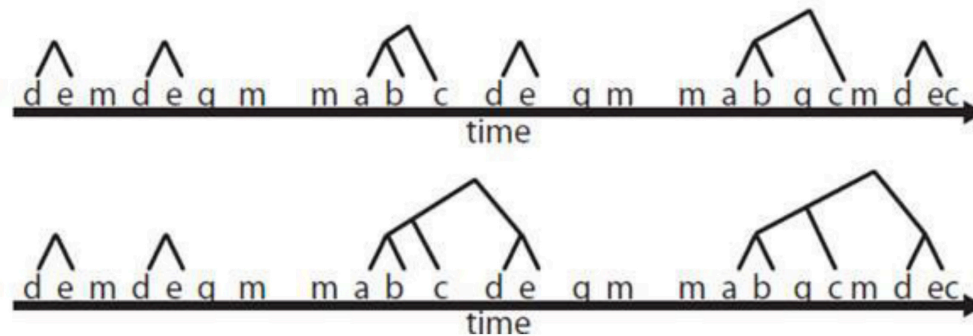
FIGURE 1 | Formation of a T-pattern from a simple T-pattern (first line) to more complex ones, such as the T-pattern at the bottom (Castañer et al., 2009, p. 859).
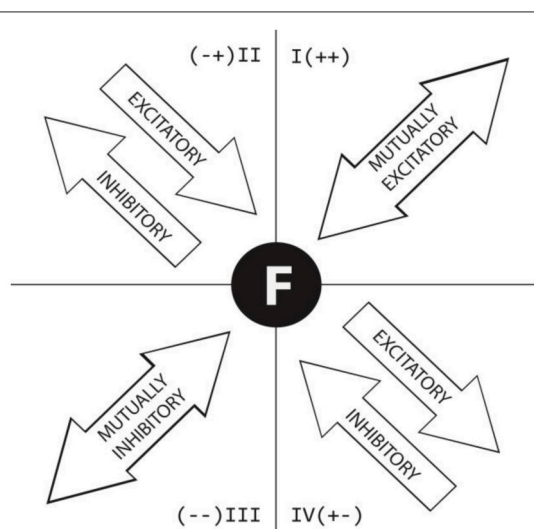
FIGURE 2 | Graphic depiction of relationships between conditional and given behaviors in polar coordinate maps according to the quadrant in which the vector is located (Castañer et al., 2016a, p. 5.).

Jonsson et al., 2006; Fernández et al., 2009; Garzón Echevarría et al., 2011; Lapresa et al., 2013, 2014; Sarmento et al., 2013; Barreira et al., 2014; Escolano-Pérez et al., 2014; Zurloni et al., 2014; Magnusson et al., 2016). Polar coordinate analysis is a powerful data reduction technique that is increasingly being used in studies of team sports (Perea et al., 2012; Robles et al., 2014; Echeazarra et al., 2015; López-López et al., 2015; Morillo-Baro et al., 2015; Sousa et al., 2015; Castañer et al., 2016a; López et al., 2016; Aragón et al., 2017). The technique provides a vectorial representation of the complex network of interrelations between carefully chosen, exhaustive and mutually exclusive defined criteria.

Participants

A total of 181 goals were analyzed, 83 scored by Lionel Messi and 98 scored by Cristiano Ronaldo (Table 1). The goals were included according to the following criteria:

i Clear observability of the sequence to the goal (Anguera and Hernández-Mendo, 2015);

ii Availability of at least two recordings of each sequence from different angles;

iii Goals scored in Champions League, *La Liga* and *Copa del Rey* were included;

iv Goals had to be from the last three seasons, namely 2013–2014, 2014–2015, and 2015–2016;

v Opponent quality criterion: Champions League is known as the most elite Union of European Football Associations (UEFA) competition, so all goals scored in this league were considered; in *La Liga* and *Copa del Rey*, the criteria of Bradley et al. (2013) were followed, i.e., only the goals scored against non-successful clubs—the last four classified at the end of each season—were not included;

vi Goals resulting directly from set pieces, including the rebound of a penalty, were not considered;

vii Goals scored in regular time were included.

Our study can thus be considered case-oriented (Sandelowski, 1996; Yin, 2014). The goals were analyzed using public television footage, in compliance with the ethical principles of the Declaration of Helsinki.

Materials
Observational Instrument

The *ad-hoc* observation instrument OSMOS-soccer player (Castañer et al., 2016a) was used with a minimal optimization of criteria. Specifically, the criterion Number of Opponents was replaced by Centre of the Game, adapted from Barreira et al. (2012, 2014, 2015), and the criterion Stability, which includes jumps, was merged with the Turn and Pivot Direction criteria. The instrument (see **Table 2**) comprised nine criteria: (1) Body Part (part of the body that the player uses to make contact with the ball); (2) Foot Contact Zone (part of the foot used to touch the ball); (3) Body Orientation (angle of the chest with respect to the side line or goal line); (4) Stability (turn direction, right vs. left; pivot foot, right vs. left; and elevation of the body); (5) Locomotion (number of steps between touches of the ball); (6) Action (common soccer technical actions); (7) Centre of the Game (number of players on both teams interacting during the

FIGURE 3 | Messi and Ronaldo event-type frequency chart.

FIGURE 4 | T-pattern shows that (a) Messi touches the ball with his left foot (LF) with the outside of this foot (OU) while facing the rival goal line (FG), tricking defenders by changing direction (DIF) in the right midfield (CR); then, (b) he continues to touch the ball with his left foot (LF) with a left body angle with respect to the rival goal line (OL) and dribbles the ball (CD) while remaining in the right midfield (CR); and then, (c) he maintains his left body angle (OL), takes three steps between touches of the ball (THR) and remains in the right midfield (CR).

striker's action); (8) Side (position of the player on the pitch); and (9) Zone (area where the player moves). Each criterion was expanded to build an exhaustive and mutually exclusive observation system that included, in total, 50 categories.

Recording Instrument

Goal-scoring sequences were coded using LINCE (v.1.2.1) (Gabín et al., 2012). This software program was also used for the data quality check.

Data Analysis Software

Two programs were used: (a) Theme software package (Magnusson et al., 2016) for T-pattern detection; (b) HOISAN v.1.6.3.2 (Hernández-Mendo et al., 2012, 2014) for the polar coordinate analysis.

Procedure

Goal-scoring sequences were analyzed from the moment the player receives the last pass to the moment he scores a goal. After appropriate training in the use of OSMOS-soccer player, two expert observers—an expert soccer analyst and a motor skills expert—recorded 30% of the total goals included for each player (**Table 3**). Intra- and inter-observer reliability was calculated in LINCE, before the full data set was coded, using a preliminary dataset of 55 and 30 goal-scoring sequences, respectively. The goals used to calculate data quality were from the 2012 to 2013 season and therefore were not included in the final sample. The resulting kappa statistic was 0.95 for inter-observer and 0.98 for intra-observer analysis, which guarantees the interpretative rigor of the coding process.

Data Analysis
T-Pattern Detection

T-pattern detection is a relevant data analysis technique in systematic observation (Anguera and Hernández-Mendo, 2015) and the THEME software is a powerful research tool for obtaining T-patterns. This software makes it possible to explore behavioral structures in detail by revealing stronger connections between successive recorded behaviors in goals than would be expected by chance. The critical interval is the key concept that

| FG,MOR,CL | OR,THR,CL | RF,OR,CL |

| (a) | (b) | (c) |
| Ronaldo is facing goal respect to the rival goal line (FG) doing more than five steps between touch and touch the ball (MOR), in the left midfield (CL) | He changes to facing right respect to the rival goal line (OR) doing three steps between touch and touch the ball (THR), while maintaining in the left midfield (CL) | He uses his right foot (RF), maintaining facing right respect to the rival goal line (OR), while maintaining in the left midfield (CL). |

FIGURE 5 | T-pattern shows that (a) Ronaldo is facing the rival goal line (FG), takes more than five steps between touches of the ball (MOR) in the left midfield (CL); then, (b) he changes to facing right with respect to the rival goal line (OR) and takes three steps between touches of the ball (THR) while remaining in the left midfield (CL); and (c) he uses his right foot (RF), remaining facing right with respect to the rival goal line (OR), while remaining in the left midfield (CL).

makes it possible to delimit the admissible temporal distances between successive identical or similar occurrences in order to consider the existence of a temporal pattern. Obtaining T-patterns is a procedure of great importance for theoretical and empirical purposes, and deriving their algorithm has involved the development of powerful new analytic techniques based on probability theory and, more specifically, on binomial distribution (Magnusson, 2000). Three criteria were applied to guarantee that any T-patterns detected were not due to random events: (a) presence of a given T-pattern in at least 25% of all sequences, (b) significance level of 0.005, and (c) redundancy reduction setting of 90% for occurrences of similar T-patterns. As Magnusson states, the idea of T-pattern analysis is to detect repeated behavioral patterns that are invisible to unaided observers. The temporal structure of complex behavioral sequences is composed of simpler or directly distinguishable event-types (Magnusson et al., 2016). Each T-data set subject to analysis consists of series of behaviors coded as occurrence times (beginning and end points) within specified observation periods (time point series; Magnusson, 1996).

More specifically, the following explanation, used in several studies, allows a clear understanding of how T-pattern detection works. For instance, in a given observation period, two repeated actions, A and B, either in this same order or simultaneously, form a minimal T-pattern (AB) if they are found more often than would be expected by chance, and if, assuming the null hypothesis of independent distributions for A and B, they are separated by approximately the same distance (time). Instances of A and B separated by this approximate distance constitute an (AB) T-pattern and their occurrence times are added to the original data. More complex T-patterns consisting of simple, already-detected patterns are subsequently added through a bottom-up detection procedure. Pairs or series of patterns can thus be detected, for example (((AB)C)(DE)) (see **Figure 1**).

The THEME software compares all patterns and retains only the most complete ones. Although, only a limited range of basic unit sizes is relevant in any study, T-patterns are, in principle, scale-independent as any basic time unit can be used. Thus, it would be fruitful in the study of Messi's and Ronaldo's goal-scoring.

Polar Coordinate Analysis

The structure of polar coordinate analysis, a technique of sequential analysis (Bakeman, 1978), is based on the complementarity between two analytical perspectives: prospective and retrospective. Polar coordinate analysis involves the detection of significant associations between *focal behavior* (the behavior of interest) and *conditional behaviors* (the other behaviors analyzed).

To define a focal behavior, it is first necessary to conduct the prospective analysis, which, depending on the aims of the study, is believed to generate or trigger a series of connections with other categories, known as conditional behaviors. The retrospective, or "backward" perspective, which incorporates what Anguera (1997) referred to as the concept of "genuine retrospectivity," reveals significant associations between the focal behavior and behaviors that occur before this behavior.

The technique of polar coordinate analysis can be applied to a series of values that are independent of each other, which is the case of adjusted residuals, whether prospective or retrospective, as they are calculated separately for each lag. Standardized Z statistics derived from adjusted residuals (Bakeman, 1978, 1991) corresponding to both prospective and retrospective lags are needed to compute prospective and retrospective Zsum statistics. These values, which can be positive or negative and are located in one of four quadrants, are then used to build maps showing the relationships between a focal behavior (Gorospe and Anguera, 2000; or a criterion behavior, as it is known in lag sequential analysis) and one or more conditional behaviors. Polar coordinate analysis involves the application of a complex procedure to provide a vector map of interrelated behaviors. The same number of prospective and retrospective lags is analyzed in each case. Prospective lags show which conditional behaviors precede the given behavior, while retrospective lags show which behaviors follow it.

As mentioned above, polar coordinate analysis merges the prospective and retrospective approaches to achieve a powerful reduction of data through the calculation of the Z_{sum} statistic $\left(\frac{\Sigma z}{\sqrt{n}} \right)$ described by Cochran (1954) and later developed by Sackett (1980). In both the prospective approach ($Z_{sum}P$) and the retrospective approach ($Z_{sum}R$), calculations are based on the frequency of the given behavior, n, and a series of mutually independent z-values for each lag. Each of these values is obtained by applying the binomial test to compute conditional probabilities (based on the number of codes recorded for each goal sequence) and unconditional probabilities (due to random effects). The length of each vector is obtained from $\sqrt{(Z_{sum}P)^2 + (Z_{sum}R)^2}$, while its angle is calculated by dividing the retrospective Z_{sum} arcsine by the radius (φ = arcsine of Y/radius). Prospective and retrospective Z_{sum} values (lags 1–5 and lags −1 to −5, respectively) can carry a positive or negative sign; these signs determine in which quadrant the resulting vectors (behaviors) are placed. To

BODY PART

MESSI   **RONALDO**

Focal category: RF (right foot)

The use of the right foot (RF) is mutually activated (QI) by following the use of the right foot (RF) and facing the chest toward the rival goal line (FG).

Focal category: RF (right foot)

The use of the right foot (RF) is activated (QII) by relative numerical superiority (NPR) and numerical equality with pressure (PE) and is mutually activated (QI) with the use of the external zone of the foot (OU) and three steps between touches of the ball (THR).
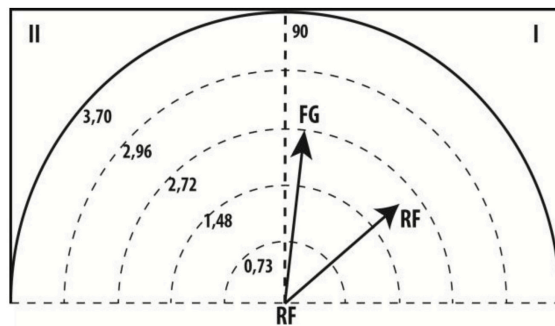
Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Bodypart_RF	I	1.41	1.41	1.99 (*)	45
Bodyorient_FG	I	0.13	2.72	2.72 (*)	87.17
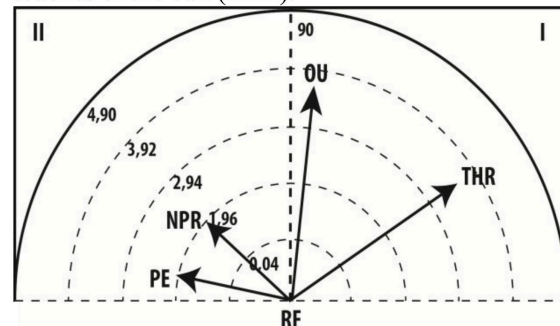
Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Centre_PE	II	-1.88	0.59	1.97 (*)	162.67
Centre_NPR	II	-1.22	1.54	1.96 (*)	128.43
Footczone_OU	I	0.46	3.65	3.68 (*)	82.87
Locomot_THR	I	2.94	1.98	3.54 (*)	33.98

Focal category: LF (left foot)

The use of the left foot (LF) is activated (QII) if the player makes a full or half turn to the left (vertical axis) (LT) and is mutually activated (QI) with the use of three steps between touches (THR), the left body orientation with respect to the rival goal line (OL) and numerical equality with no pressure (NPE).

Focal category: LF (left foot)

The use of the left foot (LF) is activated (QII) by the use of the inside of the foot (ID) and is mutually activated (QI) with the left body orientation with respect to the rival goal line (OL) and numerical equality with pressure (PE).
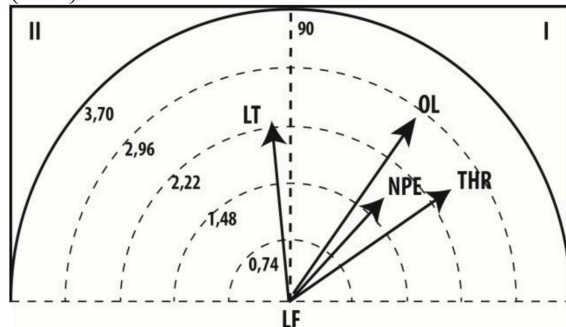
Category	Q	P.Prosp.	P.Retro.	Radius	Angle
St.Turn dir_LT	II	-0.29	2.31	2.33 (*)	97.22
Bodyorient_OL	I	1.68	2.27	2.83 (*)	53.59
Locomot_THR	I	2.16	1.27	2.50 (*)	30.41
Centre_NPE	I	1.29	1.55	2.02 (*)	50.08
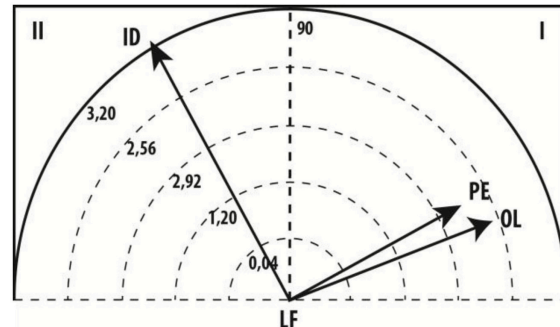
Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Footczone_ID	II	-2.01	2.70	3.36 (*)	126.71
Bodyorient_OL	I	2.35	0.95	2.54 (*)	21.96
Centre_PE	I	2.07	1.22	2.41 (*)	30.56

FIGURE 6 | Maps of polar coordinate analysis for the Body Part criterion.

FOOT CONTACT ZONE

MESSI

Focal category: OU (outside of the foot)
The use of the outside of the foot (OU) is activated (QII) by absolute numerical inferiority (PA) and is mutually activated (QI) by the use of three steps between touches of the ball (THR).

RONALDO

Focal category: OU (outside of the foot)
The use of the outside of the foot (OU) is activated (QII) by the use of the head (HD), facing backwards with respect to the rival goal line (BT) and pivoting over the right leg (RFP), and is mutually activated (QI) with the use of the external zone of the foot (OU) and the right foot (RF) and the right orientation of the body with respect to the rival goal line (OR).

Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Centre_PA	II	-1.19	2.10	2.42 (*)	119.60
Locomot_THR	I	2.13	0.72	2.25 (*)	18.74

Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Bodypart_HD	II	-3.16	0.12	3.16 (*)	177.80
Bodyorient_BT	II	-1.96	0.30	1.98 (*)	171.34
St.Pivotfoot RFP	II	-2.72	0.25	2.73 (*)	174.72
Footczone_OU	I	2.50	2.50	3.54 (*)	45
Bodypart_RF	I	3.65	0.46	3.68 (*)	7.13
Bodyorient_OR	I	2.38	0.16	2.39 (*)	3.80

FIGURE 7 | Maps of polar coordinate analysis for the Foot Contact Zone criterion.

illustrate the results, a map with four quadrants indicates the relationship (inhibitory vs. excitatory) between the focal and conditional behaviors. Thus, each quadrant reveals the following relationships:

Quadrant I (++). The given and conditional behaviors are mutually excitatory.
Quadrant II (− +). The given behavior is inhibitory and the conditional behavior is excitatory.
Quadrant III (− −). The given and conditional behaviors are mutually inhibitory.
Quadrant IV (+ −). The given behavior is excitatory and the conditional behavior is inhibitory.

As in previous research (Castañer et al., 2016a), **Figure 2** gives a graphical explanation of how to interpret the associations between given and conditional behaviors depending on the quadrant.

In each polar coordinate map, the focal behavior is placed in the middle and, depending on the quadrant in which the conditional behavior is placed, the angle of the vector is transformed as follows: quadrant I $(0 < \varphi < 90) = \varphi$; quadrant II $(90 < \varphi < 180) = 180 - \varphi$; quadrant III $(180 < \varphi < 270) = 180 + \varphi$; quadrant IV $(270° < \varphi < 360°) = 360° - \varphi$.

The HOISAN v1.6.3.2 software was used to calculate the prospective and retrospective adjusted residuals and the length and angle of the vectors and to produce a graphical representation of the results obtained.

RESULTS

T-Pattern Detection

T-pattern detection was performed using the free THEME software. Firstly, we explored the frequency of events and event sequences (**Figure 3**). The box in **Figure 3** shows the first 25 event-types with more than 2 occurrences (Messi in the left chart and Ronaldo in the right chart).

The most frequent event-types for both players were a total of nine configurations of codes. These were: facing goal, three steps between touches in the left midfield (FG,THR,CL) (Messi, $n = 16$; Ronaldo, $n = 15$); facing goal, more than five steps in the left midfield (FG,MOR,CL) (Messi, $n = 12$; Ronaldo, $n = 14$); facing goal, more than five steps in the right midfield (FG,MOR,CR) (Messi, $n = 12$; Ronaldo, $n = 6$); left orientation of the body with respect to the rival goal line, three steps in the right midfield (OL,THR,CR) (Messi, $n = 12$; Ronaldo, $n =$

BODY ORIENTATION

MESSI

Focal category: FG (chest facing rival goal line)
The orientation of chest facing the rival goal line (FG) is activated (QII) by numerical equality without pressure (NPE) and pivoting over the left leg (LFP) and is mutually activated (QI) by continuing to face the goal line (FG) and the use of the right leg (RL) and right foot (RF).

RONALDO

Focal category: FG (chest facing rival goal line)
The orientation of chest facing the rival goal line (FG) is activated (QII) by relative numerical superiority (NPR) and turning to the right (RT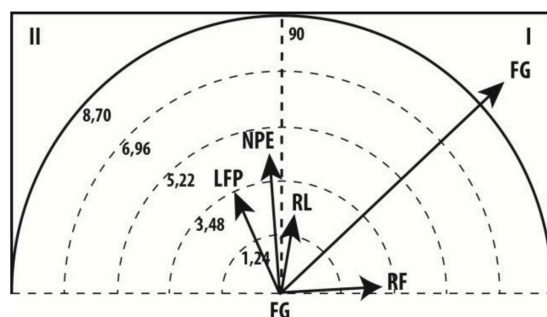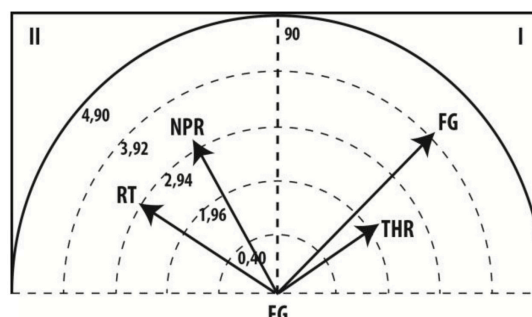) and is mutually activated (QI) by continuing to face the goal line (FG) and the use of three steps between touches (THR).

Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Centre_NPE	II	-0.46	4.36	4.39 (*)	95.97
St.Pivfoot_LFP	II	-1.87	2.73	3.31 (*)	124.47
Bodypart_RF	I	2.72	0.13	2.72 (*)	2.83
Bodypart_RL	I	0.18	1.96	1.97 (*)	84.74
Bodyorient_FG	I	6.30	6.30	8.91 (*)	45.00

Category	Q	P.Prosp.	P.Retro.	Radius	Angle
St.Turndir_RT	II	-2.86	0.82	2.98 (*)	163.92
Centre_NPR	II	-1.80	2.60	3.17 (*)	124.73
Locomot_THR	I	1.50	1.61	2.20 (*)	47.11
Bodyorient_FG	I	2.82	2.82	3.98 (*)	45

FIGURE 8 | Maps of polar coordinate analysis for the Body Orientation criterion.

11); facing goal, four steps between touches in the left midfield (FG,FOU,CL) (Messi, $n = 8$; Ronaldo, $n = 10$); left orientation of the body in the right midfield (OL,CR) (Messi, $n = 8$; Ronaldo, $n = 8$); facing goal in the right midfield (FG,CR) (Messi, $n = 7$; Ronaldo $n = 10$); facing goal, five steps in the right midfield (FG,FIV,CR) (Messi, $n = 6$, Ronaldo, $n = 6$); and facing goal, three steps in the right midfield (FG,THR,CR) (Messi, $n = 7$, Ronaldo, $n = 10$).

Other detectable aspects shown on the frequency chart are the fact that Messi used his left foot (LF) in 8 configurations and his right foot (RF) in 1 configuration. Ronaldo used his right foot (RF) in 8 configurations and his left foot (LF) did not appear in any configuration of codes. Messi used the left body orientation (OL) with respect to the rival goal line in 7 configurations of codes and the right body orientation (OR) did not appear in any configuration. Ronaldo used the right body orientation (OR) with respect to the rival goal in 7 configurations and the left body orientation (OL) appeared in 2 configurations.

Obtaining T-patterns allows us to show a broad view of the main sequences that the two players use in the process of goal-scoring. As any basic time unit can be used, the T-pattern technique selects the range of basic unit sizes that are relevant in any study. For this study, the categories that appeared in the T-patterns were: Body Part, Foot Contact Zone, Body Orientation,

Action and Side. **Figures 4**, **5** show the most complete T-patterns detected for Messi and Ronaldo, respectively.

Polar Coordinate Analysis

Given the clear understanding of the associations between focal and conditional behaviors provided by **Figure 2**, we selected quadrant II (QII), which contains the conditional categories that activate the focal category, and quadrant I (QI), which contains the categories that have mutual activation with the focal category. The maps in **Figures 6–13** show both quadrants with the length and angle of the vectors with a length of >1.96 ($p < 0.05$) for the behaviors that show statistically significant associations (activation).

Figures 6–13 show the results of polar coordinate analysis for Messi and Ronaldo concerning the categories in quadrant II (QII), which activate the focal category, and those in quadrant I (QI), which are mutually activated by the focal category. We include below each semicircle map the table of values statistically obtained. Firstly, we expose the categories that appear in the T-patterns corresponding to the following criteria: Body Part, Foot Contact Zone, Body Orientation, Locomotion and Side. Complementarily, we offer the polar coordinate analysis for the criteria Stability (turn direction) and Centre of the Game, which have also shown statistically significant activation between them.

FIGURE 9 | Maps of polar coordinate analysis for the Locomotion criterion.

DISCUSSION

The objective of this study was to perform an objective analysis of Lionel Messi's and Cristiano Ronaldo's use of motor skills prior to scoring a goal using the complementary methods of T-pattern analysis and polar coordinate analysis.

The structure of the Discussion Section is as follows. First, we comment on the polar coordinate analysis results following the order of criteria in the OSMOS-soccer player instrument. Second, we comment on the findings of the T-patterns analysis. Each section ends with clues about how experts can understand the findings in order to improve their professional work.

Body Contact with the Ball

Polar coordinate maps show great differences between the two players with regard to the use of the right foot. While there are no behaviors by Messi that activate the use of the right foot, Ronaldo's use of the right foot is promoted in situations of relative numerical superiority and numerical equality with pressure and is mutually activated by the use of the external zone of the foot and taking three steps between touches of the ball. In contrast, maps for the use of the left foot show more mutual activations between behaviors for Messi and fewer for Ronaldo. Moreover, Messi's use of the left foot and the left body orientation with respect to the rival goal line is induced by turning the body to the left. The use of three steps between touches and numerical equality with no pressure seem to be behaviors mutually activated

with the use of the left foot. These maps reinforce the notion that Ronaldo and Messi tend to use a preferred foot—right and left, respectively—in situations without high pressure and while dribbling to create advantage in attacking zones and in one-on-one situations. Moreover, the findings of Castañer et al. (2016a) related to the contralateral dominance of Messi's body orientation are corroborated. Our results also verify the findings of Carey et al. (2001), which showed that players mostly used the preferred foot when performing set pieces and the technical actions of first touching, passing, dribbling, and tackling. Furthermore, Carey et al. (2001) highlighted that players were more asymmetrical for set pieces than for the dynamic phases of the game.

The use of the inside of the foot activates Ronaldo's left foot use and this is mutually activated with numerical equality with pressure (PE) and, like Messi, the left body orientation with respect to the rival goal line.

Likewise, T-pattern analysis clearly shows the predominant use of the left foot by Messi and the right foot by Ronaldo. Despite the great differences between the two players in terms of the use of the right and left foot, the polar coordinate maps and frequency chart also show the players' versatility and adaptability in using both feet with other behaviors when necessary. Carey et al. (2001) found that very few players used both feet with equal frequency, but on those rare occasions they showed similar performance with the preferred and non-preferred feet. We therefore advise experts that the successful use of both feet, notwithstanding with different frequency, thus evidencing versatility, is an indicator of

FIGURE 10 | Maps of polar coordinate analysis for the Side criterion.
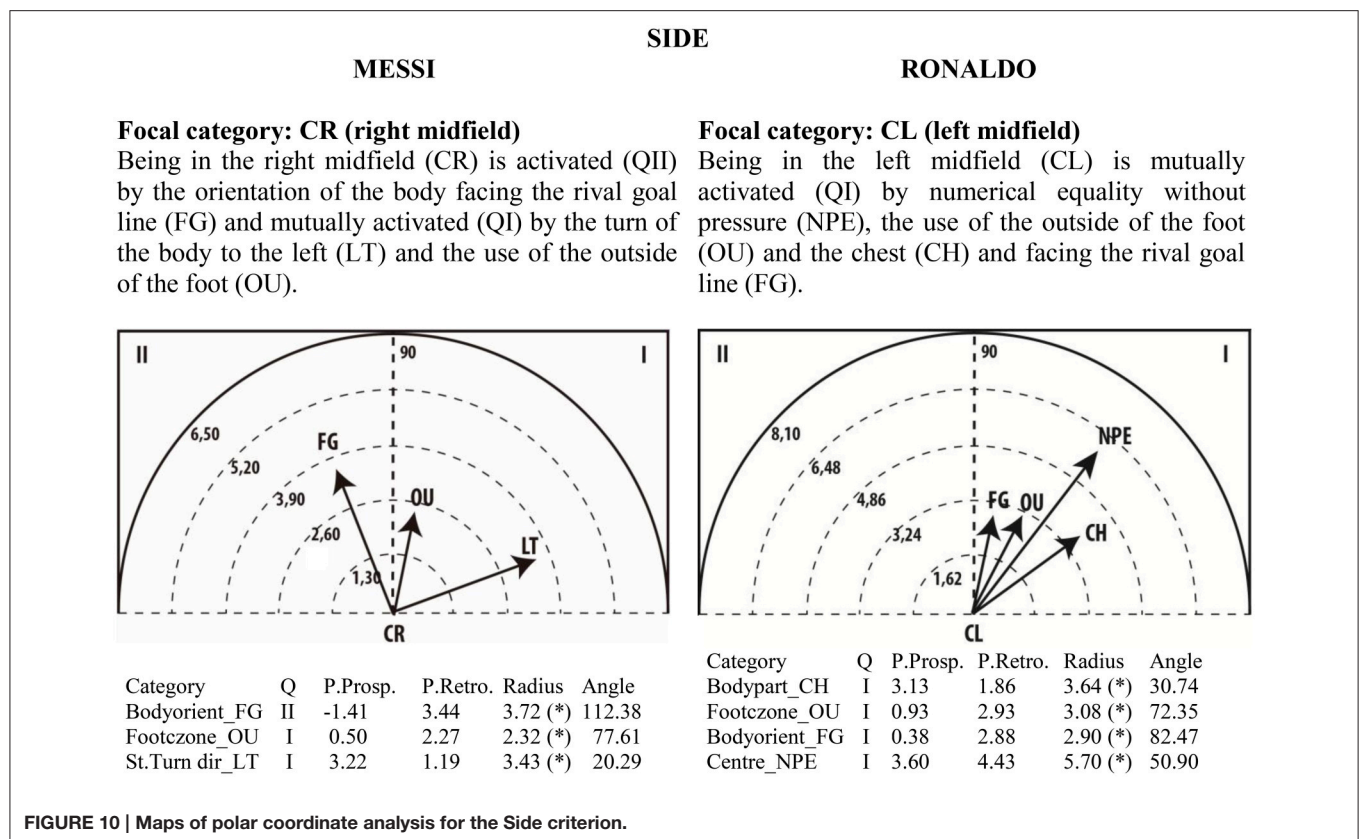
expertise in soccer and as such could be included as a coaching task in order to develop symmetrical use of both feet during dynamic interaction with the ball.

Foot Contact Zone

For Messi, the use of the outside of the foot is activated by absolute numerical inferiority and activates the use of three steps between touches of the ball. The use of the outside of the foot by Ronaldo is activated by the use of the head, facing backward with respect to the rival goal line and pivoting over the right leg, and is mutually activated with the use of the external zone of the foot and the right foot, as well as the right orientation of the body with respect to the rival goal line (OR). This finding fits with the logic of soccer play: players usually use the exterior part of the foot to run with the ball faster.

Body Orientation with Respect to the Goal Line

As the main task of strikers is goal-scoring, it is not surprising that both players use the body orientation of facing the rival goal line in interaction with other behaviors. We emphasize that contexts with no pressure induce both players to face the goal—the context of numerical equality without pressure in Messi's case and relative numerical superiority in Ronaldo's case. This result shows that expert players have great anticipation capacities, corroborating the finding of Ericsson (2003) that experts seem to be better at catching early relevant indicators of the specific

task. In our study, Messi and Ronaldo seem to create positional advantages in relation to the rival goal by using their attention abilities to better anticipate the outcomes of their actions and the actions of opponents (Afonso et al., 2012). So, in direct relation with the ball, they have already prepared conditions to have higher success in attacking situations.

Messi's goal-facing orientation is mutually activated mainly with remaining facing the goal line, with the use of the right leg and with the use of the right foot. As Messi is left-footed, the use of the right foot and leg while facing the rival goal line does not seem to us to be a paradox but rather an indication of his versatility in the use of contralateral inferior limbs, as the values of the polar coordinate analysis are very low. These findings are consistent with the findings of previous research (Castañer et al., 2016a).

Locomotion

Both polar coordinate analysis and T-pattern detection detected the locomotion behavior of taking three steps between touches of the ball. In both players, this behavior is activated by relative numerical superiority. Also, we found that Messi and Ronaldo use the outside part of the foot, a category that was activated in Ronaldo's map and mutually activated in Messi's map. These results could be interpreted to mean that in no-pressure conditions of play the exterior part of the foot is the part used most often in dribbling because with this ability the players create more speed conditions
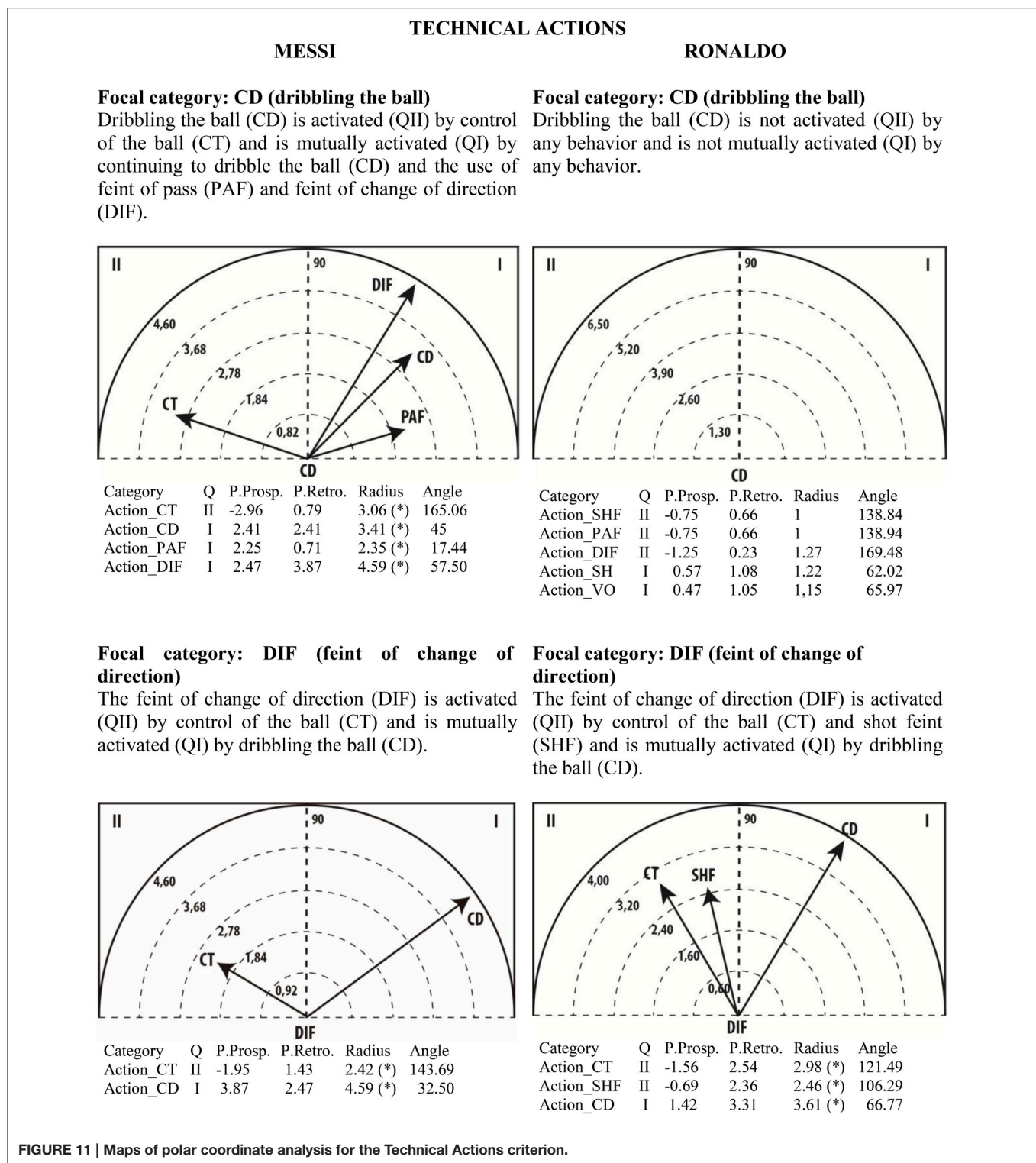
FIGURE 11 | Maps of polar coordinate analysis for the Technical Actions criterion.

The figure contains the following text and tables:

TECHNICAL ACTIONS

MESSI / **RONALDO**

Focal category: CD (dribbling the ball)

Dribbling the ball (CD) is activated (QII) by control of the ball (CT) and is mutually activated (QI) by continuing to dribble the ball (CD) and the use of feint of pass (PAF) and feint of change of direction (DIF).

Focal category: CD (dribbling the ball)

Dribbling the ball (CD) is not activated (QII) by any behavior and is not mutually activated (QI) by any behavior.

Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Action_CT	II	-2.96	0.79	3.06 (*)	165.06
Action_CD	I	2.41	2.41	3.41 (*)	45
Action_PAF	I	2.25	0.71	2.35 (*)	17.44
Action_DIF	I	2.47	3.87	4.59 (*)	57.50

Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Action_SHF	II	-0.75	0.66	1	138.84
Action_PAF	II	-0.75	0.66	1	138.94
Action_DIF	II	-1.25	0.23	1.27	169.48
Action_SH	I	0.57	1.08	1.22	62.02
Action_VO	I	0.47	1.05	1,15	65.97

Focal category: DIF (feint of change of direction)

The feint of change of direction (DIF) is activated (QII) by control of the ball (CT) and is mutually activated (QI) by dribbling the ball (CD).

Focal category: DIF (feint of change of direction)

The feint of change of direction (DIF) is activated (QII) by control of the ball (CT) and shot feint (SHF) and is mutually activated (QI) by dribbling the ball (CD).

Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Action_CT	II	-1.95	1.43	2.42 (*)	143.69
Action_CD	I	3.87	2.47	4.59 (*)	32.50

Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Action_CT	II	-1.56	2.54	2.98 (*)	121.49
Action_SHF	II	-0.69	2.36	2.46 (*)	106.29
Action_CD	I	1.42	3.31	3.61 (*)	66.77

in order to gain an advantage in space in relation to their opponents. The singular contralateral use of the feet of both players is again reinforced by these maps, which show the mutual activation of taking three steps and the use of the left foot in Messi's case and the right foot in Ronaldo's case.

Side

The right and left midfield are the categories of the Side criterion identified by polar coordinate analysis and T-pattern detection. T-patterns show clearly the difference between the two players in relation to the main uses of the midfield (the right midfield by Messi and the left midfield by Ronaldo). The presence of

STABILITY (TURN DIRECTION)

MESSI

Focal category: RT (right turn)
The right turn of the body (RT) is activated (QII) by numerical equality with pressure (PE), the use of the left leg (LL) and both, left orientation (OL) and backwards orientation with respect to the rival goal line (BT) and is mutually activated (QI) by the right orientation of the body with respect to the rival goal line (OR) and relative numerical superiority (NPR).

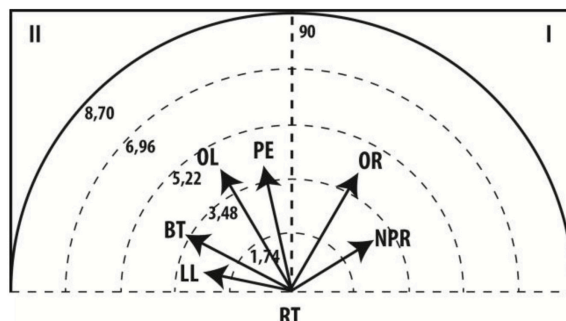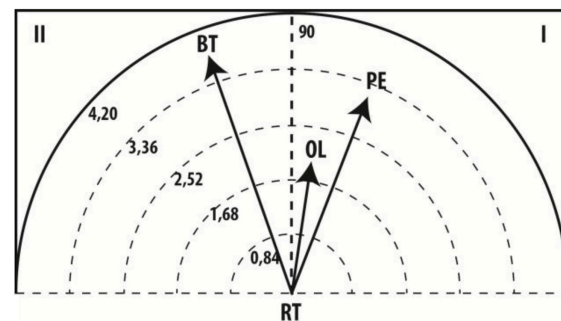| Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Bodypart_LL	II	-0.94	2.58	2.75 (*)	109.96
Bodyorient_BT	II	-0.65	3.51	3.57 (*)	100.43
Bodyorient_OL	II	-0.05	3.87	3.87 (*)	90.78
Centre_PE	II	-1.14	3.68	3.86 (*)	107.16
Bodyorient_BT	II	-0.65	3.51	3.57 (*)	100.43
Bodyorient_OL	II	-0.05	3.87	3.87 (*)	90.78
Centre_PE	II	-1.14	3.68	3.86 (*)	107.16
Bodyorient_OR	I	1.61	3.34	3.71 (*)	64.22
Centre_NPR	I	1.27	2.34	2.66 (*)	61.46

RONALDO

Focal category: RT (right turn)
The right turn of the body (RT) is activated (QII) by backwards orientation with respect to the rival goal line (BT) and is mutually activated (QI) by the left orientation of the body with respect to the rival goal line (OL) and numerical equality with pressure (PE).

Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Bodyorient_BT	II	-1.49	3.52	3.83 (*)	112.91
Bodyorient_OL	I	-0.10	2.03	2.04 (*)	87.27
Centre_PE	I	0.44	3.09	3.12 (*)	81.93

FIGURE 12 | Maps of polar coordinate analysis for the Stability (turn direction) criterion.

Messi in the right midfield is activated by the body orientation facing the rival goal line and is mutually activated by turning the body to the left and the use of the outside of the foot. The presence of Ronaldo in the left midfield is activated by numerical equality without pressure, the use of the outside of the foot, the use of the chest and facing the rival goal line. These results corroborate statistics presented by InStat Scout software about Messi's and Ronaldo's patterns of play with regard to where the players touch the ball throughout the matches: 84% of Messi's touches occur in the right wing, 8% in the mid-offensive zone, and 8% in the central attacking zones; Ronaldo touches the ball mostly in the left wing (57%), followed closely by the central attacking zones (42%). These data show that Ronaldo tends to play in interior zones of the field more frequently than Messi.

Technical Actions

T-pattern detection and the frequency chart show more use of dribbling and feint of change of direction in Messi's goal-scoring than in Ronaldo's. Polar coordinate maps also show non-statistically significant activation between dribbling and other behaviors. Contrarily, Messi's dribbling is activated by control of

the ball and is mutually activated with continuing to dribble the ball, the use of feint of pass and the use of feint of change of direction. The T-pattern detection also reinforces this behavior (**Figure 4**): Messi touches the ball with the outside part of his left foot while facing the rival goal line. To do this, Messi tricks defenders by changing direction in the right midfield and then continuing to touch the ball with his left foot with a left orientation of the body with respect to the rival goal line; then, he continues dribbling the ball while remaining in the right midfield. We therefore conclude that Messi tends to create a great diversity of individual attacking situations, a result that corroborates the conclusion of Serrado (2015): that Messi is the world's most unpredictable player. Morris (2014), studying Messi between 2010 and 2014, reported that he has 50% efficacy in dribbling and tries to perform feints on average 8 times per game. He also showed that Messi was the most successful player in assists and goals scored, having the best goals/assists ratio with 1.30 goals and 0.40 assists per game. In the same period of analysis, in passing situations Messi was the striker with the most passes performed (11,120), 84% of them successfully. Of these passes, 47% were completed to attacking zones, with 450 through balls, 30 of them permitting a goal (Morris, 2014).

CENTRE OF THE GAME

MESSI

Focal category: NPE (Numerical equality with no pressure)

Numerical equality with no pressure (NPE) is activated (QII) by the right turn of the body (RT) and is mutually activated (QI) by remaining in numerical equality with no pressure (NPE), numerical equality with pressure (PE), backwards orientation with respect to the rival goal line (BT), the use of the right leg (RL) and left foot (LF) and left pivot foot (LFP).
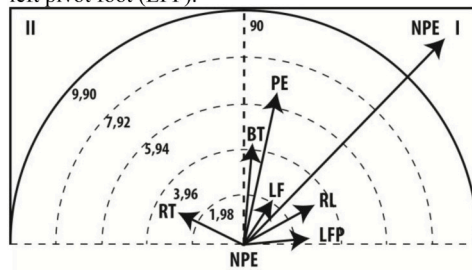
Category	Q	P.Prosp.	P.Retro.	Radius	Angle
St.Turndir_RT	II	-2.73	1.24	2.99 (*)	155.62
Bodypart_LF	I	1.55	1.29	2.02 (*)	39.92
Bodypart_RL	I	2.73	1.68	3.21 (*)	31.57
Bodyorient_BT	I	0.17	4.13	4.14 (*)	87.71
St.Pivotfoot_LFP	I	2.97	0.09	2.97 (*)	1.68
Centre_PE	I	1.05	6.42	6.51 (*)	80.71
Centre_NPE	I	8.24	8.24	11.66 (*)	45

Focal category: NPR (Relative numerical superiority)

Relative numerical superiority (NPR) is activated (QII) by the left turn of the body (LT) and is mutually activated (QI) by the relative numerical superiority (NPR), absolute numerical inferiority (PA), right orientation with respect to the rival goal line (OR), pivoting over the right foot (RFP), use of the left leg (LL), right turn of the body (RT) and use of the chest (CH).

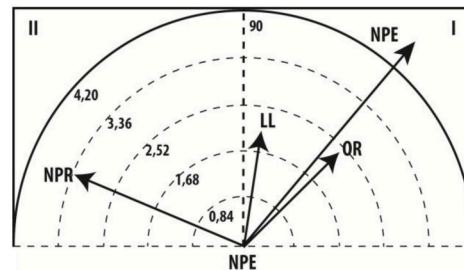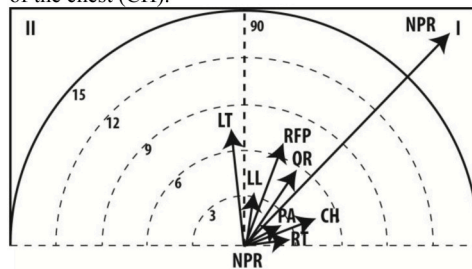| Category	Q	P.Prosp.	P.Retro.	Radius	Angle
St.Turndir_LT	II	-0.27	7.08	7.09 (*)	92.21
Bodypart_LL	I	0.55	3.28	3.33 (*)	80.48
Bodypart_CH	I	4.26	0.39	4.28 (*)	5.20
Bodyorient_OR	I	4.10	4.29	5.94 (*)	46.29
St.Turn_dir_RT	I	2.34	1.27	2.66 (*)	28.54
St.Pivofoot_RFP	I	2.72	6.16	6.74 (*)	66.18
Centre_PA	I	1.42	1.77	2.27 (*)	51.2
Centre_NPR	I	12.54	12.54	17.74 (*)	45

RONALDO

Focal category: NPE (Numerical equality with no pressure)

Numerical equality with no pressure (NPE) is activated (QII) by relative numerical superiority (NPR) and is mutually activated (QI) by remaining in numerical equality with no pressure (NPE), right orientation with respect to the rival goal line (OR) and the use of the left leg (LL).

Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Centre_NPR	II	-3.03	1.30	3.30 (*)	156.75
Bodypart_LL	I	0.24	2.12	2.13 (*)	83.53
Bodyorient_OR	I	1.83	1.62	2.44 (*)	41.46
Centre_NPE	I	3.51	3.51	4.96 (*)	..45

Focal category: NPR (Relative numerical superiority)

Relative numerical superiority (NPR) is activated (QII) by numerical equality with pressure (PE), left orientation with respect to the rival goal line (OL) and backwards orientation (BT), use of the head (HD) and pivoting over the right foot (RFP) and is mutually activated (QI) by the relative numerical superiority (NPR).
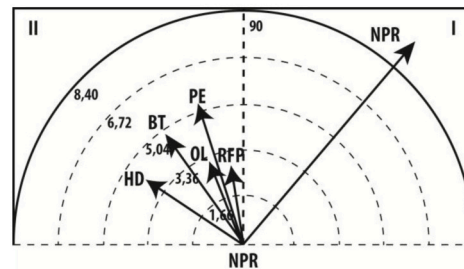
Category	Q	P.Prosp.	P.Retro.	Radius	Angle
Bodypart_HD	II	-1.97	3.95	4.42 (*)	116.51
Bodyorient_BT	II	-0.67	4.84	4.89 (*)	97.91
Bodyorient_OL	II	-1.29	2.76	3.05 (*)	115.1
St.Pivotfoot_RFP	II	-1.85	2.08	2.79 (*)	131.6
Centre_PE	II	-1.58	4.91	5.15 (*)	107.8
Centre_NPR	I	7.02	7.02	9.93 (*)	45

FIGURE 13 | Maps of polar coordinate analysis for the Centre of the Game criterion.

The maps also show that feint of change of direction is more similar in the two players. This behavior is activated by the control of the ball and is mutually activated with dribbling. For Ronaldo, it is also mutually activated by the shot feint, which corroborates the notion that Ronaldo was the top shooter in the 2010–2014 period, with 1,018 shots performed (Morris, 2014).

Stability (Turn Direction)

The maps show for Messi that the right turn of the body is activated by the use of the left leg and being oriented backwards with respect to the rival goal line and is mutually activated with the right body orientation with respect to the rival goal line and relative numerical superiority. This finding reinforces again the contralateral actions of stasis and precision of the laterality uses of the limbs (Teixeira et al., 2011). For Messi and also for Ronaldo, the map shows that the right turn of the body is activated by facing backwards with respect to the rival goal line and is mutually activated with the left body orientation with respect to the rival goal line and numerical equality with pressure for Ronaldo and right body orientation for Messi. Along similar lines, Castañer et al. (2016a) reported that the right turn of the body showed that Messi's goal-scoring was directly related to the use of the left leg because he remains steady over his right leg in order to turn the body, allowing the left leg to perform precise actions.

Centre of the Game

The most relevant aspect that can be seen in the behavior of numerical equality with no pressure is that in both players it is mutually activated by the continuation of numerical equality with no pressure. We conclude that expert players frequently create conditions, in time and space, to play in no-pressure conditions, in this case in goal-scoring situations. Anticipation is generally considered a hallmark of experts, so it should be considered on the basis of the specific tasks and contexts with knowledge of their advantages and disadvantages (Gold and Shadlen, 2007). Messi and Ronaldo, as the most expert goal scorers, seem to create better conditions to apply shooting technique.

Conclusions and Future Lines of Study

The objective of this study was to describe objectively the singular goal-scoring style of the world's top soccer players, Cristiano Ronaldo and Lionel Messi. Observational methodology allows sports scientists to obtain objective data to complement subjective judgments of soccer players' motor skill use. We used the OSMOS-soccer player observational system (Castañer et al., 2016a), applying six criteria related to the players' motor skills and three criteria related to tactics and contextual aspects. This instrument is a good fit for our study because we consider that going deeply into the motor skills that players use could be of interest to soccer studies, which are traditionally more focused on the tactical and technical analysis of teams. The combination of two powerful observational techniques, namely T-pattern detection and polar coordinate analysis, allowed us to describe the "mosaic" of motor skills and contextual aspects that make up the singular style of play of Messi and Ronaldo, two of the best soccer players in the world in the early twenty-first century.

Our findings permit us to conclude that Messi and Ronaldo exhibit motor skills that allow them to create varied conditions for goal-scoring. The cumulative use of these abilities, over the course of matches and seasons, allows them to win the top awards in soccer. Here we detail our most important results:

- The creation of no-pressure conditions in goal-scoring shows that both players use attention abilities to better anticipate the outcomes of their motor actions and the actions of their opponents, resulting in higher success in attacking.
- The players exhibited symmetry in the use of both feet with success. However, in conditions of goal-scoring, i.e., with no pressure in the centre of the game, both players mostly used the foot with better laterality precision.
- Ronaldo and Messi mainly use the exterior part of the foot to dribble faster in order to create advantage in attacking zones and in one-on-one situations.
- The players exhibited great versatility in the use of a vast variety of motor skills and technical actions in goal-scoring contexts. Messi is considered an unpredictable player in his goal-scoring actions and Ronaldo an accurate shooter with more recurring patterns.

As for the practical implications of this study, in the Discussion Section we indicated the findings that could be of interest for coaches and for further related studies. Overall, coaches may use these findings for task manipulation related to skill acquisition and improvement of goal-scoring efficacy. Also, studies of this type could be useful for establishing defensive strategies against these specific players. Thus, it would be interesting for future research to consider others types of contexts or outcomes, for example World Cup competition and shots off target, respectively, to better discriminate between the motor ability patterns of successful and unsuccessful performances.

AUTHOR CONTRIBUTIONS

MC developed the project and supervised the design of the study and the drafting of the manuscript. DB was responsible for the review of the literature and the drafting of the manuscript. OC was responsible for the T-pattern detection, data collection/handling and the critical revision of the content. MA performed the polar coordinate analysis and the method section. TF collected and codified the data. RH supervised the drafting of the manuscript. All authors approved the final, submitted version of the manuscript.

FUNDING

REFERENCES

Afonso, J., Garganta, J., and Mesquita, I. (2012). Decision-making in sports: the role of attention, anticipation and memory. *Rev. Bras. Cineantropom. Desempenho. Hum.* 14, 592–601. doi: 10.5007/1980-0037.2012v14n5p592

Ali, A. (2011). Measuring soccer skill performance: a review. *Scand. J. Med. Sci. Sports* 21, 170–183. doi: 10.1111/j.1600-0838.2010.01256.x

Anguera, M. T. (1979). Observational typology. *Qual. Quant.* 13, 449–484.

Anguera, M. T. (2003). "Observational methods," in *Encyclopedia of Psychological Assessment*, Vol. 2, ed R. Fernández-Ballesteros (London: Sage), 632–637.

Anguera, M. T. (1997). "*From Prospective Patterns in Behavior to Joint Analysis with a Retrospective Perspective,*" in *Colloque sur invitation "Méthodologie d'Analyse des Interactions Sociales"*. Paris: Université de la Sorbona.

Anguera, M. T., Camerino, O., Castañer, M., and Sánchez-Algarra, P. (2014). Mixed methods en la investigación de la actividad física y el deporte. *Rev. Psicol. Deporte* 23, 123–130.

Anguera, M. T., and Hernández-Mendo, A. (2014). Observational methodology and sports psychology: state of affairs. *Rev. Psicol. Deporte* 23, 103–109.

Anguera, M. T., and Hernández-Mendo, A. (2015). Data analysis techniques in observational studies in sport sciences. *Cuad. Psicol. Dep.* 15, 13–30. doi: 10.4321/S1578-84232015000100002

Anguera, M. T., and Jonsson, G. K. (2003). Detection of real-time patterns in sport interactions in football. *Int. J. Comput. Sci. Sport* 2, 118–121.

Aragón, S., Lapresa, D., Arana, J., Anguera, M. T., and Garzón, B. (2017). An example of the informative potential of polar coordinate analysis: sprint tactics in elite 1,500-m track events. *Meas. Phys. Educ. Exerc. Sci.* 16, 279–286. doi: 10.1080/1091367X.2016.1245192

Araújo, D., Fonseca, C., Davids, K., Garganta, J., Volossovitch, A., Brandão, R., et al. (2010). The role of ecological constraints on expertise development. *Talent Dev. Excell.* 2, 165–179.

Bakeman, R. (1978). "Untangling streams of behavior: sequential analysis of observation data," in *Observing Behavior: Data Collection and Analysis Methods*, Vol. 2, ed G. P. Sackett (Baltimore, MD: University Park Press), 63–78.

Bakeman, R. (1991). From lags to logs: advances in sequential analysis. *Mex. J. Behav. Anal.* 17, 65–83.

Barreira, D., Garganta, J., Castellano, J., Machado, J., and Anguera, M. T. (2015). How elite-level soccer dynamics has evolved over the last three decades? Input from generalizability theory. *Cuad. Psicol. Dep.* 15, 51–62. doi: 10.4321/S1578-84232015000100005

Barreira, D., Garganta, J., Guimarães, P., Machado, J., and Anguera, M. T. (2014). Ball recovery patterns as a performance indicator in elite soccer. *Proc. Inst. Mech. Eng. P J. Sports Eng. Technol.* 228, 61–72. doi: 10.1177/1754337113493083

Barreira, D., Garganta, J., Prudente, J., and Anguera, M. T. (2012). Desenvolvimento e validação de um sistema de observação aplicado à fase ofensiva em futebol: SoccerEye. *Rev. Port. Cien. Desp.* 12, 32–57.

Bishop, D. T., Wright, M. J., Jackson, R. C., and Abernethy, B. (2013). Neural bases for anticipation skill in soccer: an fMRI study. *J. Sport Exerc. Psychol.* 35, 98–109. doi: 10.1123/jsep.35.1.98

Blanco-Villaseñor, A., Losada, J. L., and Anguera, M. T. (2003). Data analysis techniques in observational designs applied to the environment-behaviour relation. *Medio Ambient. Comport. Hum.* 4, 111–126.

Bradley, P. S., Carling, C., Gómez-Díaz, A., Hood, P., Barnes, C., Ade, J., et al. (2013). Match performance and physical capacity of players in the top three competitive standards of English professional soccer. *Hum. Mov. Sci.* 32, 808–821. doi: 10.1016/j.humov.2013.06.002

Camerino, O., Castañer, M., and Anguera, M. T. (2012a). *Mixed Methods Research in the Movement Sciences: Cases in Sport, Physical Education and Dance.* Abingdon: Routledge.

Camerino, O., Chaverri, J., Anguera, M. T., and Jonsson, G. K. (2012b). Dynamics of the game in soccer: detection of T-patterns. *Eur. J. Sport Sci.* 12, 216–224. doi: 10.1080/17461391.2011.566362

Carey, D. P., Smith, G., Smith, D. T., Shepherd, J. W., Skriver, J., Ord, L., et al. (2001). Footedness in world soccer: an analysis of France '98. *J. Sport. Sci.* 19, 855–864. doi: 10.1080/026404101753113804

Carling, C., Williams, M., and Reilly, T. (2005). *Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance.* Abingdon: Routledge.

Carling, C., Bloomfield, J., Nelsen, L., and Reilly, T. (2008). The role of motion analysis in elite soccer: contemporary performance measurement techniques and work rate data. *Sports Med.* 38, 839–862. doi: 10.2165/00007256-200838100-00004

Casarrubea, M., Jonsson, G. K., Faulisi, F., Sorbera, F., Di Giovanni, G., Benigno, A., et al. (2015). T-pattern analysis for the study of temporal structure of animal and human behavior: a comprehensive review. *J. Neurosci. Methods* 239, 34–46. doi: 10.1016/j.jneumeth.2014.09.024

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Canton, A., and Hileno, R. (2016a). Goal scoring in soccer: a polar coordinate analysis of motor skills used by Lionel Messi. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Castañer, M., Camerino, O., Landry, P., and Parés, N. (2016b). Quality of physical activity of children in exergames: sequential body movement analysis and its implications for interaction design. *Int. J. Hum. Comput. Stud.* 96, 67–78. doi: 10.1016/j.ijhcs.2016.07.007

Castañer, M., Torrents, C., Anguera, M. T., Dinušová, M., and Jonsson, G. K. (2009). Identifying and analyzing motor skill responses in body movement and dance. *Behav. Res. Methods* 41, 857–867. doi: 10.3758/BRM.41.3.857

Chi, M. T. H. (2006). "Two approaches to the study of experts' characteristics," in *The Cambridge Handbook of Expertise and Expert Performance*, eds K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman (Cambridge: Cambridge University Press), 21–30.

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* 10, 417–451. doi: 10.2307/3001616

Davids, K., Button, C., Araújo, D., Renshaw, I., and Hristovski, R. (2006). Movement models from sports provide representative task constraints for studying adaptive behavior in human movement systems. *Adapt. Behav.* 14, 73–95. doi: 10.1177/105971230601400103

Duch, J., Waitzman, J. S., and Amaral, L. A. N. (2010). Quantifying the performance of individual players in a team activity. *PLoS ONE* 5:10937. doi: 10.1371/journal.pone.0010937

Echeazarra, I., Castellano, J., Usabiaga, O., and Hernández-Mendo, A. (2015). Differences in the strategic use of space in under 14 and under 16 soccer: a polar coordinate analysis. *Cuad. Psicol. Dep.* 15, 169–180. doi: 10.4321/S1578-84232015000100017

Ericsson, K. A. (2003). "Development of elite performance and deliberate practice: an update from the perspective of the expert performance approach," in *Expert Performance in Sports Advances in Research on Sport Expertise*, eds J. Starkes and K. A. Ericsson (Champaign, IL: Human Kinetics), 49–83.

Escolano-Pérez, E., Herrero-Nivela, M. L., and Echeverría-Expósito, R. (2014). Problem solving as a strategy in teaching offensive tactics at the under 8 level in 8-a-side-soccer. *Rev. Psicol. Dep.* 23, 209–217.

Feltovich, P. J., Prietula, M. J., and Ericsson, K. A. (2006). "Studies of expertise from psychological perspectives," in *The Cambridge Handbook of Expertise and Expert Performance*, eds K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman (Cambridge: Cambridge University Press), 41–68.

Fernández, J., Camerino, O., Anguera, M. T., and Jonsson, G. K. (2009). Identifying and analyzing the construction and effectiveness of offensive plays in basketball by using systematic observation. *Behav. Res. Methods* 41, 719–730. doi: 10.3758/BRM.41.3.719

Gabín, B., Camerino, O., Anguera, M. T., and Castañer, M. (2012). Lince: multiplatform sport analysis software. *Proc. Soc. Behav. Sci.* 46, 4692–4694. doi: 10.1016/j.sbspro.2012.06.320

Garzón Echevarría, B., Lapresa Ajamil, D., Anguera Argilaga, M. T., and Arana Idiakez, J. (2011). Observational analysis of the free throw shot made by grassroots basketball players. *Psicothema* 23, 851–857.

Glazier, P., and Davids, K. (2009). Constraints on the complete optimization of human motion. *Sports Med.* 39, 15–28. doi: 10.2165/00007256-200939010-00002

Glazier, P. S., and Robins, M. T. (2013). "Self-organisation and constraints in sports performance," in *Routledge Handbook of Sports Performance Analysis*, eds T. McGarry, P. O'Donoghue, and J. Sampaio (Abingdon: Routledge), 42–51.

Gold, J., and Shadlen, M. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574. doi: 10.1146/annurev.neuro.29.051605.113038

Gorospe, G., and Anguera, M. T. (2000). Retrospectivity in polar coordinates analysis: application to tennis. *Psicothema* 12, 279–282.

Hernández-Mendo, A., Castellano, J., Camerino, O., Jonsson, G., Blanco-Villaseñor, A., Lopes, A., et al. (2014). Observational software, data quality control and data analysis. *Rev. Psicol. Dep.* 23, 111–121.

Hernández-Mendo, A., López-López, J. A., Castellano, J., Morales-Sánchez, V., and Pastrana, J. L. (2012). IT program for use in methodology observacional. *Cuad. Psicol. Dep.* 12, 55–78.

Hodges, N. J., Starkes, J. L., and MacMahon, C. (2006). "Expert performance in sport: a cognitive perspective," in *The Cambridge Handbook of Expertise and Expert Performance*, eds K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman (Cambridge: Cambridge University Press), 471–488.

Hong, S., Kazama, Y., Nakayama, M., and Asai, T. (2012). Ball impact dynamics of knuckling shot in soccer. *Proc. Eng.* 34, 200–205. doi: 10.1016/j.proeng.2012.04.035

Jafari, S., and Smith, L. S. (2016). Can Lionel Messi's brain slow down time passing? *Chronobiol. Int.* 33, 462–463. doi: 10.3109/07420528.2016.1154570

Jonsson, G. K., Anguera, M. T., Blanco-Villaseñor, A., Losada, J. L., Hernández-Mendo, A., Ardá, T., et al. (2006). Hidden patterns of play interaction in soccer using SOF-CODER. *Behav. Res. Methods* 38, 372–381. doi: 10.3758/bf03192790

Lago-Ballesteros, J., Lago-Peñas, C., and Rey, E. (2012). The effect of playing tactics and situational variables on achieving score-box possessions in a professional soccer team. *J. Sport. Sci.* 30, 1455–1461. doi: 10.1080/02640414.2012.712715

Lapresa, D., Alsasua, R., Arana, J., Anguera, M. T., and Garzón, B. (2014). Observational analysis of the construction of offensive sequences that end in a shot in youth basketball. *Rev. de Psicol. Dep.* 23, 365–376.

Lapresa, D., Álvarez, L., Arana, J., Garzón, B., and Caballero, V. (2013). Observational analysis of the offensive sequences that ended in a shot by the winning team of the 2010 UEFA Futsal Championship. *J. Sport. Sci.* 31, 1731–1739. doi: 10.1080/02640414.2013.803584

Loffing, F., Solter, F., Hagemann, N., and Strauss, B. (2015). Accuracy of outcome anticipation, but not gaze behavior, differs against left- and right-handed penalties in team-handball goalkeeping. *Front. Psychol.* 6:1820. doi: 10.3389/fpsyg.2015.01820

López, J., Valero-Valenzuela, A., Anguera, M. T., and Díaz, A. (2016). Disruptive behavior among elementary students in physical education. *SpringerPlus* 5, 1154. doi: 10.1186/s40064-016-2764-6

López-López, J. A., Menescardi, C., Estevan, I., Falcó, C., and Hernández-Mendo, A. (2015). Technical-tactical analysis in taekwondo with polar coordinates through software HOISAN. *Cuad. Psicol. Dep.* 15, 131–142. doi: 10.4321/S1578-84232015000100013

Mackenzie, R., and Cushion, C. (2013). Performance analysis in football: a critical review and implications for future research. *J. Sport. Sci.* 31, 639–676. doi: 10.1080/02640414.2012.746720

Magnusson, M. S. (1996). Hidden real-time patterns in intra- and inter-individual behavior: description and detection. *Eur. J. Psychol. Assess.* 12, 112–123. doi: 10.1027/1015-5759.12.2.112

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behav. Res. Method Ins. Comput.* 32, 93–110. doi: 10.3758/BF03200792

Magnusson, M. S., Burgoon, J. K., and Casarrubea, M. (eds.) (2016). *Discovering Hidden Temporal Patterns in Behavior and Interaction: T-Pattern Detection and Analysis with THEME*[TM]. New York, NY: Springer-Verlag.

McGarry, T., Anderson, D. I., Wallace, S. A., Hughes, M. D., and Franks, I. M. (2002). Sport competition as a dynamical self-organizing system. *J. Sports Sci.* 20, 771–781. doi: 10.1080/026404102320675620

Memmert, D., Hüttermann, S., and Orliczek, J. (2013). Decide like Lionel Messi! The impact of regulatory focus on divergent thinking in sports. *J. Appl. Soc. Psychol.* 43, 2163–2167. doi: 10.1111/jasp.12159

Morillo-Baro, J. P., Reigal, R. E., and Hernández-Mendo, A. (2015). Analysis of positional attack in beach handball male and female with polar coordinates. *Rev. Int. Cienc. Deporte* 41, 226–244. doi: 10.5232/ricyde2015.04103

Morris, B. (2014). *Lionel Messi is impossible.* Available online at: https://fivethirtyeight.com/features/lionel-messi-is-impossible/

Murgia, M., Sors, F., Muroni, A. F., Santoro, I., Prpic, V., Galmonte, A., et al. (2014). Using perceptual home-training to improve anticipation skills of soccer goalkeepers. *Psychol. Sport Exerc.* 15, 642–648. doi: 10.1016/j.psychsport.2014.07.009

O'Donoghue, P. (2010). *Research Methods for Sports Performance Analysis.* Abingdon: Routledge.

Perea, A., Castellano, J., Alday, L., and Hernández-Mendo, A. (2012). Analysis of behaviour in sports through polar coordinate analysis with MATLAB®. *Qual. Quant.* 46, 1249–1260. doi: 10.1007/s11135-011-9435-z

Perić, A., Marković, G., Bradić, A., and Dizdar, D. (2013). The weighted expert system for the evaluation of actual quality of top-level soccer players. *Homo Sporticus* 15, 10–19.

Portell, M., Anguera, M. T., Chacón-Moscoso, S., and Sanduvete-Chaves, S. (2015). Guidelines for reporting evaluations based on observational methodology. *Psicothema* 27, 283–289. doi: 10.7334/psicothema2014.276

Robles, F. J., Castellano, J., and Perea, A. E. (2014). Differences of the game between the football spanish team and its rivals. *Rev. Ib. CC. Act. Fis. Dep.* 3, 1–8.

Sackett, G. P. (1980). "Lag sequential analysis as a data reduction technique in social interaction research," in *Exceptional Infant: Psychosocial Risks in Infant-Environment Transactions*, Vol. 4, eds D. B. Sawin, R. C. Hawkins, L. O. Walker, and J. H. Penticuff (New York, NY: Brunner-Mazel), 300–340.

Sánchez-Algarra, P., and Anguera, M. T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: impact of recording and coding predominating perspectives. *Qual. Quant.* 47, 1237–1257. doi: 10.1007/s11135-012-9764-6

Sandelowski, M. (1996). One is the liveliest number: the case orientation of qualitative research. *Res. Nurs. Health* 19, 525–529.

Sarmento, H., Barbosa, A., Anguera, M. T., Campaniço, J., and Leitão, J. (2013). "Regular patterns of play in the counterattack of the FC Barcelona and Manchester United football teams," in *Performance Analysis of Sport IX*, eds D. Peters, and P. O'Donogue (Abingdon: Routledge), 59–66.

Serrado, R. (2015). *Lionel Messi.* Edições Viera Da Silva.

Shergold, A. (2016). *Lionel Messi and Cristiano Ronaldo have Remarkably Both Scored 270 Goals in La Liga Since 2009.* Available online at: http://www.dailymail.co.uk/sport/football/article-4051266/Lionel-Messi-Cristiano-Ronaldo-remarkably-scored-270-goals-La-Liga-2009.html#ixzz4VHqHhUgs

Sousa, D. J., Prudente, J. N., Sequeira, P., López-López, J. A., and Hernández-Mendo, A. (2015). Analysis of 2vs2 situations in Men's European Handball Championship 2012: application of polar coordinates technique. *Cuad. Psic. Dep.* 15, 181–194. doi: 10.4321/S1578-84232015000100018

Tarragó, R., Iglesias, X., Lapresa, D., and Anguera, M. T. (2016). Complementarity between diachronic T-pattern relationships and behavioral patterns in elite male fencing. *Cuad. Psic. Dep.* 16, 113–128.

Teixeira, L. A., de Oliveira, D. L., Romano, R. G., and Correa, S. C. (2011). Leg preference and interlateral asymmetry of balance stability in soccer players. *Res. Q. Exerc. Sport* 82, 21–27. doi: 10.1080/02701367.2011.10599718

Tenga, A., Holme, I., Ronglan, L. T., and Bahr, R. (2010). Effect of playing tactics on achieving score-box possessions in a random series of team possessions from Norwegian professional soccer matches. *J. Sport. Sci.* 28, 245–255. doi: 10.1080/02640410903502766

Wallace, J. L., and Norton, K. I. (2014). Evolution of World Cup soccer final games 1966-2010: game structure, speed and play patterns. *J. Sci. Med. Sport* 17, 223–228. doi: 10.1016/j.jsams.2013.03.016

Yin, R. K. (2014). *Case Study Research: Design and Methods, 5th Edn.* Los Angeles, CA: Sage.

Zurloni, V., Cavalera, C., Diana, B., Elia, M., and Jonsson, G. K. (2014). Detecting regularities in soccer dynamics: a T-pattern approach. *Rev. Psicol. Dep.* 23, 157–164.

Check for
updates

Polar Coordinate Analysis of Relationships With Teammates, Areas of the Pitch, and Dynamic Play in Soccer: A Study of Xabi Alonso

Rubén Maneiro Dios* and Mario Amatria Jiménez

Faculty of Science of Education, Pontifical University of Salamanca, Salamanca, Spain

Research in soccer has traditionally focused on very specific aspects of the game, such as technical and physiological aspects, and has largely ignored important issues such as tactical performance and the role of individual players within the team. The aim of this study was to study the different relationships that Xabi Alonso, one of the world's best midfielders, establishes with his teammates during offensive play, and to investigate his connections with the pitch in terms of where his direct interventions started and finished, his use of technical actions, his involvement in set plays and interceptions, and his relationship with shots at goal. To do this, we analyzed all the matches played by the winner of the 2012 UEFA European Championship: Spain. We employed an observational methodology design (Anguera, 1979) using a modified version of the *ad hoc* soccer observation instrument designed by Amatria et al. (2016). The resulting data were analyzed by polar coordinate analysis (Gorospe and Anguera, 2000), which is a powerful data reduction technique with high predictive power. The results showed significant associations ($Z > 1.96$; $p < 0.05$) between Alonso and players in different positions, a wide sphere of influence on the pitch, both for the start and end of interventions, and a strong link with game interruptions and interceptions and with the use of different technical actions. No significant associations were detected for type of shot. Studies on tactical performance that take account of the multiple factors involved in soccer will lead to better decision-making by coaches and facilitate analysis of a player's true performance.

Keywords: performance analysis, observational methodology, soccer, polar coordinates, Xabi Alonso

INTRODUCTION

Research in soccer has traditionally focused on more mechanical aspects of the game, such as training, biomechanics, health and injury, and physiology. Tactical studies came later, with early studies typically analyzing isolated quantitative data that had little to offer to a game such as soccer, which is marked by constant change and a high degree of uncertainty (Svensson and Drust, 2005). The emergence of increasingly sophisticated data analysis software, however, combined with the development of rigorous methodological approaches, gave researchers the tools to prove or disprove numerous theories about soccer that had been hitherto lacking in scientific credibility.

One particularly prominent area of current research in the field of soccer is how players interact with each other on the pitch (Carling et al., 2008; Duch et al., 2010; Ric et al., 2017). The depth and sophistication of these interactions, which are directly linked to creativity and intra-team coordination (Furley and Memmert, 2015), confirm that soccer, together with its underlying structures, is a complex system (McGarry et al., 2002).

On the whole, complex data are easier to analyze when broken down into measurable units (Bakeman and y Quera, 1996). In the case of soccer, this "simplification" is the only way to validate associations and/or causal relationships that occur during play and to gain insights into behaviors and interactions from a certain distance. In this complex game, the whole is greater than the sum of its parts, hence the importance of understanding how different players interact and relate to each other. One of the main data analysis techniques for analyzing complex interactions in soccer is polar coordinate analysis applied to data gathered through systematic observation (Anguera and Hernández-Mendo, 2015; Castañer et al., 2016, 2017). Polar coordinate analysis can be used to measure the spontaneous behavior of players interacting in their natural environment from the perspective of a given behavior, known as a *focal behavior*.

Polar coordinate analysis has proven to be an effective tool for breaking down the complexity of the game (Lago and Anguera, 2002; Castellano and Hernández-Mendo, 2003; Perea et al., 2012; Robles et al., 2013), and its application to the study of two of the world's best strikers, Lionel Messi (Castañer et al., 2016) and Cristiano Ronaldo (Castañer et al., 2017), has enabled researchers to draw practical conclusions on what makes these players great and in turn to make recommendations for improving both offensive and defensive aspects of play.

Polar coordinate analysis, however, has not yet been tested in midfield players. The bulk of research on midfielders draws on subjective opinions based on both quantitative (Taylor et al., 2005) and qualitative (Wiemeyer, 2003; Thelwell et al., 2006) data analyzed out of context, i.e., with no consideration of other inputs, such as interactions with the ball or other players or the strategic use of space.

Relationships between players are typically visualized through a lens focused on their traditional role according to their position on the pitch. Midfielders are "permeable" to the flow of information from other players on their team; they are a central driving force, positioned at the crossroads between attack and defense. In a recent study on leadership roles in sport, Fransen et al. (2016) showed that players with central positions had a privileged position, which combined with their high tactical responsibilities, positioned them as team leaders. Soccer coach Jürgen Klopp, referring to a Champions League match in 2014, said the following about rival midfielder Xabi Alonso: "Our plan was to take Xabi out of the game. Because if Alonso can play as he wants it is impossible to defend against Madrid."

There have been recent claims that midfielders are the most important members of a soccer team (Duch et al., 2010), particularly at the start and end of an attack (Clemente et al., 2015), due to the highly specific and versatile nature of their position (Kannekens et al., 2010). Additional requirements of midfielders include the combination of sophisticated motor (Di Salvo et al., 2009; Carling et al., 2010) and technical skills (Bloomfield et al., 2007), even in situations when time is critical (Rampinini et al., 2007).

Considering the above, we believe that an in-depth analysis of one of the world's greatest midfield players, Xabi Alonso, in his natural environment through the study of interconnections between multiple, suitably coded, behaviors, is justified. The application of a robust methodological approach, combined with an in-depth, multidimensional analysis of rigorously coded data, will help to provide objective insights into how Alonso interacts with his environment and makes him unique. The specific aim was to analyze Alonso's spatial, technical, and tactical skills during the course of play together with his interactions with the other players on his team.

METHOD

Design

We undertook an systematic observation study (Anguera, 1979). Observational methodology is currently considered to be one of the most suitable methodologies for analyzing spontaneous interactions in sport (Castellano et al., 2012; Anguera and Hernández-Mendo, 2015).

The specific design employed was I/P/M, which stands for Idiographic, Point, and Multidimensional. It was *idiographic* (Anguera et al., 2011) because we studied a single player, *point*, because we applied intrasessional follow-up only (Sánchez-Algarra and Anguera, 2013), and *multidimensional*, because we analyzed behaviors from different dimensions in the observation unit. The observation of behavior was scientifically rigorous because the events observed were fully perceivable and the observers had a non-participatory role.

Participants

The observation sample was a convenience sample (Anguera et al., 2011) formed by a single player, Alonso, during his participation in the final phases of the 2012 UEFA European Championship (UEFA Euro 2012) as a member of the national Spanish team.

As such, the study can be considered a case study (Yin, 2014; Castañer et al., 2016). The behaviors were annotated by analyzing video footage of the matches broadcast on public television. The study thus complies with the ethical principles of the Declaration of Helsinki.

Observation Instrument

We created a modified version of the original *ad hoc* soccer observation instrument built by Amatria et al. (2016). The modifications included new subdivisions that reflected the length of the pitch (**Figure 1**) and more specific definitions of players, technical actions, types of shot, and interruptions (e.g., set plays; **Figure 2**).

The instrument is a combination of a field format and systems of categories (Anguera et al., 2007). It contains eight dimensions, each of which is broken down into a system of exhaustive, mutually exclusive categories.
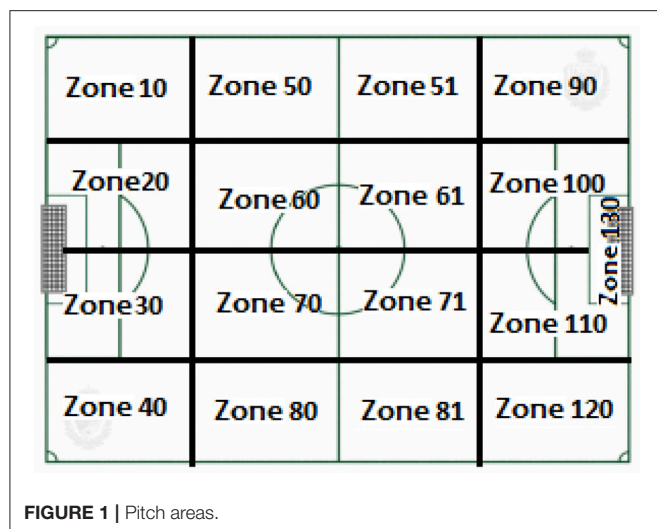
FIGURE 1 | Pitch areas.

Data Annotation and Coding

The data from the video footage were annotated (Hernández-Mendo et al., 2014) using the freely available software program LINCE (v. 1.2.1; Gabín et al., 2012). The interobserver agreement analysis yielded a kappa value of 0.95. The data were concurrent, time-based (type IV) data (Bakeman, 1978).

Two additional software programs were used: GSEQ v5.1 (Bakeman and Quera, 2011) for the lag sequential analysis and HOISAN v. 1.2 (Hernández-Mendo et al., 2012) for the polar coordinate analysis. Lag sequential analysis is needed to calculate the adjusted residual values necessary for polar coordinate analysis.

Data Analysis

Polar coordinate analysis was developed by Sackett (1980) and later improved by Anguera (1997). Although conceptual (Anguera and Losada, 1999) and empirical (Hernández-Mendo and Anguera, 1998, 1999; Gorospe, 1999; Gorospe and Anguera, 2000; Castellano and Hernández-Mendo, 2002) studies started using this method decades ago, its use in sports sciences is recent (Anguera and Hernández-Mendo, 2015). Polar coordinate analysis has attracted increasing attention in recent years, as it offers numerous features that are ideal for the type study proposed in this paper (González et al., 2013; Robles et al., 2014; Echeazarra et al., 2015; López-López et al., 2015; Morillo-Baro et al., 2015; Sousa et al., 2015; Castañer et al., 2016, 2017; López et al., 2016; Aragón et al., 2017; Tarragó et al., 2017; Prudente et al., 2018).

The starting point of any polar coordinate analysis study is the calculation of adjusted residuals using lag sequential analysis (Bakeman, 1978). These are calculated both prospectively (for each positive lag considered) and retrospectively (for each negative lag considered). Following the proposal of Sackett (1980), these residuals are then standardized and used to calculate Zsum statistics (Cochran, 1954), which, in turn, are used to produce a vector map showing the statistical relationships

between the behavior of interest, known as the *focal behavior*, and other behaviors, known as *conditional behaviors*.

The Z_{sum} statistic is calculated using the formula $Z_{sum} = \frac{\sum Z}{\sqrt{n}}$, where Z corresponds to each of the standardized adjusted residuals reflecting the relationship between the focal behavior and a conditional behavior at each of the lags considered, and where n is the number of lags. The Z_{sum} statistic is calculated separately for each prospective and retrospective lag, thereby resulting in a prospective and retrospective value for each conditional behavior. The relationship between the focal behavior and each of the conditional behaviors is depicted by the length and angle of the corresponding vectors (Sackett, 1980).

The length of each vector is equivalent to the hypotenuse of a right-angled triangle in which the respective prospective and retrospective Z_{sum} values correspond to the lengths of the sides adjacent to the right angle, in other words, $Length = \sqrt{(ZsumP)^2 + (ZsumR)^2}$. The angle is calculated via a trigonometric function, by which $\phi = \frac{arc\ sen\ ZsumR}{Length}$.

The resulting value (ϕ) is then transformed according to the quadrant in which the vector is located. The position of the vector is determined by the interplay between the positive and/or negative signs carried by the prospective and retrospective $Zsum$ values, which are respectively plotted along the X and Y axes.

The above calculations are performed in HOISAN (López-López et al., 2015), which also produces the results in the form of easy-to-interpret vector maps. Vectors located in quadrant I show focal behaviors that activate and at the same time are activated by conditional behaviors; those located in quadrant II show focal behaviors that inhibit but are not inhibited by conditional behaviors; those located in quadrant III show focal behaviors that inhibit and at the same time are inhibited by conditional behaviors; and those located in quadrant IV show focal behaviors that activate but are not activated by conditional behaviors (**Figure 3**).

RESULTS

As mentioned in the Methods section, before performing the polar coordinate analysis, we first applied lag sequential analysis to investigate the statistical associations between the categories from the observation instrument.

We have grouped our results into six sections, each describing a different aspect of Alonso's performance and interaction with his teammates and environment during UEFA Euro 2012:

a) Relationship between Alonso and other team members
b) Relationship between Alonso and intervention initiation zone
c) Relationship between Alonso and intervention conclusion zone
d) Relationship between Alonso and game situations (interruptions and interceptions)
e) Relationship between Alonso and technical action (ball contact)
f) Relationship between Alonso and type of shot he takes

No.	Dimension	Categories: codes and brief description
1	Ball possession	PO) possession of ball by team being observed; PC) possession of ball by rival team; Inob) unobservable
2	Player	J0) unidentified player; J1) player 1; J2) player 2; J3) player 3; J4) player 4; J5) player 5; J6) player 6; J7) player 7; J8) player 8; J9) player 9; J10) player 10; J11) player 11; J12) player 12; J13) player 13; J14) player 14; J15) player 15; J16) player 16; J17) player 17; J18) player 18; J19) player 19; J20) player 20; J21) player 21; J22) player 22; J23) player 23; JR) rival player.
3	Move initiation zone	ZI10, ZI20, ZI30, ZI40-safety sector-, ZI50, ZI60, ZI70, ZI80-creation sector in own half, ZI51, ZI61, ZI71, ZI81-creation sector in rival's half -ZI90, ZI100, ZI110, ZI120, ZI130 –definition sector.
4	Move conclusion zone	ZF10, ZF20, ZF30, ZF40-safety sector-, ZF50, ZF60, ZF70, ZF80-creation sector in own half, ZF51, ZF61, ZF71, ZF81-creation sector in rival's half -ZF90, ZI100, ZF110, ZF120, ZF130 –definition sector.
5	Contact with ball	C1) single contact with ball and regulatory throw-in/kick-in; C12) attempt to control the ball with 2 or more touches resulting in loss of ball; C2) control of ball (including catching of ball by goalkeeper) followed by a shot - regardless of whether the ball reaches a team member or is recovered by an opponent; C23) control of ball, followed by dribbling, and loss of ball; C24) control of ball, followed by dribbling, attempt to go around one or more opponents, and loss of ball; C3) control of ball, followed by dribbling and shot - regardless of whether the ball reaches a team member or is recovered by an opponent; C4) control of ball, passing of one or more opponents, and shot – regardless of whether the ball reaches a team member or is recovered by an opponent; C5) Header.
6	Interruptions	GTO) goal by team being observed; GATO) goal against team being observed, FKTO) free kick for team being observed; OTO) offside for team being observed; TITO) throw-in for team being observed; CKTO) corner kick or team being observed; GKTO) goal kick for team being observed; FKATO) free kick against team being observed; OATO) offside against team being observed; TIATO) throw-in against team being observed; CKATO) corner kick against team being observed; GKATO) goal kick against team being observed; NK) kick-off/neutral kick; KO) kick-off; EFH) end of first half; EM) end of match.
7	Interceptions	LB) loss of ball; RB) recovery of ball; OIC) occasional interception with continuation of play
8	Type of shot	SG) shot resulting in goal; SI) shot intercepted by opponent other than the goalkeeper; SBP) shot between the posts not resulting in a goal; SWP) shot wide of the posts; SSG) shot saved or cleared by the goalkeeper; HEG) header ending in a goal; HIG) header intercepted by opponent other than the goalkeeper; HBP) Header between the posts not resulting in a goal; HWP) Header wide of the posts; HBG) Header blocked or cleared by the goalkeeper.

FIGURE 2 | Observation instrument. Source: Modified from Amatria et al. (2016).

Alonso (code J14) was established as the focal behavior for all the above analyses.

Relationship Between Alonso and Other Players

For this analysis, we investigated the relationship between Alonso/J14, defined as the focal behavior or category, and the other players on the Spanish national team—J0 (unidentified player), J1 (Iker Casillas), J2 (Raúl Albiol), J3 (Gerard Piqué), J4 (Javi Martinez), J5 (Juanfran), J6 (Iniesta), J7 (Pedro), J8 (Xavi Hernández), J9 (Fernando Torres), J10 (Cesc Fábregas), J11 (Álvaro Negredo), J12 (Víctor Valdés), J13 (Juan Mata), J15 (Sergio Ramos), J16 (Busquets), J17 (Arbeloa), J18 (Jordi Alba), J19 (Fernando Llorente), J20 (Cazorla), J21 (David Silva), J22 (Jesús Navas), and J23 (Reina)—and rival players (JR), defined as the conditional behaviors or categories. The
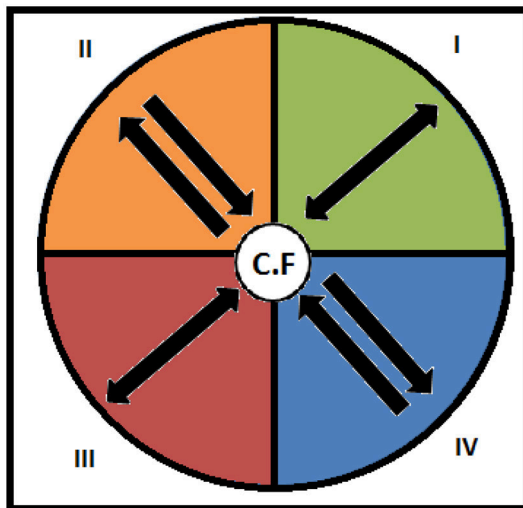
FIGURE 3 | Activating and inhibitory relationships between focal and conditional behaviors.

TABLE 1 | Polar coordinate analysis results for the relationship between the focal category Alonso (J14) and other players.

Category	Quadrant	Prospective perspective	Retrospective perspective	Radius	Angle
J0	IV	0.34	−0.11	0.36	342.01
J1	II	−0.99	2.89	3.05*	108.89
J3	II	−1.53	2.11	2.61*	126.06
J4	III	−3.73	−3.58	5.17*	223.84
J6	IV	2.22	−0.33	2.25*	351.53
J7	IV	0.97	−0.6	1.14	328.42
J8	I	0.95	0.11	0.96	6.69
J9	III	−1.01	−2.88	3.05*	250.63
J10	IV	1.44	−2.21	2.64*	302.96
J11	I	3.12	1.46	3.44*	25.07
J13	II	−0.72	1.77	1.91	112.18
J15	II	−2.51	0.64	2.59*	165.7
J16	I	1.5	1.36	2.02*	42.13
J17	IV	0.5	−0.05	0.5	353.83
J18	I	1	1.37	1.69	53.92
J20	III	−2.8	−3.36	4.38*	230.14
J21	IV	0.73	−0.72	1.03	315.53
J22	I	0.13	0	0.13	0
JR	III	−2.58	−0.86	2.72*	198.41

*$p < 0.05$.

aim was to investigate how Alonso interacted with teammates and rivals during the matches analyzed in UEFA Euro 2012.

The results (**Table 1** and **Figure 4**) show that both J11 (Álvaro Negredo), with a radius of 3.44 and an angle of 25.07°, and J16 (Sergio Busquets), with a radius of 2.02 and an angle of 42.13, were located in quadrant I, where the focal behavior activates the conditional behavior both prospectively and retrospectively (mutual activation).

J1 (Iker Casillas), with a radius of 3.05 and an angle of 108.89°, J3 (Gerard Piqué), with a radius of 2.61 and an angle of 126.06°, and J15 (Sergio Ramos), with a radius of 2.59 and an angle of 165.7°, were positioned in quadrant II, where the focal behavior inhibits but is not inhibited by the conditional behavior.

Quadrant III contains the conditional categories J4 (Javi Martinez), J9 (Fernando Torres), J20 (Cazorla), and JR (rival player), with respective radii of 5.17, 3.05, 4.38, and 2.72, and respective angles of 223.84, 250.63, 230.14, and 198.41°. In this quadrant the focal and conditional behaviors are mutually inhibited.

Finally, J6 (Iniesta), with a radius of 2.25 and an angle of 351.53°, and J10 (Cesc Fábregas), with a radius of 2.64 and an angle of 302.96°, were located in quadrant IV, where the focal behavior activates but is not activated by conditional behaviors -ZI71 and ZI10-.

Relationship Between Alonso and Intervention Initiation Zone

In this analysis, we investigated the relationship between Alonso (J14) and the different areas of the pitch in which an intervention involving this player was launched (ZI10, ZI20, ZI30, ZI40, ZI50, ZI60, ZI70, ZI80, ZI51, ZI61, ZI71, ZI81, ZI90, ZI100, ZI110, ZI120, and ZI130). The aim was to identify Alonso's use of space during the course of play.

FIGURE 4 | Vector map showing relationships between focal category (Alonso/J14) and other players.

The results (**Table 2** and **Figure 5**) show that both ZI51, with a radius of 2.49 and an angle of 55.17°, and ZI71, with a radius of 2.07 and an angle of 44.74°, were located in quadrant I (mutual activation between focal and conditional behaviors).

TABLE 2 | Polar coordinate analysis results for the relationship between the focal category Alonso (J14) and intervention initiation zones.

Category	Quadrant	Prospective perspective	Retrospective perspective	Radius	Angle
ZI10	II	−2.06	2.41	3.18*	130.49
ZI20	II	−1.37	1.74	2.21*	128.35
ZI30	II	−1.83	0.04	1.83	178.6
ZI40	III	−2.99	−0.9	3.12*	196.75
ZI50	I	0.34	0.46	0.57	53.67
ZI60	IV	0.28	−0.65	0.71	293.34
ZI70	I	1.45	0.21	1.47	8.23
ZI80	III	−3.11	−0.85	3.22*	195.29
ZI51	I	1.42	2.04	2.49*	55.17
ZI61	IV	5.18	−0.59	5.21*	353.5
ZI71	I	1.47	1.46	2.07*	44.74
ZI81	IV	0.82	−0.37	0.9	335.72
ZI90	III	−1.5	−1.26	1.96*	220.09
ZI100	III	−1.76	−1.1	2.07*	211.94
ZI110	IV	0.08	−0.8	0.81	275.71
ZI120	III	−3.58	−3.2	4.8*	221.83
ZI130	II	−0.64	1.04	1.22	121.72

*$p < 0.05$.

FIGURE 5 | Vector map showing relationships between the focal category (Alonso/J14) and intervention initiation zones.

Quadrant II, which shows conditional behaviors that are inhibited by but do not inhibit the focal behavior, contains the two zones located to the left of the Spanish team's goal: ZI10, with a radius of 3.18 and an angle of 130.49°, and ZI20, with a radius of 2.21 and an angle of 128.35°.

Quadrant III (mutual inhibition quadrant) contained categories ZI40, with a radius of 3.12 and an angle of 196.5°; ZI80, with a radius of 3.22 and an angle of 195.29°; ZI90, with a radius of 1.96 and an angle of 220.09°; ZI100, with a radius of 2.07 and an angle of 211.94°; and ZI120, with a radius of 4.8 and an angle of 221.83°.

Finally, ZI61, with a radius of 5.21 and an angle of 353.5°, was located in quadrant IV, where the focal category activates but is not activated by the conditional behavior.
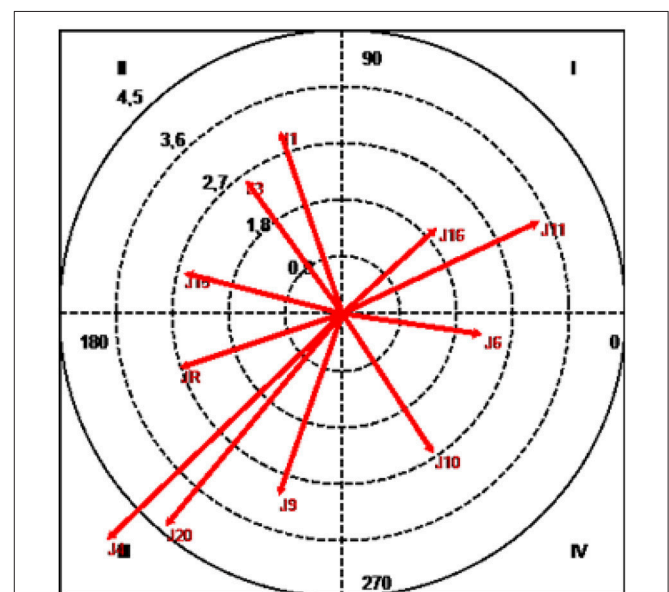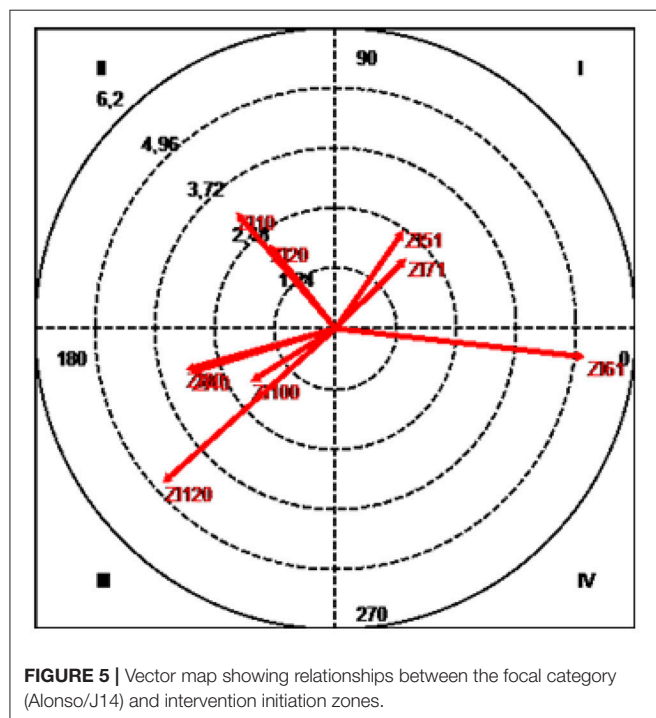
Relationship Between Alonso and Initiation Conclusion Zones

In this analysis, we studied the relationship between Alonso (J14) and the different areas of the pitch in which interventions in which he was involved ended (ZF10, ZF20, ZF30, ZF40, ZF50, ZF60, ZF70, ZF80, ZF51, ZF61, ZF71, ZF81, ZF90, ZF100, ZF110, ZF120, and ZF130). The aim was to analyze how Alonso interacted with these zones when the Spanish team was attacking.

The results (**Table 3** and **Figure 6**) show that quadrant I (mutual activation) contained categories ZF61, with a radius of 4.05 and an angle of 25.63°; ZF50, with a radius of 2.07 and an angle of 89.01°; ZF60, with a radius of 2.47 and an angle of 77.22°; and ZF70, with a radius of 1.98 and an angle of 56.17°.

Quadrant II, in turn, which shows conditional behaviors that are inhibited by but do not inhibit the focal behavior, contained ZF10 (safety sector in team's own half), with a radius of 2.17 and an angle 146.51°.

Quadrant III (mutual inhibition quadrant) contained ZF30, with a radius of 2.06 and an angle of 200.56°; ZF40, with a radius of 2.91 and an angle of 207.35°; ZF80, with a radius of 3.29 and an angle of 216.5°; ZF100, with a radius of 2.85 and an angle of 243.11°; and ZF120, with a radius of 4.53 and an angle of 241.31°.

Finally, quadrant IV, which shows conditional behaviors that are activated by but do not activate the focal behavior, contained ZF81, with a radius of 2.54 and an angle of 283.62, and ZF90, with a radius of 3.42 and an angle of 275.18°.

Relationship Between Alonso and Game Interruptions and Interceptions

For this analysis, we analyzed the relationship between Alonso (J14) and different aspects related to game interruptions and interceptions—GTO (goal by team being observed), GATO (goal against team being observed), FKTO (free kick for team being observed), OTO (offside for team being observed), TITO (throw-in for team being observed), CKTO (corner kick or team being observed), GKTO (goal kick for team being observed), FKATO (free kick against team being observed), OATO (offside against team being observed), TIATO (throw-in against team being observed), CKATO (corner kick against team being observed), GKATO (goal kick against team being observed), NK (kick-off/neutral kick), KO (kick-off), EFH (end of first half), EM (end of match), LB (loss of ball), RB (recovery of ball), and OIC (occasional interception with continuation of play)—The aim was to investigate Alonso's involvement in these situations.

The results (**Table 4** and **Figure 7**) show that GKTO (goal kick for team being observed), with a radius of 2.07 and an angle of

TABLE 3 | Polar coordinate analysis results for the relationship between the focal category Alonso (J14) and intervention conclusion zones.

Category	Quadrant	Prospective perspective	Retrospective perspective	Radius	Angle
ZF10	II	−1.81	1.2	2.17*	146.51
ZF20	II	−1.38	0.52	1.48	159.26
ZF30	III	−1.93	−0.72	2.06*	200.56
ZF40	III	−2.58	−1.34	2.91*	207.35
ZF50	I	0.04	2.07	2.07*	89.01
ZF60	I	0.55	2.41	2.47*	77.22
ZF70	I	1.1	1.64	1.98*	56.17
ZF80	III	−2.65	−1.96	3.29*	216.5
ZF51	I	1.03	1.57	1.88	56.76
ZF61	I	3.65	1.75	4.05*	25.63
ZF71	I	0.88	1.74	1.95	63.14
ZF81	IV	0.6	−2.47	2.54*	283.62
ZF90	IV	0.31	−3.4	3.42*	275.18
ZF100	III	−1.29	−2.54	2.85*	243.11
ZF110	IV	0.35	−1.75	1.78	281.28
ZF120	III	−2.17	−3.97	4.53*	241.31
ZF130	II	−0.96	0.7	1.19	144.04

*$p < 0.05$.

TABLE 4 | Polar coordinate analysis results for the relationship between focal category Alonso (J14) and game situations (interruptions and interceptions).

Category	Quadrant	Prospective perspective	Retrospective perspective	Radius	Angle
GTO	II	−0.36	0.51	0.62	125.3
FKTO	II	−1.15	3.18	3.38*	109.87
OTO	I	0.5	0.54	0.74	47.21
TITO	III	−2.43	−3.17	3.99*	232.55
CKTO	III	−3.64	−0.28	3.65*	184.43
GKTO	I	0.52	2	2.07*	75.48
FKATO	II	−0.14	0.64	0.66	102.61
OATO	II	−1.83	0.13	1.83	175.8
TIATO	II	−1.2	1.37	1.82	131.12
GKATO	I	2.37	0.56	2.44*	13.37
NK	III	−0.19	−1.07	1.09	259.84
NC	I	1.22	0.5	1.32	22.13
EFH	I	2.47	0.34	2.5*	7.93
EM	II	−0.52	0.62	0.81	130.29
LB	IV	3.79	−0.32	3.81*	355.21
RB	I	1.12	0.47	1.22	22.51
OIC	III	−1.6	−1.95	2.52*	230.62

*$p < 0.05$.

FIGURE 6 | Vector map showing relationships between focal category (Alonso/J14) and intervention conclusion zones.

FIGURE 7 | Vector map showing the relationships between the focal category (Alonso/J14) and game situations (interruptions and interceptions).

75.48°; GKATO (goal kick against team being observed), with a radius of 2.44 and an angle of 13.37°; and EFH (end of first half), with a radius of 2.5 and an angle of 7.93°, were all located in the mutual activation quadrant I.

Quadrant II, in which conditional behaviors are inhibited by but do not inhibit the focal behavior, contained FKTO (free kick

for team being observed), with a radius of 3.38 and an angle of 109.87°.

Quadrant III, where the focal and conditional behaviors mutually inhibit each other, contained TITO (throw-in for team being observed) with a radius of 3.99 and an angle of 232.55°; CKTO (corner kick for team being observed) with a radius of

3.65 and an angle of 184.43°; and OIC (occasional interception with continuation of play) with a radius of 2.52 and an angle of 230.62°.

Finally, the conditional category LB (loss of ball), with a radius of 3.81 and an angle of 355.21°, was located in quadrant IV, where focal behaviors activate but are not activated by conditional behaviors.

Relationship Between Alonso and Technical Actions (Ball Contact)

For this analysis, we studied the relationship between Alonso and the different categories in the ball contact dimension C1 (single contact with ball and regulatory throw-in/kick-in), C12 (attempt to control the ball with 2 or more touches resulting in loss of ball), C2 (control of ball, including catching of ball by goalkeeper, followed by a shot—regardless of whether the ball reaches a team member or is recovered by an opponent), C23 (control of ball, followed by dribbling, and loss of ball), C24 (control of ball, followed by dribbling, attempt to go around one or more opponents, and loss of ball), C3 (control of ball, followed by dribbling and shot—regardless of whether the ball reaches a team member or is recovered by an opponent), C4 (control of ball, passing of one or more opponents, and shot—regardless of whether the ball reaches a team member or is recovered by an opponent), C5 (header). The aim was to investigate Alonso's technical skills.

Quadrant II, where the focal behavior inhibits the presence of the conditional behavior prospectively and activates it retrospectively, contained the category C5 (header), with a radius of 3.06 and an angle of 170.58° (**Table 5** and **Figure 8**).

The mutual inhibition quadrant, quadrant III, contained the conditional behavior C1 (single contact with ball and regulatory throw-in), with a radius of 2.29 and an angle of 192.65°.

Relationship Between Alonso and Type of Shot

For this analysis, we studied the relationship between Alonso and the different categories in the type of shot dimension (SG, SI, SBP, SWP, SSG, HEG, HIG, HBP, HWP, and HBG). The aim was to

identify significant associations between Alonso and the type of shots he takes.

As shown in **Table 6**, the results were not significant.

DISCUSSION

Polar coordinate analysis has been shown on multiple occasions to be an ideal technique for dissecting the complexity underlying the relationships between different players of a team. It is a powerful data reduction technique that produces a manageable set of vectorial parameters that are graphically represented in a vector map that shows both activating and inhibitory relationships between variables of interest (Gorospe and Anguera, 2000). Specifically, the map shows how a given category, the *focal behavior*, is connected to all other categories within a category system. In this vector map, the angle of the

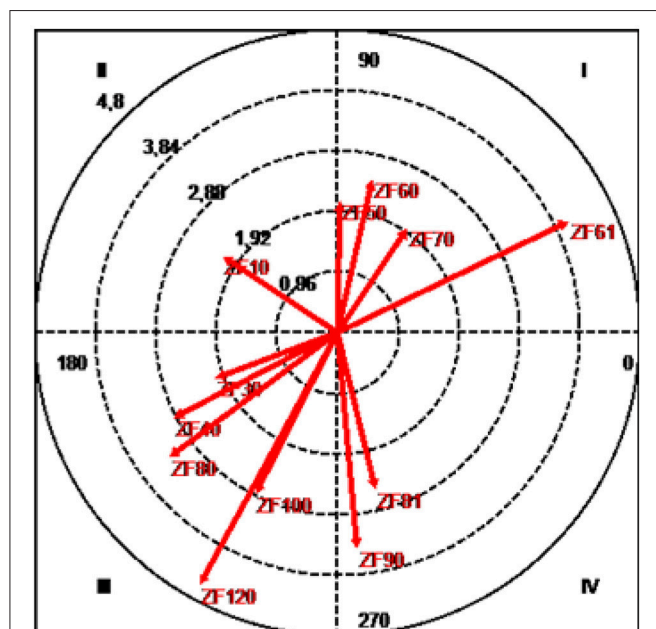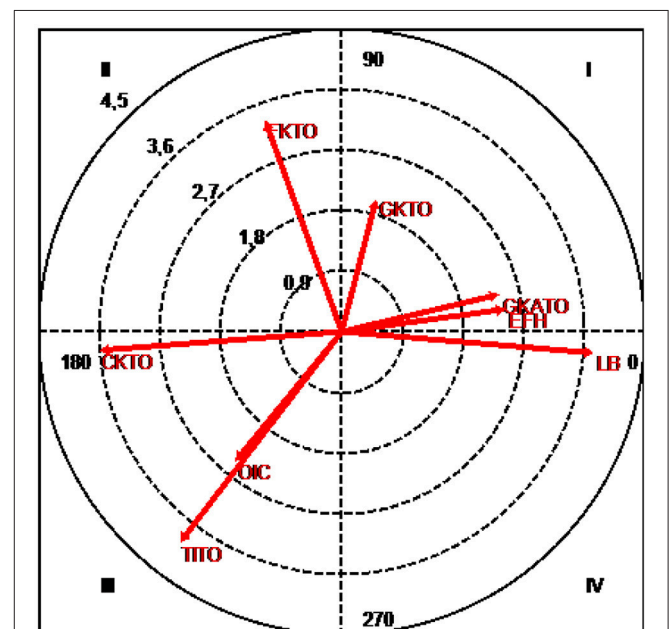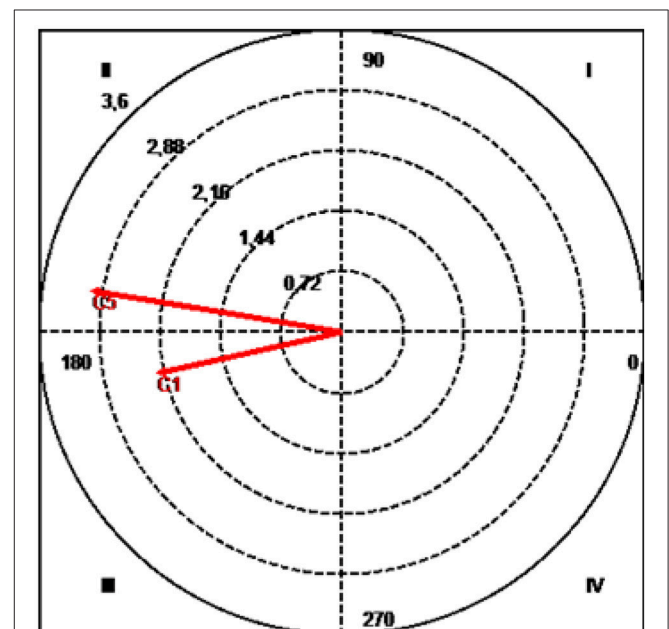**FIGURE 8 |** Vector map showing relationships between focal category (Alonso/J14) and type of contact.

TABLE 5 | Polar coordinate analysis results for the relationship between the focal category (Alonso/J14) and technical actions.

Category	Quadrant	Prospective perspective	Retrospective perspective	Radius	Angle
C1	III	−2.23	−0.5	2.29*	192.65
C12	I	1.36	1.07	1.74	38.2
C2	I	1.04	0.01	1.04	0.49
C23	IV	1.78	−0.26	1.79	351.69
C24	IV	0.11	−0.55	0.57	281.4
C3	IV	1.57	−0.29	1.59	349.64
C4	II	−0.09	1.04	1.04	95.17
C5	II	−3.02	0.5	3.06*	170.58

*p < 0.05.

TABLE 6 | Polar coordinate analysis results for focal category (Alonso/J14) and type of shot.

Category	Quadrant	Prospective perspective	Retrospective perspective	Radius	Angle
SG	III	−1.17	−0.35	1.22	196.58
SI	III	−0.41	−0.99	1.07	247.49
SWP	I	1.21	0.91	1.51	36.84
PT	I	1.16	0.22	1.18	10.71
HEG	III	−0.78	−0.01	0.78	180.65
HIG	II	−0.98	0.04	0.98	177.92
HWP	III	−1.19	−0.1	1.19	184.94
HBG	I	0.46	0.33	0.57	35.7

*p < 0.05.

vector indicates the nature of the relationship between two behaviors or categories and the radius indicates the strength of the relationship (Anguera and Hernández-Mendo, 2015; Morillo-Baro et al., 2015).

Most studies of performance in soccer to date have focused on the performance of the team as a whole, with little attention given to individual performance or interactions between team members. The most notable studies to date of individual performance are two studies by Castañer et al. (2016, 2017) that used polar coordinate analysis to study and compare Messi and Ronaldo's use of motor skills in relation to goal scoring. Methodologically rigorous studies of the tactics used by midfielders are also lacking. As pointed out by Sampaio and Maças (2012), Memmert (2010) ,Memmert et al. (2017) and Castañer et al. (2016, 2017), in order to understand the performance of a team as a whole, it is first necessary to understand how the different members of the team interact with each other.

Alonso and His Team Mates

Our analysis revealed mutual activation between Alonso (J14) and both Negredo (J11) and Busquets (J16), showing that Alonso's interactions with his teammates extend beyond the midfield area into both the attacking and shooting areas of the pitch. This finding supports the observation by Kannekens et al. (2010) that midfielders must act as the link between a team's attackers and defenders.

We also observed an interesting retrospective activation of categories J3 (Piqué) and J15 (Ramos), two defenders with outstanding tactical and technical prowess who have an important role in the building of attacks. This finding suggests that Alonso establishes "microsocieties" in the early stages of attack by first interacting with Piqué and Ramos in a small area and then with players at the first line of attack, as shown by the prospective activation of J10 (Cesc Fàbregas) and J6 (Iniesta) in quadrant IV. These interactions reveal mastery of two skills that are of tactical significance in midfield play. The first is high proficiency in mesostructural aspects of the game, such as the ability to maintain possession of the ball in small spaces (Rampinini et al., 2007), while the second is mastery of the 360° turn (Bloomfield et al., 2007), which is a highly valued skill in a midfield player, as it enhances peripheral vision and spatial intelligence (Gardner, 2016) and allows players to adapt to the constant changes around them. To master these two skills, the player must not only excel in the technical skills required of a midfielder (Taylor et al., 2005) but also demonstrate creative and attacking qualities, such as versatility, good on-the-ball movement, and tactical positioning.

Coaches should take these factors into account in order to design effective defensive tactics. Applying pressure to Alonso and likely receivers of a pass from him when the team is in possession of the ball, for example, could help to increase the effectiveness of the defense. In addition, these pressure maneuvers could cause the observed player to move into residual spaces of the playing field away from their usual range of action, where their role within the team could move to a less important role.

Alonso and His Interaction With Different Areas of the Pitch

As shown in the vector map in **Figure 5**, we detected mutual activation between Alonso and zones ZI51 and ZI71, which are two areas of creative play. Activation of these zones in quadrant I is interesting, as Alonso is largely considered to be a defensive or holding midfielder, and it further supports the idea that Alonso's role extends beyond what is traditionally expected of a defensive midfielder (Wiemeyer, 2003). The above interactions, combined with Alonso's use of space, suggest that Alonso represents a new style of midfielder, one who masters multiple aspects of offensive and defensive play.

The relationship between Alonso and areas of the pitch where attacks are launched, zones ZI10 and ZI20, was relatively unremarkable, with respective vector radius lengths of 3.18 and 2.21. These vectors were located in quadrant II, where the focal behavior activates the presence of the conditional behavior retrospectively but not prospectively. This observation is directly related to the early launch of an attack in the defensive areas of the pitch. When the backline defenders gain possession of the ball, Alonso activates his movements at the next line of play (the midfield area) in the hope of receiving the ball and moving it up to the next line of attack. Alonso's position at the tip of the triangle formed by these areas and their respective occupants does not appear to be arbitrary. Recent studies claim that these geometric patterns are consistent with Voronoi diagrams, which are diagrams based on mathematical spatial partitioning algorithms that have been recently applied to the study of soccer (Sumpter, 2016).

On analyzing the areas of the pitch where interventions involving Alonso ended, we observed that the mutual activation between Alonso and zones ZF50, ZF60, and ZF70 in quadrant I is consistent with the findings of James et al. (2002), who stated that the safety pass was one of the most frequently used technical actions by defensive midfielders. Soccer thus is not just about ball control but also about creativity and the strategic use of space, as indicated by the retrospective activation of ZF81 in quadrant IV. In view of the above findings, it is conceivable that the microsociety formed by Ramos (J15), Piqué (J3), and Alonso (J14) is designed to create space in distant areas of the pitch by drawing defenders to the passing triangle formed by these three players. The tactic is clear: the team distracts the defenders through on-the-ball interactions and then delivers the ball to the opposite side of the pitch (ZF81) to continue the attack. Essentially, this tactical play involves changes of direction, which were described in Garganta's (1997) study of offensive tactics in elite soccer some 20 years ago. Distraction and dummy moves are two important tactics employed by Alonso.

At a practical level, coaches looking to defend against such tactics should focus on studying their rivals' use of space. Placing more defenders in strategic areas, including those located some distance from the ball, will improve a team's chances of a successful defense.

Alonso's Role in Game Situations (Interruptions and Interceptions)

Versatility in set-play situations is one of Alonso's greatest attributes, as indicated by the mutual activation observed between Alonso and GKTO (goal kick for Spanish team) and GKATO (corner kick against Spanish team) in quadrant I. Alonso excels in aerial play and is a particular good header. This attribute is particularly interesting for the Spanish team, which had the shortest average squad at UEFA Euro 2012. The most interesting observation related to the association between Alonso and set-play situations was the retrospective inhibitory relationship in quadrant II, which showed that Alonso tended to be the first player to receive the ball after an interruption of play. We believe that this finding is interesting for two reasons. First, Alonso would appear to be the player of choice to receive the ball following a set play and second, it supports claims by Fransen et al. (2016) that players in central positions are the most prominent players in the building of an attack due to their ability to coordinate the flow of information between the different members of the team. In this case, Alonso would appear to be responsible for triggering the flow of information that then diffuses through the other members of the team.

At a practical level, placing more defenders on Alonso could limit his freedom to move and the time he has to take decisions. Another possible action would be a possible individual marking on the player.

Alonso and Technical Actions

Based on the data from quadrant III, where focal and conditional behaviors are mutually inhibitory, we cannot draw any conclusions on the relationship between Alonso and technical actions during UEFA Euro 2012. Our findings in general, however, illustrate that Alonso draws on a wide repertoire of technical resources that are not typically associated with midfield players (Wiemeyer, 2003; Taylor et al., 2005; Thelwell et al., 2006), and is closely involved in set plays, such as free kicks and corner kicks. Alonso is a multidisciplinary and highly skilled player. The decision on which technical action to execute needs to be taken fast, but this speed of decision needs to be accompanied by accurate execution, which in turn requires precision of both motor and cognitive skills. It is this fine balance that makes a great midfielder. We agree with Kannekens et al. (2009) that soccer training models require a paradigm shift and need to center on tactical aspects of the game. Broadly speaking, we suggest the incorporation of training drills focused on teaching players to take fast, innovative decisions in pressure situations in order to adapt to the changing circumstances of the game.

Increasing players' understanding of technical and tactical concepts will help defenders adapt to the new style of midfielder represented by Xabi Alonso. Tactics such as pressure, dissuasion, and timing are likely to be more effective than the traditional tackle.

Alonso and Shots

Although Alonso scored two goals at UEFA Euro 2012, our analysis did not detect any significant results for the relationship between Alonso and type of shot.

Nonetheless, we can draw on the data compiled to characterize this relationship. When the Spanish team is in possession of the ball, Alonso establishes polyvalent relationships with players in different positions, attaining thus a wide sphere of influence. In this respect, his role differs considerably from that of a classic midfielder, although he retains his defensive function. In addition, he combines rapid decision-making with high precision, as would be expected of a midfielder with a wide repertoire of skills. Finally, Alonso is the player of choice to receive the ball during set plays.

CONCLUSIONS AND FUTURE WORK

We have analyzed the different relationships that Alonso establishes with his teammates on the Spanish national soccer team and studied his use of space and technical-tactical skills during the course of play. In addition, we have made some practical recommendations based on our findings that could be of interest to coaches at different levels. Our study also highlights the potential of observational methodology (in particular polar coordinate analysis) as a means of studying the spontaneous behavior of players in their natural setting. Future studies should continue to analyze the qualities of Alonso and other players to further uncover on the complex structures underlying interactions between different members of a team. The findings will undoubtedly result in a greater understanding of soccer as a whole.

AUTHOR CONTRIBUTIONS

RM: Collected the data, reviewed the literature, and wrote the manuscript; MA: Collected and analyzed data and performed statistical analyzes.

ACKNOWLEDGMENTS

REFERENCES

Amatria, M., Lapresa, D., Arana, J., Anguera, M. T., and Garzón, B. (2016). Optimization of game formats in U-10 soccer using logistic regression analysis. *J. Hum. Kinet.* 24, 163–171. doi: 10.1515/hukin-2016-0047

Anguera, M. T. (1979). Observational typology. *Qual. Quant.* 13, 449–484.

Anguera, M. T. (1997). "From prospective patterns in behaivor to joins analysis with retrospective perspective," in *Colloque sur Invitation "Méthodologie Dánalyse des Interactions Sociales"* (París: Université de la Soborna).

Anguera, M. T., Blanco-Villaseñor, A., Hernández-Mendo, A., and Losada, J. L. (2011). Diseños observacionales: ajuste y aplicación en psicología del deporte. [Observational designs: Their suitability and application in sport psychology.] *Cuad. Psico. Dep.* 11, 63–76.

Anguera, M. T., and Hernández-Mendo, A. (2015). Data analysis techniques in observational studies in sport sciences. *Cuad. Psico. Dep.* 15, 13–30. doi: 10.4321/S1578-84232015000100002

Anguera, M. T., and Losada, J. L. (1999). "Reducción de datos en marcos de conducta mediante la técnica de coordenadas polares," in *Observación de la Conducta Interactiva en Situaciones Naturales: Aplicaciones,* ed M. T. Anguera (Barcelona: E.U.B.), 163–188.

Anguera, M. T., Magnusson, M. S., and Jonsson, G. K. (2007). Instrumentos no estándar. [Non-standard instruments]. *Avances en Medición* 5, 63–82.

Aragón, S., Lapresa, D., Arana, J., Anguera, M. T., and Garzón, B. (2017). An example of the informative potential of polar coordinate analysis: sprint tactics in elite 1500 m track events. *Meas. Phys. Educ. Exerc. Sci.* 16, 279–286. doi: 10.1080/1091367X.2016.1245192

Bakeman, R. (1978). "Untangling streams of behavior: sequential analysis of observation data," in *Observing Behavior: Data Collection and Analysis Methods, Vol. 2,* ed G. P. Sackett (Baltimore, MD: University Park Press), 63–78.

Bakeman, R., and y Quera, V. (1996). *Análisis de la Interacción: Análisis Secuencial con SDIS y GSEQ [Analyzing Interaction: Sequential Analysis with SDIS and GSEQ].* Madrid: Ra–Ma.

Bakeman, R., and Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences.* Cambridge, UK: Cambridge University Press.

Bloomfield, J., Polman, R., and O'Donoghue, P. (2007). Physical demands of different positions in FA Premier League soccer. *J. Sport Sci. Med.* 6, 63–70.

Carling, C., Espié, V., Le Gall, F., Bloomfield, J., and Jullien, H. (2010). Work-rate of substitutes in elite soccer: a preliminary study. *J. Sci. Med. Sport* 13, 253–255. doi: 10.1016/j.jsams.2009.02.012

Carling, C. J., Bloomfield, L., Nelsen, and Reilly, T. (2008). The role of motion analysis in elite soccer: contemporary performance measurement techniques and work rate data. *Sports Med.* 38, 839–862. doi: 10.2165/00007256-200838100-00004

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Canton, A., and Hileno, R. (2016). Goal scoring in soccer: a polar coordinates analysis of motor skills used by Lionel Messi. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Fernandes, T., and Hileno, R. (2017). Mastery in goal scoring, T-pattern detection and polar coordinate analysis of motor skills used by Lionel Messi and Cristiano Ronaldo. *Front. Psychol.* 8:741. doi: 10.3389/fpsyg.2017.00741

Castellano, J., Casamichana, D., and Lago, C. (2012). The use of match statistics that discriminate between successful and unsuccessful soccer teams. *J. Hum. Kinet.* 31, 139–147. doi: 10.2478/v10078-012-0015-7

Castellano, J., and Hernández-Mendo, A. (2002). Aportaciones del análisis de coordenadas polares en la descripción de las transformaciones de los contextos de interacción defensivos en fútbol [Contributions of polar coordinate analysis to the description of transformations during defensive interactions in soccer]. *Kronos: Revista Universitaria de la Actividad Física y el Deporte* 1, 42–48.

Castellano, J., and Hernández-Mendo, A. (2003). El análisis de coordenadas polares para la estimación de relaciones en la interacción motriz en fútbol [Polar coordinate analysis for estimating relationships between motor skill interactions in soccer]. *Psicothema* 15, 569–574.

Clemente, F. M., Martins, F. M. L., Wong, D. P., Kalamaras, D., and Mendes, R. S. (2015). Midfielder as the prominent participant in the building attack: a network analysis of national teams in FIFA World Cup 2014. *Int. J. Peform. Anal. Sport* 15, 704–722. doi: 10.1080/24748668.2015.11868825

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* 10, 417–451.

Di Salvo, V., Gregson, W., Atkinson, G., Tordoff, P., and Drust, B. (2009). Analysis of high intensity activity in Premier League soccer. *Int. J. Sports Med.* 30, 205–212. doi: 10.1055/s-0028-1105950

Duch, J., Waitzman, J. S., and Amaral, L. A. N. (2010). Quantifying the performance of individual players in a team activity. *PLoS ONE* 5:e10937. doi: 10.1371/journal.pone.0010937

Echeazarra, I., Castellano, J., Usabiaga, O., and Hernández-Mendo, A. (2015). Diferencias en el uso estratégico del espacio en categorías infantil y cadete en fútbol: una aplicación del análisis de coordenadas polares [Differences in the strategic use of space in under 14 and under 16 soccer: a polar coordinate analysis]. *Cuad. Psico. Dep.* 15, 169–180. doi: 10.4321/S1578-84232015000100017

Fransen, K., Haslam, S. A., Mallett, C. J., Steffens, N. K., Peters, K., and Boen, F. (2016). Leading from the centre: a comprehensive examination of the relationship between central playing positions and leadership in sport. *PLoS ONE* 11:e0168150. doi: 10.1371/journal.pone.0168150

Furley, P., and Memmert, D. (2015). Creativity and working memory capacity in sports: working memory capacity is not a limiting factor in creative decision making amongst skilled performers. *Front. Psychol.* 6:115. doi: 10.3389/fpsyg.2015.00115

Gabín, B., Camerino, O., Anguera, M. T., and Castañer, M. (2012). Lince: multiplatform sport analysis software. *Proc. Soc. Behav. Sci.* 46, 4692–4694. doi: 10.1016/j.sbspro.2012.06.320

Gardner, H. (2016). *Inteligencias Múltiples. La Teoría en la Práctica [Multiple Intelligence. Theory in Practice].* Barcelona: Paidós.

Garganta, J. (1997). *Modelaçao táctica do jogo de futebol. Estudo da Organizaçao da Fase Ofensiva em Equipas de Alto Rendimento [The Study of the Organization of the Offensive Phase in Elite Soccer Teams].* Doctoral thesis, University of Porto, Porto.

González, A., Botejara, J., Puñales, L., Trejo, A., and Ruy, E. (2013). Análisis de la finalización del ataque en partidos igualados en balonmano de alto nivel mediante coordenadas polares [Analysis of attack completion in close elite handball matches using polar coordinate analysis]. *Rev. Int. Cienc. Dep.* 9, 71–89.

Gorospe, G. (1999). *Observation and Analysis in Singles Tennis: Contributions from Sequential Analysis and Polar Coordinate Analysis.* Doctoral thesis, University of País Vasco, San Sebastián.

Gorospe, G., and Anguera, M. T. (2000). Retrospectivity in polar coordinates analysis: application to tennis. *Psicothema* 12, 279–282.

Hernández-Mendo, A., and Anguera, M. T. (1998). "Análisis de coordenadas polares en el estudio de las diferencias individuales de la acción de juego [Polar coordinate analysis to study individual differences during play]," in *Perspectivas Actuales en la Investigación de las Diferencias Individuales,* eds M. P. Sánchez and M. A. López Quiroga Estévez (Madrid: Centro de Estudios Ramón Areces), 84–88.

Hernández-Mendo, A., and Anguera, M. T. (1999). "Aportaciones de análisis de coordenadas polares a los deportes de equipo [Contributions of polar coordinate analysis to team sports]," in *La Psicología del Deporte en España al Final del Milenio,* ed F. Guillén (Las Palmas: Universidad de Las Palmas de Gran Canaria), 169–175.

Hernández-Mendo, A., Castellano, J., Camerino, O., Jonsson, G., Blanco-Villaseñor, A., Lopes, A., et al. (2014). Observational software, data quality control and data analysis. *Rev. Psicol. Dep.* 23, 111–121.

Hernández-Mendo, A., López-López, J. A., Castellano, J., Morales-Sánchez, V., and Pastrana, J. L. (2012). IT program for use in methodology observacional. *Cuad. Psicol. Dep.* 12, 55–78.

James, N., Mellalieu, S. D., and Hollely, C. (2002). Analysis of strategies in soccer as a function of European and domestic competition. *Int. J. Peform. Anal. Sport* 2, 85–103. doi: 10.1080/24748668.2002.11868263

Kannekens, R., Elferink-Gemser, M. T., and Visscher, C. (2009). Tactical skills of world-class youth soccer teams. *J. Sports Sci.* 27, 807–812. doi: 10.1080/02640410902894339

Kannekens, R., Elferink-Gemser, M. T., and Visscher, C. (2010). Positioning and deciding: key factors for talent development in soccer. *Scand. J. Med. Sci. Sports* 21, 846–852. doi: 10.1111/j.1600-0838.2010.01104.x

Lago, C., and Anguera, M. T. (2002). Use of the polar coordinates technique to study interactions among professional soccer players. *Rev. Port. Ciênc. Desporto.* 2, 21–40. doi: 10.5628/rpcd.02.04.21

López, J., Valero, A., Anguera, M. T., and Díaz, A. (2016). Disruptive behavior among elementary students in physical education. *Springerplus* 5, 1154. doi: 10.1186/s40064-016-2764-6

López-López, J. A., Menescardi, C., Estevan, I., Falcó, C., and Hernández-Mendo, A. (2015). Análisis técnico-táctico en Taekwondo con coordenadas polares a través del software HOISAN [Technical-tactical analysis in taekwondo with polar coordinates through software HOISAN]. *Cuad. Psico. Dep.* 15, 131–142. doi: 10.4321/S1578-84232015000100013

McGarry, T., Anderson, D. I., Wallace, S., Hughes, M. D., and Franks, I. (2002). Sport competition as a dynamical self-organizing system. *J. Sports Sci.* 20, 771–781. doi: 10.1080/026404102320675620

Memmert, D. (2010). Testing of tactical performance in youth elite soccer. *J. Sports Sci. Med.* 9, 199–205.

Memmert, D., Lemmink, K. A., and Sampaio, J. (2017). Current approaches to tactical performance analyses in soccer using position data. *Sports Med.* 47, 1–10. doi: 10.1007/s40279-016-0562-5

Morillo-Baro, J. P., Reigal, R. E., and Hernández-Mendo, A. (2015). Analysis of positional attack in beach handball male and female with polar coordinates. *Rev. Int. Cienc. Deporte* 41, 226–244. doi: 10.5232/ricyde2015.04103

Perea, A., Castellano, J., Alday, L., and Hernández-Mendo, A. (2012). Analysis of behaviour in sports through Polar Coordinate Analysis with MATLAB®. *Qual. Quant.* 46, 1249–1260. doi: 10.1007/s11135-011-9435-z

Prudente, J., Sousa, D., Sequeira, P., López-López, J. A., and Hernández-Mendo, A. (2018). Analyzing the influence of playing time and partial score on the tactical behavior in the duel 2 vs 2 in the offensive process in handball, using the polar coordinates technique. *Anal. Psychol.* 33, 515–529. doi: 10.6018/analesps.33.3.271071

Rampinini, E., Impellizzeri, F. M., Castagna, C., Abt, G., Chamari, K., Sassi, A., et al. (2007). Factors influencing physiological responses to small-sided soccer games. *J. Sports Sci.* 25, 659–666. doi: 10.1080/02640410600811858

Ric, A., Torrents, C., Gonçalves, B., Torres-Ronda, L., Sampaio, J., and Hristovski, R. (2017). Dynamics of tactical behaviour in association football when manipulating players' space of interaction. *PLoS ONE* 12:e0180773. doi: 10.1371/journal.pone.0180773

Robles, F., Castellano, J., Perea, A., Martinez-Santos, R., and Casamichana, D. (2013). "Spatial strategy used by the world champion in South Africa 2010," in *Science and Football VII: The Proceedings of the Seventh World Congress on Science and Football* (London: Routledge), 161–167.

Robles, F. J., Castellano, J., and Perea, A. E. (2014). Differences of the game between the football spanish team and its rivals. *Rev. Ib. CC. Act. Fis. Dep.* 3, 1–8.

Sackett, G. P. (1980). "Lag sequential analysis as a data reduction technique in social interaction research," in *Exceptional Infant: Psychosocial Risks in Infant-Environment Transactions, Vol. 4,* eds D. B. Sawin, R. C. Hawkins, L. O. Walker, and J. H. Penticuff (New York, NY: Brunner-Mazel), 300–340.

Sampaio, J., and Maças, V. (2012). Measuring tactical behaviour in football. *Int. J. Sports Med.* 33, 395–401. doi: 10.1055/s-0031-1301320

Sánchez-Algarra, P., and Anguera, M. T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: impact of recording and coding predominating perspectives. *Qual. Quant.* 47, 1237–1257. doi: 10.1007/s11135-012-9764-6

Sousa, D. J., Prudente, J. N., Sequeira, P., López-López, J. A., and Hernández-Mendo, A. (2015). Análisis de las situaciones de juego 2vs2 en el campeonato europeo masculino de balonmano 2012: Aplicación de la técnica de coordenadas polares [Analysis of 2vs2 situations in Men's European Handball Championship 2012: application of polar coordinates technique]. *Cuad. Psico. Dep.* 15, 181–194.

Sumpter, D. (2016). *Soccermatics. Fussball und die Magie der Zahlen.* Uppsala: Ecowin.

Svensson, M., and Drust, B. (2005). Testing soccer players. *J. Sports Sci.* 23, 601–618. doi: 10.1080/02640410400021294

Tarragó, R., Iglesias, X., Lapresa, D., Anguera, M. T., Ruiz-Sanchis, L., and Arana, J. (2017). Analysis of diachronic relationships in successful and unsuccessful behaviors by world fencing champions using three complementary techniques. *Anal. Psychol.* 33, 471–485. doi: 10.6018/analesps.33.3.271041

Taylor, J. B., Mellalieu, S. D., and James, N. (2005). A comparison of individual and unit tactical behaviour and team strategy in professional soccer. *Int. J. Peform. Anal. Sport* 5, 87–101. doi: 10.1080/24748668.2005.11868329

Thelwell, R. C., Greenlees, I. A., and Weston, N. J. (2006). Using psychological skills training to develop soccer performance. *J. Appl. Sport Psychol.* 18, 254–270. doi: 10.1080/10413200600830323

Wiemeyer, J. (2003). Who should play in which position in soccer? Empirical evidence and unconventional modelling. *Int. J. Peform. Anal. Sport* 3, 1–18. doi: 10.1080/24748668.2003.11868269

Yin, R. K. (2014). *Case Study Research: Design and Methods, 5th Edn.* Los Angeles, CA: Sage.

Check for updates

Decision-Making by Handball Referees: Design of an *ad hoc* Observation Instrument and Polar Coordinate Analysis

Juan P. Morillo, Rafael E. Reigal, Antonio Hernández-Mendo*, Alejandro Montaña and Verónica Morales-Sánchez

Departamento de Psicología Social, Trabajo Social, Antropología Social y Estudios de Asia Oriental, Facultad de Psicología, Universidad de Málaga, Málaga, Spain

Referees are essential for sports such as handball. However, there are few tools available to analyze the activity of handball referees. The aim of this study was to design an instrument for observing the behavior of referees in handball competitions and to analyze the resulting data by polar coordinate analysis. The instrument contained 6 criteria and 18 categories and can be used to monitor and describe the actions of handball referees according to their role/position on the playing court. For the data quality control analysis, we calculated Pearson's (0.99), Spearman's (0.99), and Tau Kendall's (1.00) correlation coefficients and Cohen's kappa (entre 0.72 y 0.75) and Phi (entre 0.83 y 0.87) coefficients. In the generalizability analysis, the absolute and relative generalizability coefficients were 0.99 in both cases. Polar coordinate analysis of referee decisions showed that correct calls were more common for central court and 7-meter throw calls. Likewise, calls were more likely to be incorrect (in terms of both errors of omission and commission) when taken from the goal-line position.

Keywords: refereeing, handball, polar coordinates, decision-making, systematic observation

INTRODUCTION

Referees have a key role in elite sports competitions (Cruz, 1997; Dohmen and Sauermann, 2015). Officiating a match between two teams is a difficult task that is further complicated by the need to take decisions on a range of events that occur within a short space of time (Plessner, 2005; Mascarenhas and Smith, 2011). The decisions taken by referees can influence the unfolding of events during a match and even decide the outcome (Philippe et al., 2009). It is therefore important to analyze and improve referee performance in these contexts.

Good referees share certain qualities. They must be very knowledgeable about the rules of the game, have a good level of physical fitness, position themselves correctly on the court or pitch, have good visual and auditory acuity, and be highly motivated and capable of taking on-the-spot decisions and controlling their emotions (Weinberg and Richardson, 1990; Mascarenhas et al., 2005; Simmons, 2011). These qualities, can, however, be modified by various factors that can affect decision-making processes (Weston et al., 2012), such as previous experiences with teams and/or players and even player reputation and gender.

Refereeing in team handball is a complex task, as handball is a fast, physical game involving continuous contact and offensive and defensive actions (Souchon et al., 2009). To meet the

demands of officiating a match and withstand the pressure generated by players, crowds, and critical moments, referees need to be sufficiently prepared, both psychologically and technically (Gimeno et al., 1998; Debanne, 2014). Insufficient preparation can lead to attention and concentration difficulties, doubts regarding decisions, increased anxiety levels, and a greater risk of making mistakes (Estrada and Pérez, 2008; Debanne, 2014).

Decision-making by handball players has received considerable attention in recent years, and numerous tools have been created to analyze what occurs in game situations (Luckwü and Guzmán, 2011; Martín et al., 2013; Loffing et al., 2015; Weigel et al., 2015; Helm et al., 2016). Tools have also emerged to analyze the activity of coaches, who have an important influence on match tactics and outcomes (Debanne and Fontayne, 2009; Debanne, 2014). There is, however, a need for reliable, accurate tools for analyzing the performance of handball referees, as very few have been developed (Souchon et al., 2009).

Research in this area has sought to identify different elements that can help to interpret decision-making in sport (Araujo et al., 2016). Systematic observation, for instance, offers a range of techniques for analyzing behavior in natural settings (González et al., 2013; Lapresa et al., 2013; Anguera and Hernández-Mendo, 2014; Sousa et al., 2014). Numerous studies have shown that observational methodology is an adequate methodology for analyzing behavior in sport (Anguera and Hernández-Mendo, 2013). It is (a) non-intrusive, (b) has a high level of ecological validity (i.e., it analyzes natural behaviors in natural settings), and (c) offers high analytical specificity through the construction of *ad hoc* observation instruments designed specifically for analyzing specific game situations in the environment in which they occur (Araujo, 2011, 2013; Pinder et al., 2011).

One technique that has shown great potential in this area in recent years is polar coordinate analysis (Sackett, 1980). It is among the most informative techniques (Araujo et al., 2016) and is particularly powerful when the concept of genuine retrospectivity is applied (Anguera, 1997). Recent years have seen a rapid uptake of polar coordinate analysis in the field of Sports Sciences, where it has been used to analyze a range of sports, including soccer, tennis, and handball (Castellano et al., 2007; Perea et al., 2012; Morillo and Hernández-Mendo, 2015; Morillo et al., 2015; Castañer et al., 2016; López et al., 2016; Santoyo et al., 2017; Tarragó et al., 2017).

To analyze decisions taken in sport, it is necessary to analyze the different actions that occur during a game (Pinder et al., 2011). Polar coordinate analysis is a suitable technique for identifying and helping to understand these actions. Prudente et al. (2017), for example, used this technique to show how playing time influenced tactical decisions made by handball players. Polar coordinate analysis has also been used in beach volleyball to identify erroneous behaviors in relation to passes and receptions (Morillo et al., 2015). Finally, the technique has been successfully applied to analyzing tactical decisions taken in track events.

Polar coordinate analysis is a powerful technique that reduces the volume of data to be processed without losing important information. It is used to identify significant relationships between a behavior of interest, known as the focal behavior, and other behaviors, known as conditional behaviors, and presents these in an easy-to-interpret vector format (Hernández-Mendo and Anguera, 1998; Anguera and Losada, 1999; Gorospe and Anguera, 2000). The technique involves using adjusted residuals derived from sequential analysis (z scores) to calculate Z_{sum} statistics ($Z_{sum} = \sum z/\sqrt{n}$) (Cochran, 1954). This computation is possible, as both the frequency of the focal behavior (n) and the Z scores for each of the lags considered are known. These Z scores are independent of each other, as they are computed using the binomial test, which compares observed probabilities (corresponding to textual units derived from observation of the teachers' discourse) with expected probabilities (chance occurrences). The relationship between the focal behavior and the conditional behaviors is estimated using the angle of the resulting vector, while the strength is estimated using the vector radius (Anguera et al., 1997; Castellano and Hernández-Mendo, 2003). A crucial component of polar coordinate analysis is that its powerful data reduction feature permits the consideration of both retrospective and prospective perspectives. In other words, it shows what happens before and after the behavior of interest.

Given the scarcity of tools available for analyzing the activity of handball referees, the main aim of this study was to design a tool that could be used to objectively analyze referee behavior and performance in competition situations. A second aim was to test the tool using data from three matches at the 2013 World Men's Handball Championship held in Spain.

MATERIALS AND METHODS

As we used an observation instrument that combined field formats with category systems, the observational design was multidimensional (Morillo et al., 2015; Prudente et al., 2017). The specific design was follow-up/idiographic/multidimensional, which fits into quadrant I of the systematic observation designs described by Anguera et al. (2011).

Participants

A group of six observers, all male referees who officiated regional handball matches in Andalusia, Spain, participated in the data quality control phase. They were aged between 22 and 26 years (mean = 23.50; SD = 1.26) and had between 5 and 8 years' refereeing experience. For this analysis, the observers studied the semi-final between Spain and Slovenia at the 2013 World Men's Handball Championship.

The polar coordinate analysis was performed using data coded by a single observer from three final-stage matches in the same championship: the semi-final between Spain and Slovenia, the semi-final between Denmark and Croatia, and the third place game between Slovenia and Croatia.

The ethical requirements of observational methodology were applied to the current study and performed in accordance with the ethical standards laid down in the Declaration of Helsinki.

Instruments

The observation instrument used to analyze and code the referees' actions in the matches analyzed was designed within the framework of observational methodology (Anguera, 1990, 1991; Anguera and Hernández-Mendo, 2014). Given the scarcity of existing theoretical constructs and the multidimensional nature of handball, the coding system (observation instrument) was built using an empirical-inductive approach (Castellano et al., 2000; Morillo et al., 2015).

The instrument comprised a combination of a field format system for each criterion (Anguera, 1979; Anguera and Hernández-Mendo, 2013) and a system of exhaustive, mutually exclusive categories. The final instrument contained 6 criteria and 18 categories (**Table 1**).

For the polar coordinate analysis, we chose three focal behaviors that would permit analysis of individual interventions by referees, as this is a key aspect of refereeing. The categories chosen were related to decision (responsibility). Although the other categories are also important, we chose the three categories that could provide the most useful information for the aim of the study. These were:

MDO_ACI: Correct call by referee responsible for making the call,
MDO_ERO: Failure to make a call by referee responsible for the call,
MDO_ERC: Call made by referee not responsible for making it.

The data were coded and analyzed using HOISAN (Hernández-Mendo et al., 2012, 2014), a software program that performs polar coordinate analysis and presents the output in the form of vector maps. The generalizability analysis was performed in SAGT (Hernández-Mendo et al., 2016).

Procedure

A generalizability analysis was used to test the validity and accuracy of the *ad hoc* observation instrument (Blanco-Villaseñor et al., 2000, 2014; Castellano et al., 2000). Generalizability coefficients provide an estimate of how the observed mean compares with the mean of all possible observations (Blanco-Villaseñor et al., 2000, 2014). Inter-observer agreement was assessed to estimate reliability (Anguera, 1990; Morillo and Hernández-Mendo, 2015).

For the data quality control analysis, three different moments of the semi-final between Spain and Slovenia were analyzed by previously trained observers. Two of the moments were observed by the same team of observers and the third one was observed by a second team. To maximize inter-observer agreement, the observers were trained (Morillo and Hernández-Mendo, 2015) and provided with a purpose-designed observation protocol. In addition, the data were coded using the consensus agreement method described by Anguera (1990). Cohen's kappa coefficients, generalizability analysis, and correlation coefficients were used to measure intra- and inter-observer agreement; the results in all cases were higher than 0.90. In the subsequent full data collection phase, 328 behaviors were coded in the three matches analyzed.

Handball matches are officiated by two referees with the same level of responsibility. In each match, the actions of the referees were coded simultaneously by three previously trained observers using the consensus agreement method. The observers were all regional-level handball referees. There are two referees in handball, a court referee and a goal-line referee, and these generally position themselves opportunely to cover critical areas of the playing court at any given time.

Polar coordinate analysis, through the calculation of Z_{sum} statistics derived from adjusted residuals corresponding to prospective and retrospective lags, indicates the nature of the relationship between a focal and a conditional behavior, which can be excitatory or inhibitory. The type of relationship is determined by the quadrant in which the corresponding vector is located, and the focal behavior will always be excitatory or inhibitory. The meaning of the four quadrants is shown below:

(1) Quadrant I: Mutual excitation between focal and conditional behavior (i.e., prospective and retrospective activation),
(2) Quadrant II: Inhibitory focal behavior and excitatory conditional behavior (i.e., prospective inhibition and retrospective activation),
(3) Quadrant III: Mutual inhibition between focal and conditional behavior (i.e., prospective and retrospective inhibition),
(4) Quadrant IV: Excitatory focal behavior and inhibitory conditional behavior (i.e., prospective activation and retrospective inhibition).

The following events were excluded from the analysis and were therefore not recorded as correct calls: goals, throw-offs (recorded as an error if incorrectly executed), whistle for a free throw, throw-in, or goalkeeper throw. As one of the criterion was a whistle signal by a referee, application of the advantage rule was not recorded as a correct call.

RESULTS

Data Quality

The correlation coefficients in **Table 2** show that the *ad hoc* observation instrument allowed for the reliable and accurate recording of data.

Generalizability Analysis

Generalizability analysis is used to estimate accuracy, validity, reliability, and sample size (Blanco-Villaseñor et al., 2014). The analysis consists of analyzing potential sources of variation that might be affecting an observational measurement or measurement design and estimating the generalizability of the design with respect to the particular conditions of a theoretical value (Blanco-Villaseñor et al., 2014).

The results for the measurement design [Criteria] [Categories]/[Observers] are shown in **Tables 3**, **4**. The largest source of variation was associated with the interaction [Criteria] [Categories].

The results of the generalizability analysis show optimal values for absolute and relative generalizability coefficient values, in

TABLE 1 | Observation instrument: Criteria and corresponding categories and codes.

Criterion	Category
Position of referee	POS_CEN: center of court POS_FON: goal line
Responsibility: Whether or not the call that was made or should have been made was the responsibility of the referee being observed	PER_SI: Yes, the call was his responsibility PER_SI: No, the call was not his responsibility
Whistle: Whether or not the whistle was blown	PIT_SI: Yes, the whistle was blown PIT_NO: No, the whistle was not blown
Decision (responsibility)	MDO_ACI: The call was correct and was the responsibility of the referee being observed MDO_ERO: Error of omission; the call was the responsibility of the referee being observed but was not made MDO_ERC: Error of commission; the call was not the responsibility of the referee being observed but was made
Type of infraction	TIP_TECGF: Technical, free throw TIP_TEC7: Technical, 7-meter throw foul TIP_DIS: Warning, progressive punishment, direct dismissal, disqualification or disqualification with written report TIP_T-D: Technical foul-punishment; free throw or 7-meter throw fouls that include a personal punishment TIP_T-T: Technical-tactical; loss of possession without punishment TIP_TTD: Technical-tactical with punishment; loss of possession with punishment
Decision (action)	MDA_ACI: Correct call MDA_ERR: Incorrect call

TABLE 2 | Intra and inter-observer agreement.

Coefficient for entire session	Intra-observer agreement (Obs. 1 vs. Obs. 1bis)	Inter-observer agreement (Obs. 1 vs. Obs. 2)
Pearson's	0.9981	0.9982
Spearman's (p)	0.9987	0.9975
Kendall's tau-b	1	1
Kappa	0.7222	0.7573
Phi	0.8345	0.8782

TABLE 3 | Sources of variation, sum of squares, degrees of freedom, mean squares, %, and standard error.

Sources of variation	Sum of squares	GC	Mean squares	%	Standard error
Observers	4.675	2	2.337	0.006	0.020
Criteria	64.032	5	12.806	0.000	6.280
[Observers][Criteria]	1.516	10	0.152	0.000	0.010
Categories	2,5796.778	13	1,984.368	4.764	42.837
[Observers][Categories]	16.437	26	0.632	0.000	0.035
[Criteria][Categories]	99,194.079	65	1,526.063	95.048	87.888
[Observers][Criteria][Categories]	126.706	130	0.975	0.182	0.120

GC, Generalizability Coefficient.

addition to a linear tendency for the SDs of each design. In all cases, the relative SD was lower than the absolute SD.

Polar Coordinate Analysis

The vector maps for the three focal behaviors selected for the polar coordinate analysis are shown below. The following results

TABLE 4 | Absolute generalizability coefficient, relative generalizability coefficient, absolute SD, and relative SD in relation to measurement design.

Measurement design	Absolute generalizability coefficient	Relative generalizability coefficient	Absolute SD	Relative SD
[Criteria][Categories][Observers]	0.999	0.999	0.579	0.570
[Observers][Criteria][Categories]	0.001	0.001	6.181	6.032

were obtained for MDO-ACI (correct call by right referee) (**Table 5**, **Figure 1**).

In quadrant I, the following conditional behaviors were significantly associated (>1.96) with a correct call made by the right referee (MDO_ACI): the call was made by the court referee (POS_CEN), the referee was responsible for the call (PER_SI), and the whistle was blown (PIT_SI). Relationships in quadrant I are mutually excitatory, i.e., the focal and conditional behaviors activate each other.

There were no significant relationships in quadrant II.

In quadrant III, the following behaviors were significantly associated with MDO_ACI: the call was made by the goal-line referee (POS_FON), the call was not the responsibility of the referee (PER_NO), the whistle was not blown (PIT_NO), error of omission (MDO_ERO), error of commission (MDO_ERC), and incorrect call (MDA_ERR). As expected, the focal behavior inhibited the other two categories in the same criterion, as they are mutually exclusive. It also inhibited not blowing the whistle, as for MDO-ACI to occur, the referee has to use his whistle.

In quadrant IV, just one behavior was significantly associated with MDO-ACI: TIP_TEC7 (7-meter throw foul). This shows

TABLE 5 | Relationships between focal behavior MDO_ACI and conditional behaviors.

Category	Quadrant	Prospective	Retrospective	Ratio	Radius	Angle
POS_CEN	I	1.27	2.34	0.88	2.66*	61.42
PER_SI	I	1.81	2.38	0.8	2.99*	52.7
PIT_SI	I	2.21	1.27	0.5	2.55*	29.84
MDO_ACI	I	2.46	2.46	0.71	3.48*	45
TIP_TECGF	I	1.09	1.11	0.71	1.56	45.56
TIP_T-T	I	0.26	0.03	0.1	0.26	5.75
TIP_TTD	I	0.44	0.45	0.71	0.63	45.6
MDA_ACI	I	2.05	2.66	0.79	3.36*	52.35
TIP_DIS	II	−0.26	0.49	0.88	0.56	118.08
TIP_T-D	II	−0.06	0.42	0.99	0.42	98.19
POS_FON	III	−1.27	−2.34	−0.88	2.66*	241.42
PER_NO	III	−1.81	−2.38	−0.8	2.99*	232.7
PIT_NO	III	−2.21	−1.27	−0.5	2.55*	209.84
MDO_ERO	III	−2.21	−1.27	−0.5	2.55*	209.84
MDO_ERC	III	−0.93	−2.63	−0.94	2.79*	250.44
MDA_ERR	III	−2.05	−2.66	−0.79	3.36*	232.35
TIP_TEC7	IV	0.63	−2.2	−0.96	2.29*	285.94

*$p < 0.05$.

FIGURE 1 | Vector map for focal behavior MDO_ACI.

that the likelihood of this foul being called correctly by the right referee is very high.

The following results were obtained for the error of omission category MDO-ERO, which is when a referee should have made a call but did not (**Table 6**, **Figure 2**).

The following conditional behaviors were associated with MDO_ERO (error of omission) in quadrant I: the call was made by the goal-line referee (POS_FON), the call was not the responsibility of the referee (PER_NO), error of commission (MDO_ERC), and incorrect call (MDA_ERR). As they are

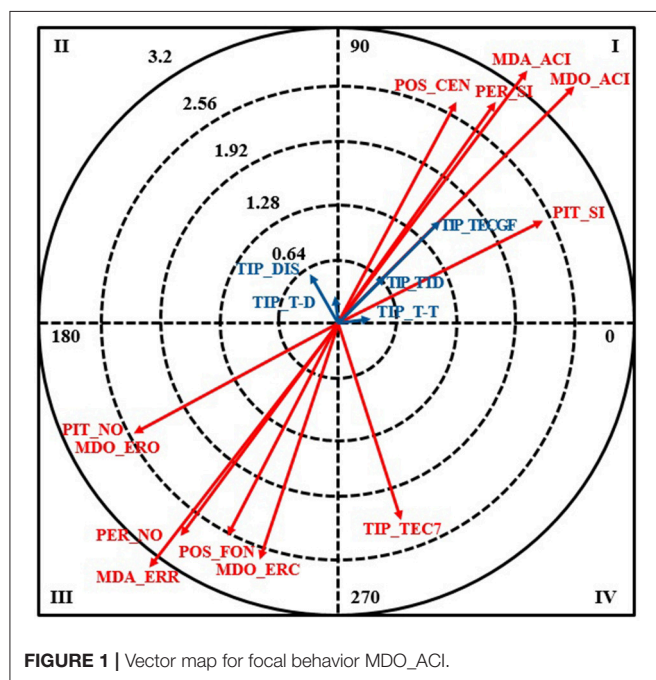located in quadrant I, the focal and conditional behaviors are mutually excitatory.

No significant relationships were detected in quadrant II or IV.

In quadrant III, the following behaviors were significantly associated with MDO_ERO: the call was made by the court referee (POS_CEN), the call was not the responsibility of the referee (PER_SI), correct call made by the right referee (MDO_ACI), and correct call (MDA_ACI). As expected for quadrant III, the focal behavior (incorrect call) inhibited correct calls.

The following results were obtained for the error of commission category MDO-ERC, which is when a referee made a call that was not his responsibility (**Table 7**, **Figure 3**).

The following conditional behaviors were all significantly associated with the focal behavior in quadrant I: goal-line position (POS_FON), whistle not blown (PIT_NO), and error of omission (MDO_ERO).

No significant relationships were detected in quadrant II or IV.

MDO_ERO was significantly associated with several behaviors in quadrant III: central-court position (POS_CEN), whistle blown (PIT_SI), correct call made by right referee (MDO_ACI), and correct call (MDA_ACI). Again, the focal behavior inhibited behaviors related to correct calls.

DISCUSSION

We have presented a new tool for observing, coding, and analyzing the actions of referees in handball competitions. Although decisions made by referees are influenced by contextual factors (Debanne, 2014), the *ad hoc* observation instrument described in this study was designed to provide an objective means of recording, describing, and analyzing actions taken by handball referees according to their role and position on the court. While observational methodology has been used to analyze handball, studies to date have focused on game situations from the players' perspective (González et al., 2013; Sousa et al., 2014).

The reliability, generalizability, and correlation results in the data quality control analysis attest to the suitability of the data obtained. The observation instrument thus would appear to be an adequate tool for obtaining reliable datasets for performing sequential and other analyses of the performance of court and goal-line referees during handball competitions. In this respect, it is similar to observation instruments designed for other sports, such as soccer (Sarmento et al., 2010), basketball (Garzón et al., 2011), waterpolo (Santos et al., 2014), and beach handball (Morillo and Hernández-Mendo, 2015). A recent study by Araujo et al. (2016) addressed the issue of decision-making in sport and argued that the use of observation to analyze specific actions and behaviors could provide complementary insights into this complex process. The conceptual vector maps presented in this study show how the referees responded to events based on their use of the whistle. The instrument presented has numerous applications. It could be used, for example, to identify streams of behavior or specific actions that cause greater difficulties for

TABLE 6 | Relationships between focal behavior MDO_ERO and conditional behaviors.

Category	Quadrant	Prospective	Retrospective	Ratio	Radius	Angle
POS_FON	I	0.47	1.92	0.97	1.97*	76.29
PER_NO	I	1.86	3.23	0.87	3.73*	60.02
PIT_NO	I	1.01	1.01	0.71	1.43	45
MDO_ERO	I	1.01	1.01	0.71	1.43	45
MDO_ERC	I	0.74	2.42	0.96	2.53*	72.92
MDA_ERR	I	1.38	1.78	0.79	2.26*	52.22
TIP_TEC7	II	−0.38	1.43	0.97	1.48	104.94
POS_CEN	III	−0.47	−1.92	−0.97	1.97*	256.29
PER_SI	III	−1.86	−3.23	−0.87	3.73*	240.02
PIT_SI	III	−1.01	−1.01	−0.71	1.43	225
MDO_ACI	III	−1.29	−2.14	−0.86	2.5*	238.96
TIP_TECGF	III	−0.78	−0.53	−0.56	0.94	214.35
TIP_DIS	III	−0.68	−1.03	−0.83	1.24	236.6
TIP_TTD	III	−0.19	−0.16	−0.64	0.25	219.83
MDA_ACI	III	−1.38	−1.78	−0.79	2.26*	232.22
TIP_T-D	IV	0.46	−0.27	−0.51	0.53	329.64
TIP_T-T	IV	0.36	−0.12	−0.33	0.38	341.03

*p < 0.05.

TABLE 7 | Relationships between focal behavior MDO_ERO and conditional behaviors.

Category	Quadrant	Prospective	Retrospective	Ratio	Radius	Angle
POS_FON	I	1.5	1.51	0.71	2.13*	45.3
PIT_NO	I	2.42	0.74	0.29	2.53*	17.08
MDO_ERO	I	2.42	0.74	0.29	2.53*	17.08
MDO_ERC	I	0.47	0.47	0.71	0.66	45
TIP_DIS	I	1.74	0.4	0.22	1.78	12.99
MDA_ERR	I	1.39	1.95	0.81	2.4*	54.49
PER_SI	II	−0.2	0.95	0.98	0.97	101.78
TIP_T-T	II	−1.1	0.28	0.25	1.13	165.73
TIP_TEC7	II	−0.62	1.49	0.92	1.61	112.45
POS_CEN	III	−1.5	−1.51	−0.71	2.13*	225.3
PIT_SI	III	−2.42	−0.74	−0.29	2.53*	197.08
MDO_ACI	III	−2.41	−0.91	−0.35	2.57*	200.66
TIP_TECGF	III	−0.61	−1.14	−0.88	1.3	241.88
TIP_T-D	III	−0.79	−0.3	−0.36	0.84	200.91
TIP_TTD	III	−0.58	−0.6	−0.72	0.83	226.04
MDA_ACI	III	−1.39	−1.95	−0.81	2.4*	234.49
PER_NO	IV	0.2	−0.95	−0.98	0.97	281.78

*p < 0.05.
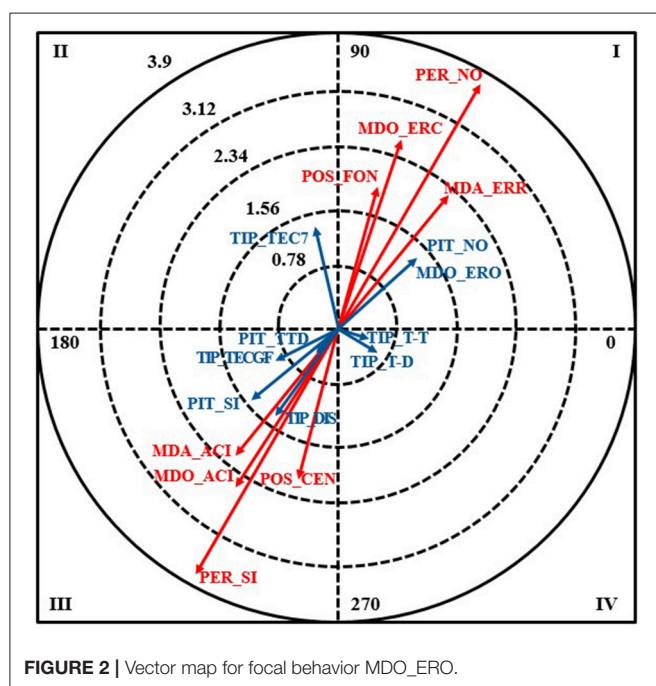
FIGURE 2 | Vector map for focal behavior MDO_ERO.
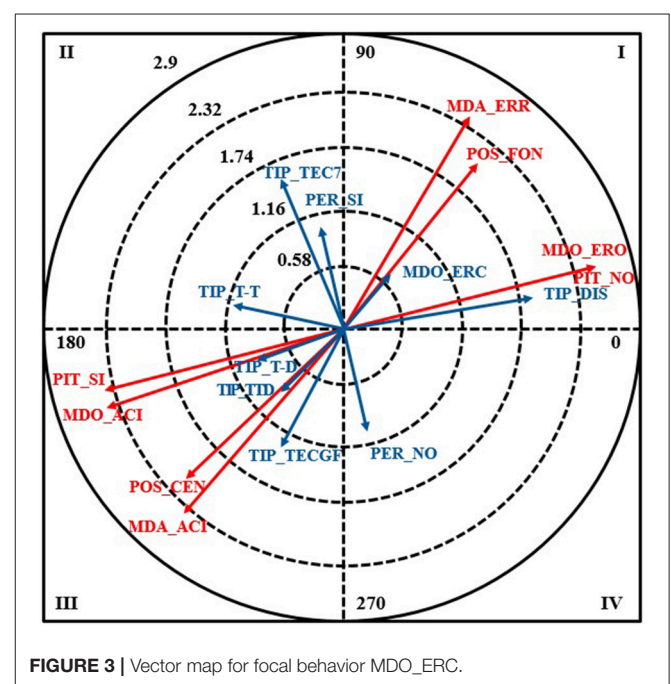
FIGURE 3 | Vector map for focal behavior MDO_ERC.

referees or situations that are prone to more error, regardless of level of physical fitness. Handball refereeing has been reported to require moderate levels of fitness and does not appear to be limited by aerobic capacity (Fernandes da Silva et al., 2010).

Most of the correct calls were made from the central court position. This is logical, as court referees are generally responsible for making more calls than goal-line referees and have to deal with less conflictive situations. Goal-line referees, by contrast, have to deal with multiple interactions in short spaces of time and are therefore more likely to make incorrect calls, even

though they use their whistle less. We also found that the referees observed made a high percentage of correct calls. Seven-meter throw fouls, for example, were correctly called by the right referee (the goal-line referee) in all cases. This again is logical, as fouls of this type are generally the responsibility of the referee at the end of the court and are rarely called by the court referee.

Handball, unlike other sports such as basketball, does not use instant-replay or similar technology to facilitate the work of referees. The installation of court-side cameras to watch

instant or near-instant replays of dubious play or the use of goal-line sensors to check whether or not the ball completely crossed the goal-line could lead to interesting improvements in the game. Such measures, however, also have drawbacks. The technology is costly and perhaps should only be considered for elite competitions. In addition, the use of these systems could hurt the credibility of referees and cause them to lose confidence in their calls, particularly in the case of less experienced referees. Novice referees have been found to perform less well than "expert referees" with greater knowledge, experience, and expert memory (Abdeddaim et al., 2016).

More studies of decision-making by handball referees are needed to assess the possible advantages of redistributing responsibilities and zones between both referees and even perhaps of using a third referee in areas with high error rates. Our study highlights some limitations that could be overcome in future studies. It would be interesting, for example, to analyze more areas of the court and to divide the court into specific zones to analyze the actions of referees according to the number of players on the court at a given time and the position of the defense. The distribution of responsibilities is more complicated in open and man-to-man defences, as it is less clear in such cases who is responsible for calling what. It is also complicated to determine whether a referee chose not to make a call or decided to apply the advantage rule, as there are no official hand signals for this decision. The rules do, however, specify that referees should refrain from interrupting the game prematurely to allow continuity of play where possible. Accordingly, there may be some overlap between application of the advantage rule and errors of omission.

Although some research has already been done on how player gender can influence decision-making by referees in handball (Souchon et al., 2010), more work in this area is necessary. Finally, it would be interesting to analyze different championships over time to monitor the influence of new rules and regulations and changes in refereeing practice and performance.

AUTHOR CONTRIBUTIONS

AH, VM and RR: design of the work; acquisition, analysis, and interpretation of data for the work. JM: acquisition, analysis, and interpretation of data for the work; AM: acquisition and analysis of data for the work. All authors: Drafting the work or revising, final approval of the version and agreement to be accountable for all aspects of the work.

ACKNOWLEDGMENTS

REFERENCES

Abdeddaim, A., Sebbane, M., Zitouni, A., Zubiaur, M., and Boumesdji, A. (2016). Efecto de la base de conocimientos y de la memoria experta en una tarea de detección de errores de arbitraje en balonmano [Effect of the base of knowledge and of the memory expert in error detection task of refereeing in handball]. *Rev. Iberoam. Psicol. Ejerc. Deporte* 11, 239–245.

Anguera, M. T. (1979). "Observación de la conducta especial" [Observation of spatial behavior]," in *VI Congreso Nacional de Psicología* (Pamplona).

Anguera, M. T. (1990). "Metodología observacional," in *Metodología de la Investigación en Ciencias del Comportamiento [Research Methodology in Behavioral Sciences],* eds J. Arnau, M. T. Anguera, and J. Gómez (Murcia: Secretariado de Publicaciones de la Universidad de Murcia), 125–236.

Anguera, M. T. (1991). "Proceso de categorización" [Categorization process], in *Metodología Observacional en la Investigación Psicológica,* Vol. 1, ed M. T. Anguera (Barcelona: P.P.U), 115–167.

Anguera, M. T. (1997). "From prospective patterns in behavior to joint analysis with a retrospective perspective," in *Collaque Sur Invitation ≪Méthologie d'Analyse des Interactions Sociales≫* (Paris: Université de la Sorbona).

Anguera, M. T., Blanco, A., and Losada, J. L. (1997). "Aportaciones de la técnica de coordenadas polares en diseños mixtos [Contributions of the technique of polar coordinates in mixed designs]," in *IV Simposio de Metodología de las Ciencias del Comportamiento* (Murcia: Universidad de Murcia), 583.

Anguera, M. T., Blanco-Villaseñor, A., Hernández-Mendo, A., and Losada, J. L. (2011). Diseños observacionales: ajuste y aplicación en psicología del deporte [Observational designs: their suitability and application in sports psychology]. *Cuad. Psicol. Deporte* 11, 63–76.

Anguera, M. T., and Hernández-Mendo, A. (2013). La metodología observacional en el ámbito del deporte [Observational methodology in sport sciences]. E-Balonmano.com: *Rev. Cien. Deporte* 9, 135–160.

Anguera, M. T., and Hernández-Mendo, A. (2014). Metodología observacional y psicología del deporte: estado de la cuestión [Observational methodology and sports psychology: state of affairs]. *Rev. Psicol. Deporte* 23, 103–119.

Anguera, M. T., and Losada, J. L. (1999). "Reducción de datos en marcos de conducta mediante la técnica de coordenadas polares," [Reduction of data in behavioral frameworks using the polar coordinate technique] in *Observación de la Conducta Interactiva en Situaciones Naturales: Aplicaciones*, ed M. T. Anguera (Barcelona: E.U.B), 165–188.

Araujo, D. (2011). De la toma de decisiones, al curso de las decisiones [From decision-making, to the course of decisions]. *Rev. Psicol. Deporte* 20, 639–643.

Araujo, D. (2013). The study of decision-making behavior in sport. *Rev. Int. Cien. Deporte* 9, 1–4. doi: 10.5232/ricyde2013.031

Araujo, D., Teques, P., Hernández-Mendo, A., Reigal, R., and Anguera, M. T. (2016). La toma de decisión, ¿es una conducta observable?: discusión sobre diferentes perspectivas teóricas utilizadas en el estudio del rendimiento deportivo [Decision-making, is it an observable behaviour?: discussion of contrasting theoretical perspectives in use in the study of sport performance]. *Cuad. Psicol. Deporte* 16, 183–196.

Blanco-Villaseñor, A., Castellano, J., and Hernández-Mendo, A. (2000). Generalizabilidad de las observaciones en la acción del juego en el fútbol [Observations of the football game action: a generalizability study] *Psicothema* 12, 81–86.

Blanco-Villaseñor, A., Castellano, J., Hernández-Mendo, A., Sánchez-López, C. R., and Usabiaga, O. (2014). Aplicación de la TG en el deporte para el estudio de la fiabilidad, validez y estimación de la muestra [Application of the generalizability theory in sport to study the validity, reliability and estimation of samples]. *Rev. Psicol. Deporte* 23, 131–137.

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Canton, A., and Hileno, R. (2016). Goal scoring in soccer: a polar coordinate analysis of motor skills used by Lionel Messi. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Castellano, J., and Hernández-Mendo, A. (2003). El análisis de coordenadas polares para la estimación de relaciones en la interacción motriz en fútbol

[Polar coordinates analysis to estimate the relationships in the motor interaction in soccer]. *Psicothema* 15, 569–579.

Castellano, J., Hernández-Mendo, A., Gómez de Segura, P., Fontetxa, E., and Bueno, I. (2000). Sistema de codificación y análisis de la calidad del dato en el fútbol de rendimiento [Codification system and data quality analysis in top level soccer]. *Psicothema* 12, 636–641.

Castellano, J., Hernández-Mendo, A., Morales-Sánchez, V., and Anguera, M. T. (2007). Optimizing a probabilistic model of the development of play in soccer. *Qual. Quant.* 41, 93–104. doi: 10.1007/s11135-005-3148-0

Cochran, W. G. (1954). Some methods for streghning the common χ² test. *Biometrics* 10, 417–451 doi: 10.2307/3001616

Cruz, J. (1997). "Asesoramiento psicológico en el arbitraje y juicios deportivos" [Psychological counseling in arbitration and sport trials]. in *Psicología del Deporte* ed J. Cruz (Madrid: Síntesis), 245–269.

Debanne, T. (2014). Techniques used by coaches to influence referees in professional team handball. *Int. J. Sports Sci. Coach* 9, 433–446. doi: 10.1260/1747-9541.9.3.433

Debanne, T., and Fontayne, P. (2009). A study of a successful experienced elite handball coach's cognitive processes in competition situations. *Int. J. Sports Sci. Coach* 4, 1–16. doi: 10.1260/1747-9541.4.1.1

Dohmen, T., and Sauermann, J. (2015). *Referee Bias* (No. 8857). Bonn: Institute for the Study of Labor (IZA).

Estrada, O., and Pérez, E. (2008). Palabras e imágenes positivas en la respuesta de ansiedad en deportistas de competición [Positive words and images on the anxiety response in competition sports]. *Cuad. Psicol. Deporte* 8, 31–45.

Fernandes da Silva, J., Castagna, C., Carminatti, L. J., Foza, V., Guglielmo, L., and de Oliveira, F. R. (2010). Physiological demands of team-handball referees during games. *J. Strength Cond. Res.* 24, 1960–1962. doi: 10.1519/JSC.0b013e3181ddb019

Garzón, B., Lapresa, D., Anguera, M. T., and Arana, J. (2011). Análisis observacional del lanzamiento de tiro libre en jugadores de baloncesto base [Observational analysis of the free throw shot made by grassroots basketball players]. *Psicothema* 23, 851–857.

Gimeno, F., Buceta, J. M., Lahoz, D., and Sanz, G. (1998). Evaluación del proceso de toma de decisiones en el contexto del arbitraje deportivo: Propiedades psicométricas de la adaptación española del cuestionario DMQ II en árbitros de Balonmano [Evaluation of the decision-making process in the context of sports refereeing: Psychometric properties of the Spanish adaptation of the DMQ II questionnaire in handball referees]. *Rev. Psicol. Deporte* 7, 249–258.

González, A., Botejara, J., Puñales, L., Trejo, A., and Ruy, E. (2013). Análisis de la finalización del ataque en partidos igualados en balonmano de alto nivel mediante coordenadas polares [Analysis of the finalization of the attack in equalized games of high level handball using polar coordinates]. *E-balonmano.com: Rev. Cien. Deporte* 9, 71–89.

Gorospe, G., and Anguera, M. T. (2000). Modificación de la técnica clásica de coordenadas polares mediante un desarrollo distinto de la retrospectividad: aplicación al tenis [Retrospectivity in polar coordinates analysis: application to tennis]. *Psicothema* 12, 279–282.

Helm, F., Reiser, M., and Munzert, J. (2016). Domain-specific and unspecific reaction times in experienced team handball goalkeepers and novices. *Front. Psychol.* 7:882. doi: 10.3389/fpsyg.2016.00882

Hernández-Mendo, A., and Anguera, M. T. (1998). "Análisis de coordenadas polares en el estudio de las diferencias individuales de la acción de juego" [Polar coordinate analysis in the study of individual differences of the game action] in *Perspectivas Actuales en la Investigación de las Diferencias Individuales* eds M. P. Sánchez-López and M. A. Quiroga-Estévez (Madrid: Centro de Estudios Ramón Areces), 84–88.

Hernández-Mendo, A., Blanco-Villaseñor, A., Pastrana, J. L., Morales-Sánchez, V., and Ramos-Pérez, F. J. (2016). SAGT: aplicación informática para análisis de generalizabilidad [SAGT: New software for generalizability analysis]. *Rev. Iberoam. Psicol. Ejerc. Deporte* 11, 77–89.

Hernández-Mendo, A., Castellano, J., Camerino, O., Jonsson, G., Blanco-Villaseñor, Á., Lopes, A., et al. (2014). Programas informáticos de registro, control de calidad del dato y análisis de datos [Observational software, data quality control and data analysis]. *Rev. Psicol. Deporte* 23, 111–121.

Hernández-Mendo, A., López-López, J. A., Castellano, J., Morales-Sánchez, V., and Pastrana, J. L. (2012). Hoisan 1.2: programa informático para uso en metodología observacional [Hoisan 1.2: Computer program for

use in observational methodology]. *Cuad. Psicol. Deporte* 12, 55–78. doi: 10.4321/S1578-84232012000100006

Lapresa, D., Arana, J., Anguera, M. T., and Garzón, B. (2013). Comparative analysis of sequentiality using SDIS-GSEQ and THEME: a concrete example in soccer. *J. Sports Sci.* 31, 1687–1695. doi: 10.1080/02640414.2013.796061

Loffing, F., Sölter, F., Hagemann, N., and Strauss, B. (2015). Accuracy of outcome anticipation, but not gaze behavior, differs against left-and right-handed penalties in team-handball goalkeeping. *Front. Psychol.* 6:1820. doi: 10.3389/fpsyg.2015.01820

López, J., Valero, A., Anguera, M. T., and Díaz, A. (2016). Disruptive behavior among elementary students in physical education. *SpringerPlus* 5:1154. doi: 10.1186/s40064-016-2764-6

Luckwü, R. M., and Guzmán, J. F. (2011). Deportividad en balonmano: un análisis desde la teoría de la autodeterminación [Sportsmanship in handball: an analysis based on the self-determination theory]. *Rev. Psicol. Deporte* 20, 305–320.

Martín, I., González, A., Cavalcanti, L. A., Chirosa, L. J., and Aguilar, J. (2013). Fiabilidad y optimización del programa PROTODEBA v 1.0 para la observación de la Toma de decisiones en balonmano [Reliability and optimization of PROTODEBA v 1.0 software for the observation of the decision making in handball]. *Cuad. Psicol. Deporte* 13, 63–70. doi: 10.4321/S1578-84232013000100007

Mascarenhas, D. R. D., Collins, D., and Mortimer, P. (2005). Elite refereeing perfomance: developing a model for sport science support. *Sport. Psychol.* 19, 364–379. doi: 10.1123/tsp.19.4.364

Mascarenhas, D. R. D., and Smith, N. C. (2011). "Developing the performance brain: decision making under pressure," in *Performance psychology – Developing a peak performance culture,* eds D. Collins, H. Richards and C. Button (Edinburgh: Elsevier), 245–267.

Morillo, J. P., and Hernández-Mendo, A. (2015). Análisis de la calidad del dato de un instrumento para la observación del ataque en balonmano playa [Analysis of quality data a tool for observing the attack in beach handball]. *Rev. Iberoam. Psicol. Ejerc. Deporte* 10, 15–22.

Morillo, J. P., Reigal, R. E., and Hernández-Mendo, A. (2015). Análisis del ataque posicional de balonmano playa masculino y femenino mediante coordenadas polares [Analysis of positional attack in beach handball male and female with polar coordinates]. *Rev. Int. Cien. Deporte* 11, 226–244. doi: 10.5232/ricyde2015.04103

Perea, A., Castellano, J., Alday, S., and Hernández-Mendo, A. (2012). Analysis of behaviour in sports through polar coordinate analysis with MATLAB. *Qual. Quant.* 46, 1249–1260. doi: 10.1007/s11135-011-9435-z

Philippe, F. L., Vallerand, R. J., Andrianarisoa, J., and Brunel, P. (2009). Passion in referees: examining their affective and cognitive experiences in sports situations. *J. Sport Exerc. Psychol.* 31, 77–96. doi: 10.1123/jsep.31.1.77

Pinder, R., Davids, K., Renshaw, I., and Araujo, D. (2011). Manipulating informational constraints shapes movement reorganization in interceptive actions. *Atten. Percept. Psychophys.* 73, 1242–1254. doi: 10.3758/s13414-011-0102-1

Plessner, H. (2005). "Positive and negative effects of prior knowledge on referee decisions in sports," in *The Routines of Decision-Making* eds T. Betsch, and S. HAberstroh (Hillsdale: Lawrence Erlbaum), 311–324.

Prudente, J., Sousa, D., Sequeira, P., López-López, J. A., and Hernández-Mendo, A. (2017). Analyzing the influence of playing time and partial score on the tactical behavior in the duel 2 vs. 2 in the offensive process in handball, using the polar coordinates technique. *An. Psychol.* 33, 515–529. doi: 10.6018/analesps.33.3.271071

Sackett, G. P. (1980). "Lag sequential analysis as a data reduction technique in social interaction research," in *Exceptional Infant. Phychosocial Risks in Infant-Environment Transactions,* eds D. B. Sawin, R. C. Hawkins, L. O. Walker and J. H. Penticuff (New York, NY: Brunner/Mazel), 300–340.

Santos, S., Sarmento, H., Alves, J., and Campaniço, J. (2014). Construcción de un instrumento para la observación y el análisis de las interacciones en el waterpolo [Constructing an instrument to observe and analyse interactions in waterpolo]. *Rev. Psicol. Deporte* 23, 191–200.

Santoyo, C., Jonsson, G. K., Anguera, M. T., and López-López, J. A. (2017). Observational analysis of the organization of on-task behavior in the classroom using complementary data analyses. *An. Psychol.* 33, 497–514. doi: 10.6018/analesps.33.3.271061

Sarmento, H., Anguera, M. T., Campaniço, J., and Leitão, J. (2010). Development and validation of a notational system of study the offensive process in football. *Medicina (Kaunas)* 46, 401–407.

Simmons, P. (2011). Competent, dependable and respectul: football refereeing as a model for communicating fairness. *Ethical Space: Int. J. Commun. Ethics* 8, 33–42.

Souchon, N., Cabagno, G., Traclet, A., Dosseville, F., Livingstone, A., Jones, M., et al. (2010). Referees' decision-making and player gender: the moderating role of the type of situation. *J. Appl. Sport Psychol.* 22, 1–16. doi: 10.1080/10413200903250476

Souchon, N., Cabagno, G., Traclet, A., Trouilloud, D., and Maio, G. (2009). Referees' use of heuristics: the moderating impact of standard of competition. *J. Sports Sci.* 27, 695–700. doi: 10.1080/026404109028 74729

Sousa, D., Prudente, J., Sequeira, P., and Hernández-Mendo, A. (2014). Análise da qualidade dos dados de um instrumento para observação do 2 vs 2 no andebol [Analysis of data quality in an instrument to observe 2 vs. 2 in handball]. *Rev. Iberoam. Psicol. Ejerc. Deporte* 9, 173–190.

Tarragó, R., Iglesias, X., Lapresa, D., Anguera, M. T., Ruiz-Sanchís, L., and Arana, J. (2017). Analysis of diachronic relationships in successful and unsuccessful behaviors by world fencing champions using three complementary techniques. *An. Psychol.* 33, 471–485. doi: 10.6018/analesps.33.3. 271041

Weigel, P., Raab, M., and Wollny, R. (2015). Tactical decision making in team sports-a model of cognitive processes. *Int. J. Sports Sci.* 5, 128–138. doi: 10.5923/j.sports.20150504.03

Weinberg, R. S., and Richardson, P. A. (1990). *Psychology of Officiating*. Champaign, IL: Leisure Press.

Weston, M., Castagna, C., Impellizzeri, F. M., Bizzini, M., Williams, A. M., and Gregson, W. (2012). Science and medicine applied to soccer refereeing. *Sports Med.* 42, 615–631. doi: 10.2165/11632360-000000000-00000

Development of a Work Climate Scale in Emergency Health Services

Susana Sanduvete-Chaves[1], José A. Lozano-Lozano[2], Salvador Chacón-Moscoso[1,2]* and Francisco P. Holgado-Tello[3]

[1] Departamento de Psicología Experimental, Universidad de Sevilla, Seville, Spain, [2] Departamento de Psicología, Universidad Autónoma de Chile, Santiago de Chile, Chile, [3] Departamento de Metodología de las Ciencias del Comportamiento, Universidad Nacional de Educación a Distancia, Madrid, Spain

An adequate work climate fosters productivity in organizations and increases employee satisfaction. Workers in emergency health services (EHS) have an extremely high degree of responsibility and consequent stress. Therefore, it is essential to foster a good work climate in this context. Despite this, scales with a full study of their psychometric properties (i.e., validity evidence based on test content, internal structure and relations to other variables, and reliability) are not available to measure work climate in EHS specifically. For this reason, our objective was to develop a scale to measure the quality of work climates in EHS. We carried out three studies. In Study 1, we used a mixed-method approach to identify the latent conceptual structure of the construct *work climate*. Thus, we integrated the results found in (a) a previous study, where a content analysis of seven in-depth interviews obtained from EHS professionals in two hospitals in Gibraltar Countryside County was carried out; and (b) the factor analysis of the responses given by 113 EHS professionals from these same centers to 18 items that measured the work climate in health organizations. As a result, we obtained 56 items grouped into four factors (work satisfaction, productivity/achievement of aims, interpersonal relationships, and performance at work). In Study 2, we presented validity evidence based on test content through experts' judgment. Fourteen experts from the methodology and health fields evaluated the representativeness, utility, and feasibility of each of the 56 items with respect to their factor (theoretical dimension). Forty items met the inclusion criterion, which was to obtain an Osterlind index value greater than or equal to 0.5 in the three aspects assessed. In Study 3, 201 EHS professionals from the same centers completed the resulting 40-item scale. This new instrument produced validity evidence based on the internal structure in a second-order factor model with four components ($RMSEA = 0.079$, $GFI = 0.97$, $AGFI = 0.97$, $CFI = 0.97$; $NFI = 0.95$, and $NNFI = 0.97$); absence of Differential Item Functioning (DIF) in 80% of the items; reliability ($\alpha = 0.96$); and validity evidence based on relations to other variables, specifically the test-criterion relationship ($\rho = 0.680$). Finally, we discuss further developments of the instrument and its possible implications for EHS workers.

Keywords: work climate, emergency health services, mixed methods, content validity, reliability, construct validity, criterion validity

INTRODUCTION

One of the main priorities of modern organizations is fostering a positive work climate because it promotes greater productivity, satisfaction, stability, and commitment to the organization (Lozano-Lozano et al., 2013; Lee et al., 2016; Meneghel et al., 2016). There are three main types of definitions for the *work climate* construct: (1) those based on objective and structural characteristics of organizations (Schneider et al., 2013), (2) those that emphasize individual psychological features (Zadow et al., 2017), and (3) those focused on both organizational and individual levels. This last perspective emphasizes workers' perception of the structure and processes occurring in work groups (Schulz et al., 2017). This work is based on the third group of definitions because it is the most frequently used and the most complete, as it considers both organizational and individual points of view.

Independent of the type of definition, there is no consensus on what the main components that form the construct *work climate* are. In this sense, we find that measurement instruments are based on various numbers of components as follows: **three:** clarity, support and challenge (Stringer, 2002; Perry et al., 2005) and job satisfaction, organizational commitment, and motivation to continue working (Zacher and Yang, 2016); **four:** authority, efficiency, innovation, and adaptation (Payne and Mansfield, 1978); **five:** communication, work conditions, job involvement, self-realization, and supervision (Torres and Zegarra, 2015) and culture, climate, burnout, engagement, and psychosomatization (Uribe-Prado et al., 2015); **six:** organizational clarity, rewards, decisions, leadership, social interaction, and opening (Gómez, 2004); **seven:** ability, recognition, internal organization, satisfaction, information received, knowledge of management, and goals and management responsiveness (García et al., 2010); **eight:** authority, motivation, communication, influence, decision, planning, control, and performance (Likert, 1967) and compensation and justice, teamwork, quality and effectiveness, communication, environmental sustainability, trust, security, and support (Zenteno and Durán, 2016) and relationships, management style, sense of belonging, remuneration, availability of resources, stability, clarity and consistency in management, and shared values (Fernández-Argüelles et al., 2015); **nine:** structure, responsibility, reward, challenge, relationships, cooperation, standards, conflict, and identity (Litwin and Stringer, 1968); **ten:** involvement, cohesion, support, autonomy, organization, pressure, clarity, control, innovation, and comfort (Moos and Insel, 1974); and **twelve:** conflict, equipment, failure, cohesion, autonomy, management, stress, cooperation, social life, lack of shared social life, logging, and marginality (Delgado-Sánchez et al., 2006) and relationship with the boss, work environment, desire for changes, work satisfaction, capacity for making decisions, tolerance, communication and support, opportunity for training, flexibility of schedules, satisfaction with benefits, resources, and stress (Rojas et al., 2011).

The disparity of the definitions of the construct *work climate* inspires the use of diverse instruments to measure it (e.g., Insel and Moos, 1974). Additionally, there are other factors that increase the variability of the instruments to measure work climate: (1) the different levels of specification of work climate, such as an overwork climate (Mazzetti et al., 2016) or competitive work group climate (Fletcher and Nusbaum, 2010); and (2) the adaptation of instruments to specific work contexts and professions, such as civil servants (Popoola, 2016).

In the specific area of health services, we found some consolidated models, such as that proposed by Perry et al. (2005), where a work climate is understood as a part of the organization mainly generated by individual human behavior and interactions. Each work team has its own climate defined as the quality of the internal environment experienced by its members that influences their behavior. It is said that a change in the work climate of an organization implies changes in other aspects in that organization, such as the effectiveness or the quality of the care given to patients (Castaneda and Scanlan, 2015; Safi et al., 2016).

When establishing models of work climate in health services into specific components, we find a huge disparity in aspects such as work satisfaction, which is understood as the feeling of contentment by health workers with their job; this aspect correlates negatively with absenteeism and professional abandonment (Mendoza-Llanos, 2015); interpersonal relationships (the way the different components of a work group relate), productivity (the degree to which the work group is able to achieve their professional goals efficiently) and performance of the health staff (performing daily tasks), components that are positively correlated (Brown and Calnan, 2016); or commitment (level of involvement) established with the patient to achieve their professional goals (Chiang et al., 2017).

Referring to the tools used to measure the work climate in health services, we find several different instruments (Elmi et al., 2017), some of which are specific to concrete professions, such as nursing (Olsen et al., 2017) or anesthesiology (Rama-Maceiras et al., 2012), and certain fields such as mental health (Ehrhart et al., 2014).

There are substantial differences in the characteristics of work in the various departments of hospitals. In many cases, emergency health services (EHS) establish initial contact with the patient (Gill et al., 2017) and serve a large number of patients (Hunt et al., 2006) who improperly use the service on some occasions (Carret et al., 2007) and, in general, have a high incidence of morbidity (Billings and Raven, 2013). EHS are usually structured as three main areas: (1) the emergency department, an area for the triage and prioritization of care needs, observation and assessment of the patient, and referral to other hospital services (Godoi et al., 2016); (2) the intensive care unit, which usually has an isolation zone where healthcare professionals care for patients who, due to serious health problems, require critical care (Leung and Gomersall, 2016); and (3) the emergency ambulance service, which is a mobile unit where the emergency has occurred and, in some cases, avoids transport to the overloaded emergency departments (Wankhade and Mackway-Jones, 2015). Patients who receive care in the EHS are in extremely serious condition, so workers suffer from a high degree of responsibility and consequent stress (Estryn-Behar et al., 2011). They are subject to a number of external

pressures, such as the need for short waiting times, and they experience detrimental impacts to their quality of life (including their own health) due to shift work (Vedaa et al., 2016), even occasionally suffering from posttraumatic stress disorder (Arora et al., 2013; Bragard et al., 2015) because of their daily experience with situations requiring critical decisions in a matter of seconds (Borg Xuereb et al., 2016). It has been verified that emergency personnel work with higher levels of stress than other health professionals and that they are a unique population with greater autonomy (Johnston et al., 2016). In this context, it is necessary to design an instrument adapted to these specific characteristics.

On the one hand, an adequate work climate in EHS allows workers to successfully share the common purpose and responsibility of professional teams, where each member clearly understands their role and combines their skills and knowledge to provide better care for their patients (Ajeigbe et al., 2013). On the other hand, an inadequate work climate in this context can cause significant stress in professionals (Laposa et al., 2003) and personal dissatisfaction, which can have an impact on the quality of health care provided (Hooper et al., 2010), increase the perception of fatigue and distress at work (Adriaenssens et al., 2011), and affect labor productivity (Engelen et al., 2016).

Measuring a work climate would be useful as a first step before acting to improve it to detect weaknesses in the work group that need strengthening. Additionally, in an indirect way, the work climate affects other important variables, such as alcohol consumption in workers (Carreño et al., 2006); safety climate (Sexton et al., 2016); motivation (Li et al., 2016); patient outcomes and nurses' occupational health (Taylor et al., 2011); and job involvement, effort, and performance (Brown and Leigh, 1996).

Some scales to measure work climate in EHS are already available in the literature (Davenport et al., 2007). However, such scales show poor psychometric properties and present only results about their reliability and, in some cases, validity evidence based on relations to other variables (Biggs et al., 2016). We also found proposals for scales that, in addition to reliability, provide data regarding validity evidence based on internal structure, such as the Safety Attitudes Questionnaire (Sexton et al., 2016) applied in different contexts (Patterson et al., 2010), but they do not provide data about validity evidence based on test content or relations to other variables. Other scales applied to EHS propose to measure something other than work climate, such as the Perceptions of Safety Climate and Adherence to Safe Work Practices (Eliseo et al., 2012), which aims to measure safety perception in emergency room conditions, or those designed exclusively to measure the work environment for nursing staff (Swiger et al., 2017).

Due to the poor consensus in the definition and measure of work climate and due to the fact that there is no specific instrument to measure this construct in EHS with adequate psychometric properties, the aim of this study was to develop an instrument to measure the quality of the work climate in EHS specifically. For this purpose, we carried out our research in three stages (American Educational Research Association et al., 2014): (1) we identified the latent conceptual structure of *work climate* using a mixed-method approach based on the information obtained from in-depth interviews using the grounded theory (Lozano-Lozano et al., 2013) and factor analysis; (2) we presented validity evidence of the resultant instrument based on the test content through experts' judgment; and (3) we assessed the psychometric properties of the final version of the instrument: we carried out studies on reliability; validity evidence based on internal structure, specifically Confirmatory Factor Analysis (CFA) and Differential Item Functioning (DIF); and validity evidence based on relations to other variables, specifically test-criterion relationships.

STUDY 1: IDENTIFICATION OF THE LATENT CONCEPTUAL STRUCTURE OF WORK CLIMATE: A MIXED-METHOD APPROACH

Method

Participants

One hundred thirteen EHS workers from two different hospitals in Gibraltar Countryside County, chosen by incidental sampling, participated voluntarily. The inclusion criteria for the sample selection integrated (a) currently working in EHS when the study was carried out, and (b) having worked in EHS for at least 6 months. In total, 59.2% were women and 40.8% were men (age: $M = 37.68$ and $SD = 8.79$). Fifty percent were nurses, 22.1% were nursing assistants, 21.2% were doctors, 5.3% were orderlies, and 1.4% were administrative officers; 59.3% worked in the emergency department, 37.2% in the critical care unit and 3.5% in the emergency ambulance service. Concerning the type of contract, 36.1% had an indefinite contract, 34.1% had an interim contract, 27.8% had a temporary contract, and 2% had a contract for work and services. They had a mean of 12.83 years of experience in their profession ($SD = 7.97$); 51% had more than 5 years of experience in their profession, 21.1% had 2–5 years, and 23.8% had <2 years of experience. They had a mean of 9.38 years of experience in the current workplace ($SD = 8.48$); 52.4% had more than 5 years of experience, 28.8% had 2–5 years of experience, and 18.8% <2 years of experience.

Instruments

We used a list of 38 items to measure work climate in EHS specifically, which were obtained from a previous study (Lozano-Lozano et al., 2013). To produce this list, all EHS workers in the two hospitals in Gibraltar Countryside County (including doctors, nurses, nursing assistants, orderlies, administrative officers, security stuff and caretakers) were asked to voluntarily participate in answering in-depth individual interviews related to aspects such as the operation of the service, the organization, job satisfaction, the needs of the service, communication, productivity, relationship with authorities, conflicts and their resolutions, innovation and training. Eighteen EHS professionals responded: 9 doctors, 5 nurses, and 4 nursing assistants. After excluding the interviews that did not provide additional information to the built category system, the sample was formed by four doctors and three nurses. The inductive-deductive process proposed by Strauss and Corbin (1998) was followed to extract the information gathered from these interviews.

Supplementary Table 1 presents the list of 38 items and their origins (codes and coding families).

Additionally, the questionnaire we used to measure the work climate in EHS was formed by 18 items rated from 1 (not at all) to 5 (to a great degree) on a five-point rating scale, available in Perry et al. (2005) to measure work climate in organizations in general, which was translated into Spanish according to the International Test Commission Guidelines for translating and adapting tests (International Test Commission, 2005; Barbero-García et al., 2008). Specifically, the following back-translation method was applied to the original English version: (a) the original version was translated into Spanish by a bilingual three-component expert group, (b) the new version was again translated into English by another bilingual translator who was not among those who formed the expert group in the first stage, and (c) the discrepancies that arose were discussed and the appropriate corrections of the new version were made. **Table 1** presents the items in English. The Spanish translation is available in Supplementary Data 1. We chose these specific items for three main reasons (Perry et al., 2005): (1) the authors based their construction on a consolidated theoretical model (Stringer, 2002); (2) they presented evidence of adequate reliability, construct validity, and invariance across gender, management status, and educational level; and (3) although they are context independent, their validation was carried out in a public health organization (similar to our context, the EHS).

Procedure

Survey methodology

We contacted the management of the EHS of Gibraltar Countryside County (two different hospitals) and presented the reasons for the research and the characteristics of the study. Once we obtained the corresponding authorization, the 18-item questionnaire obtained from Perry et al. (2005) was administered to the staff of the EHS (250 participants). To ensure anonymity, the workers completed the questionnaire in a room in the hospital and, once finished, they deposited it in a ballot box.

With the data gathered, we carried out an exploratory factor analysis (EFA) (Pérez-Gil et al., 2000). SPSS 24 was used to store the data; PRELIS and LISREL 9.2 were used to carry out the data analysis. First, we created a polychoric correlation matrix between all of the variables that came into the analysis (Holgado-Tello et al., 2010). Second, we checked whether such a matrix met the assumptions necessary to be able to develop an EFA (Yela, 1957) by calculating the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity. Third, we performed an EFA using a principal components factor extraction and Kaiser's varimax method for orthogonal rotation.

Mixed-method approach

With the survey methodology, we calculated the factor structure of the 18 items obtained from Perry et al. (2005) in a specific EHS context. To obtain more detailed information about the important aspects that form a work climate in this specific context, we integrated these 18 items considering their factor structure and the 38 items obtained from in-depth interviews (Lozano-Lozano et al., 2013) considering their coding families.

TABLE 1 | Factor weights of the questionnaire to measure work climate (obtained from Perry et al., 2005).

Items	F1	F2	F3	F4
1. We take pride in our work	0.761			
2. Our work group is known for quality work	0.683			
3. We have a common purpose	0.599			
4. We seek to understand the needs of our clients	0.585			
5. We readily adapt to new circumstances	0.571			
6. We strive to achieve successful outcomes	0.487			
7. We understand the relevance of the job of each member in our group		0.803		
8. We understand each other's capabilities		0.768		
9. We strive to improve our performance		0.623		
10. We pay attention to how well we are working together		0.618		
11. We are recognized for our individual contributions			0.740	
12. We have the resources we need to do our jobs well			0.722	
13. We have a plan that guides our activities			0.617	
14. We participate in the decisions of our work group			0.528	
15. Our work group is productive			0.511	
16. Our work is important				0.824
17. We develop our skills and knowledge				0.719
18. We are clear about what is expected in our work				0.511

The reliability of the assignment item—factor was checked using SPSS 24 in two ways: (a) by intracoder, where one author, JALL, completed the assignment and repeated the same task after 1 month; and (b) by intercoder, where two different authors, SCM and SSC, independently performed the same task. In both cases, Cohen's kappa (κ) coefficient was calculated.

Results

Survey Methodology

Of the potential 250 participants, 113 answered, which is a participation rate of 45.2%. Assumptions to develop an EFA were accepted: the KMO was 0.851, and Bartlett's test of sphericity resulted in $\chi^2_{(153)} = 804.174$, $p < 0.001$.

The EFA provided a four-factor solution that explained 61.33% of the common variance. Factor 1 (F) *Work satisfaction* produced an eigenvalue of 6.57 and explained 36.54% of the common variance. F2 *Productivity/achievement of aims* produced an eigenvalue of 1.93 and explained 10.79% of the common variance. F3 *Interpersonal relationships* produced an eigenvalue of 1.34 and explained 7.44% of the common variance. Finally, F4 *Performance at work* produced an eigenvalue of 1.17 and explained 5.47% of the common variance.

Table 1 shows the highest factor loading for each item. All values were higher than 0.48.

Mixed-Method Approach

Table 2 presents the scale used to measure the work climate in EHS after combining the results obtained with the in-depth interviews and the survey methodology. The assignment item—factor produced an intracoder reliability κ of 0.922 and an

TABLE 2 | Resultant work climate scale in emergency health services after finishing the mixed-method approach (Study 1, in-depth interviews—survey) and Osterlind indexes obtained in Study 2, validity evidence based on test content.

Item	R	U	F
FACTOR 1. WORK SATISFACTION			
1. We take pride in our work	0.50	0.59	0.55
2. Our work group is known for quality work	0.58	0.75	0.77
3. We have a common purpose	0.58	0.67	0.55
4. We seek to understand the needs of our clients	0.79	0.59	0.64
5. We readily adapt to new circumstances	0.63	0.54	0.55
6. We strive to achieve successful outcomes	0.96	0.88	0.73
7. We have the experience to do our work well	0.79	0.77	0.65
8. Our workday is adequate to develop our work	0.68	0.63	0.50
9. Time to perform my work well	0.54	0.64	**0.45**
10. Time to attend to our users	0.58	0.55	**0.31**
11. Similar amount of work as in other centers	**0.17**	**0.18**	**0.25**
12. Good relations with the other services	0.75	0.60	0.50
13. We have a good relationship with our patients	**0.46**	**0.32**	**0.40**
14. Good relationship with relatives	**0.14**	**0.09**	**0.33**
15. Patients have good relationship with hospital	**0.46**	**0.30**	**0.33**
FACTOR 2. PRODUCTIVITY/ACHIEVEMENT OF AIMS			
16. Relevance of the job of each member	0.58	0.73	0.50
17. We understand each other's capabilities	0.71	0.63	0.68
18. We strive to improve our performance	0.63	0.55	**0.35**
19. We pay attention to how well we are working	0.68	0.65	**0.35**
20. We have the necessary infrastructure	0.71	0.63	0.68
21. We receive the necessary training	0.67	0.88	0.60
22. The characteristics of our service are appropriate	0.63	0.70	0.61
23. Our service works correctly	0.71	0.90	0.56
24. Our work group is known for its productivity	0.79	0.79	0.70
25. Our work is guided by protocols for action	**0.32**	**0.15**	0.60
26. We feel motivated doing our work	0.79	0.79	0.50
27. The merit of our good job is recognized	0.71	0.73	0.50
28. Our colleagues value our profession	0.77	0.73	0.60
29. We are appreciated for the work we do	0.83	0.86	0.60
30. Our specialization is recognized	0.75	0.68	0.55
31. Our expectations have been fulfilled	0.83	0.80	0.60
32. Our type of patient fits with the service	0.75	0.60	0.61
33. We always attend to patients in emergencies	**0.35**	**0.45**	**0.44**
34. We know our patients' characteristics	0.59	0.70	0.50
35. We coordinate with the other hospital services	0.65	0.50	0.78
FACTOR 3. INTERPERSONAL RELATIONSHIPS			
36. We are recognized for our individual contributions	0.71	0.82	0.55
37. We have the resources we need	**0.42**	**0.45**	0.70
38. We have a plan that guides our activities	0.50	0.55	0.60
39. We participate in the decisions of our group	0.67	0.70	0.60
40. Our work group is productive	**0.33**	**0.32**	0.50
41. We have good communication within the group	0.79	0.82	0.65
42. Good relationship between members	0.83	0.86	0.61
43. I feel comfortable with my work group	0.82	0.80	0.55
44. I have good personal relationships	0.67	0.64	0.50
45. We work in a good work group climate	0.67	0.77	0.61
46. Before entering, I knew some of its members	**0.33**	**0.10**	**0.40**

(Continued)

TABLE 2 | Continued

Item	R	U	F
47. We know how to manage conflict	**0.33**	**0.40**	0.50
48. Good relationship with our work group chief	0.71	0.77	**0.40**
FACTOR 4. PERFORMANCE AT WORK			
49. Our work is important	0.92	0.86	0.70
50. We develop our skills and knowledge	0.83	0.82	0.60
51. We know what is expected in our work	0.79	0.73	0.65
52. I know my professional shortcomings	0.54	0.64	0.50
53. We know the functions of the members	0.75	0.75	0.72
54. Our patients fit the specialty of our service	0.79	0.73	0.70
55. We know our shortcomings as a group	0.75	0.60	0.50
56. We can make proposals to improve our work	0.55	**0.25**	**0.44**

Some items are presented in short format. The full version is available in Supplementary Data 2. N = 14. Items from the survey method are marked in italics. The rest come from the in-depth interviews. R, representativeness; U, utility; F, feasibility. Values under the cut point (<0.5) are marked in bold.

intercoder reliability κ of 0.879. Both results were considered adequate.

F1 *Work satisfaction* refers to feelings evoked in workers by their job and their conditions: self-confidence due to the experience, the adequacy of the workday or time for each patient, contentment with relationships with other professionals outside the group and patients and their relatives, pride, success, cohesion, or nervousness facing new circumstances. F1 is formed using 15 items, six from the survey method (items 1–6) and nine from the in-depth interviews (items 7–15, corresponding to items 1–9 in Supplementary Table 1).

F2 *Productivity/achievement of aims* refers to the perception of workers having everything they need to do their job or, on the contrary, lacking what they need to achieve their goals: understanding of the relevance, capabilities, and specialization of others; the value of working in a group; motivation and fulfillment of expectations; recognition of their work as a group; self-improvement; infrastructure; training and the characteristics and functioning of their service; patients' characteristics fitting with their specialization and knowledge of such characteristics; protocols; and coordination with other hospital services. F2 is formed by 20 items, four from the survey method (items 16–19) and 16 from the in-depth interviews (items 20–35, corresponding to items 10–25 in Supplementary Table 1).

F3 *Interpersonal relationships* refer to the feelings when workers relate to other members of the group and aspects that influence such feelings: the quality of the communication, their relationship, the level of comfort, their friendship, their conflicts, being recognized for their individual contributions, having the resources they need, following a plan, participating in decision-making, and productivity. Thirteen items, five from the survey method (items 36–40) and eight from the in-depth interviews (items 41–48, corresponding to items 26–33 in Supplementary Table 1) form F3.

Finally, F4 *Performance at work* includes everything related to the development of workers' job placement: the perceived importance of their job and capacity to decide how to improve

their performance; the skills and knowledge they use; and their knowledge of their tasks and others' tasks, their individual and group limitations, and their patients. F4 is formed by eight items, three from the survey method (items 49–51) and five from the in-depth interviews (items 52–56, corresponding to items 34–38 in Supplementary Table 1).

STUDY 2: VALIDITY EVIDENCE OF THE RESULTANT INSTRUMENT BASED ON TEST CONTENT

Method
Participants

We requested 27 experts' collaboration. The inclusion criterion was to have more than 3 years of experience in social and health sciences methodology and/or in EHS as a health professional. Fourteen experts answered, which is considered a moderate number of participants ($10 \leq N \leq 30$) for a content validity study (Prieto and Muñiz, 2000).

Eight experts (57.1%) were men and six (42.9%) were women. Their mean age was 48.83 ($SD = 8.99$). Regarding their professions, nine (64.3%) were professors specializing in methodology, design, psychometrics and/or data analysis, four (28.6%) were physicians from EHS, and one (7.1%) was a senior technician in clinical analysis and emergency and intensive care medicine. They were in their professions a mean of 23.67 years ($SD = 9.46$).

Instruments

The questionnaire used to obtain validity evidence based on test content through experts' judgment was composed of the 56 items obtained in the previous mixed-method study ordered in the four factors or dimensions found (see Supplementary Data 2). Each item presented three five-point Likert scales (Sanduvete-Chaves et al., 2013) to measure their representativeness (R), utility (U), and feasibility (F) (Chacón-Moscoso et al., 2016). Additionally, there was a final open-format question to receive comments for improvements for the proposed scale. Supplementary Data 3 presents the Spanish version of the questionnaire used to obtain validity evidence based on test content, which was completed by native Spanish speakers.

Procedure

The questionnaire to obtain validity evidence based on test content was sent by e-mail to 27 experts. After the third request, a total of 14 experts gave their answers. Anonymity was assured.

The Osterlind index of congruence (1998) was used to quantify the consensus between experts regarding the adequacy item—factor (theoretical dimension) (Glück et al., 2013). The formula used was

$$I_{ik} = \frac{(N-1)\sum_{j=1}^{n} X_{ijk} + N\sum_{j=1}^{n} X_{ijk} - \sum_{j=1}^{n} X_{ijk}}{2(N-1)n}$$

where N = the number of dimensions of the instrument; X_{ijk} = each score given by each expert to each item referring to

each aspect (R, U, and F); and n = the number of experts. The results could be from -1 to $+1$, 0 being the highest possible level of disagreement between experts. The inclusion criterion was to produce at least 0.5 (Osterlind, 1998) in the three aspects evaluated (R, U, and F).

Results

Considering that we requested assessments from 27 experts and 14 answered, we obtained a 51.9% rate of participation.

Forty items met the inclusion criterion; 16 did not meet it. **Table 2** presents the Osterlind indexes for each item and each aspect (R, U, and F). The most frequently undervalued aspect was F: from the 16 items excluded, 14 were rated under 0.5 in this aspect.

Additionally, several experts proposed a reorganization of some items with respect to its factor assignment through the open-format question: four and three experts, respectively, suggested including item 49, referring to the perceived importance of their work, and item 50, referring to the development of their skills and knowledge at work, in F1 *Work satisfaction* instead of F4 *Performance at work*. Four experts suggested including item 2, referring to the quality of work and item 3, referring to the existence of a common purpose, in F2 *Productivity/achievement of aims* instead of F1. Two experts proposed including item 51, referring to the knowledge of what is expected in their work, in F2 instead of F4 *Performance at work*; and vice versa, five experts proposed including item 17, referring to the understanding of each other's capabilities, in F4, instead of F2.

The four authors of this work considered these proposals individually. After debating, we accepted all the suggestions given by consensus, based on the adequacy item—factor (dimension), from a substantive point of view.

STUDY 3: RELIABILITY AND VALIDITY EVIDENCE BASED ON INTERNAL STRUCTURE AND RELATIONS TO OTHER VARIABLES

Method
Participants

Two hundred and one EHS professionals from the same two hospitals in Study 1 participated voluntarily. The inclusion criteria for the sample selection included (a) working in EHS when the study was carried out, and (b) having worked in EHS for at least 6 months. In total, 61.7% were women and 38.3% were men (age: $M = 41.6$ and $SD = 10.23$). Of these, 40.3% were nurses, 33.3% were doctors, 22.4% were nursing assistants, and 4% were orderlies. Sixty-seven percent worked in the emergency department, 30.5% to the critical care unit, and 2.5% to the emergency ambulance service; 44.6% had a temporary contract, 26.3% had an indefinite contract, 17.1% had an interim contract, 10.9% were residents, and 1.1% had a contract for work and services. Regarding their years of experience in their profession, they had a M of 15.12 ($SD = 9.64$); 49.2% had more than 5 years in their profession, 22.4% had 2–5 years of experience, and 28.4%

had <2 years of experience. Regarding the number of years in their current workplace, the mean was 8.52 ($SD = 8.31$); 81.6% had more than 5 years of experience, 13.8% had 2–5 years of experience, and 4.6% had <2 years of experience.

Instruments

The scale used was formed via instructions for participants and the items that met the inclusion criterion in Study 2 (content validity), grouped into factors per expert suggestion (see Supplementary Data 4 for the English version and Supplementary Data 5 for the Spanish version). Each item was valued from 1 (strongly disagree) to 5 (strongly agree).

Furthermore, we added one extra omnibus item: *As a whole, the work climate of my work group is good;* in the Spanish version, *De manera global, el clima laboral de mi grupo de trabajo es bueno;* translation carried out according to the International Test Commission recommendations (International Test Commission, 2005; Barbero-García et al., 2008), which was also valued from 1 (strongly disagree) to 5 (strongly agree) and was used as a criterion in Study 3. This item was used as a criterion based on the following reasons (Holgado-Tello et al., 2015): (a) it can be considered an appropriate direct measure to relate to another indirect measure (the resulting scale) in the sense that both refer to the same construct (work climate in EHS); (b) the Likert-type scale of the item permits obtaining a linear monotonic function with respect to the attitude measured as the item characteristic curve, and individual differences in participants' attitude provoke variation in responses; and (c) no instrument with its psychometric properties tested that measures work climate in EHS was available as a criterion.

Procedure

The information was gathered using Google for Work applications: Drive, Forms, and Spreadsheets. The scale was administered by two procedures: (a) using a laptop with internet access in the hospital and (b) sending emails to all the EHS professionals with the link to the scale; in this case, they answered outside the work context.

SPSS 24 was used to store the data and calculate the internal consistency of the test, the average discrimination index, and validity evidence based on a test-criterion relationship and needed assumption tests. PRELIS and LISREL 9.2 were used to estimate the polychoric correlation matrix to verify bivariate normality and to carry out the CFA.

The internal consistency of the items was calculated using Cronbach's alpha coefficient, following criteria established by George and Mallery (2003). Values > 0.9 were considered excellent, 0.8–0.9 good, and 0.7–0.8 acceptable. Following criteria by Tavakol and Dennick (2011), values equal to or higher than 0.7 were considered appropriate.

The average discrimination index was also calculated: values >0.4 were considered excellent (Sabri, 2013), 0.3–0.4 good, and 0.2–0.3 adequate (Barbero-García, 1993). Obtaining adequate values in internal consistency and on the average discrimination index was considered an essential requirement before proceeding to study the factor structure.

To evaluate the factor structure of the scale, we estimated the polychoric correlations and the asymptotic variance-covariance matrix. A Pearson correlation matrix was not estimated because the items were in ordinal scales, and therefore, the responses cannot be treated as if they were quantitative because all participants situated at different points of the interval may be assigned the same score; the use of Pearson correlations for ordinal scales would undervalue the real correlations (Holgado-Tello et al., 2010). In these cases, polychoric correlations are the most consistent and robust estimator (Morata-Ramírez and Holgado-Tello, 2013).

The use of the polychoric correlation matrix is only appropriate if we previously accept the assumption of bivariate normal distribution. For this purpose, we calculated the chi-square test (χ^2) and the percentage of tests that rejected the null hypothesis of bivariate normality for each pair of correlations, assuming a 95% confidence level and the Bonferroni correction, calculating the value of α to use in the comparison of each contrast with the formula α/c ($\alpha = 0.05$ corresponding to a 95% confidence level and c as the number of contrasts [$c =$ (number of items x number of items – 1)/2]. Due to the sensitivity of χ^2 in large samples, we also calculated the root mean square error of approximation (RMSEA). We concluded that the parameter estimation was not significantly affected when RMSEA values did not exceed 0.1 (Hooper et al., 2008).

Once we tested the assumption of bivariate normal distribution, we then tested the resultant model from Studies 1 and 2 using a CFA (Bagozzi and Yi, 2012). The tested model was a second-order factor model that measures work climate in EHS formed by four factors (**Table 2**): F1 *Work satisfaction* (items 1–10); F2 *Productivity/achievement of aims* (items 11–30); F3 *Interpersonal relationships* (items 31–35); and F4 *Performance at work* (items 36–40).

The estimation method used was the unweighted least squares, which is appropriate for polychoric correlations and ordinal variables distributed asymmetrically (Jöreskog, 2003; Morata-Ramírez et al., 2015).

The lambda parameter corresponding to the relationship of the first item with each factor was fixed at 1 to (a) solve the problem of identification of the model and (b) establish the measurement scale of the latent variables.

The standardized factor loadings were calculated. Additionally, several fit indices were used to reach conclusions about the adequacy of the model: (a) the χ^2 test, where the acceptance of the null hypothesis ($p \geq 0.05$) implied a good fit of the model; (b) the consistent Akaike information criterion (CAIC) with which the model was considered appropriate when the value of the index was closer to the value for the saturated model than the independent one (the smaller the values, the better the fit) (Bandalos, 1993); (c) the root mean square error of approximation (RMSEA) (Hooper et al., 2008), where values lower than 0.05 were considered a good fit, values between 0.08 and 0.1 a reasonable fit, and values greater than 0.1 unfit (Browne et al., 1993); (d) the goodness-of-fit index (GFI); (e) the adjusted goodness-of-fit index (AGFI) (Hooper et al., 2008); (f) the comparative fit index (CFI) (Byrne, 1998); (g) the normed fit index (NFI); and (h) the non-normed fit

index (NNFI) (Hoe, 2008). Indices (d)–(h) were interpreted as indicators of good fit if the values were above 0.9 (Bendayan et al., 2013).

We also studied DIF to obtain additional validity evidence based on the internal structure of the scale. EASY-DIF (González et al., 2011) was used to perform the Mantel-Haenszel procedure for ordinal items. The DIF was calculated by gender; seniority at work (0–15 years and 16–40 years of experience); age (18–42 and 43–65 years old); and type of employment relationship (eventual and permanent). The matching method used was the minimum cell frequency.

Additionally, we calculated the validity evidence based on the test-criterion relationship, correlating the global score (X, the sum of the scores given in the 40 items) and those obtained in each factor with the score given in the omnibus item 41: *As a whole, the work climate of my work group is good* (criterion Y) (Holgado-Tello et al., 2015). Previously, we tested several assumptions to check that the use of the Pearson correlation coefficient (r), a parametric test, was adequate: (a) normality using the Kolmogorov Smirnov test, where $p > 0.05$ implied the acceptance of the assumption (Chakravarti et al., 1967); (b) a test of linearity, where linearity $p < 0.05$ implied the acceptance of the assumption (Field, 2000); and (c) independence of errors using Durbin Watson (d), a statistic that ranged from 0 to 4 and was expected to produce a value close to 2; thus, values between 1.5 and 2.5 implied acceptance of the assumption (Kutner et al., 2004). If one of the three assumptions is rejected, we would opt for a non-parametric correlation test (Spearman's correlation coefficient, ρ). Both the Pearson and Spearman would be interpreted as validity evidence based on a test-criterion relationship when a $p < 0.05$ showed a statistically significant relationship between the global scores and Y.

Results
Participation
From the potential 250 participants, 201 completed the scale, producing a participation percentage of 80.4%; in particular, 199 (79.6%) were recorded in the hospital, and two (0.8%) by email outside the work context.

Internal Consistency
The global internal consistency was excellent ($\alpha = 0.96$). Factors also produced appropriate values, all over 0.7; F1 produced a good result ($\alpha = 0.848$); F2 and F3 produced excellent results ($\alpha = 0.943$ and 0.907, respectively); and F4 produced acceptable, close to good results ($\alpha = 0.791$).

Average Discrimination Index
The global average discrimination index produced an excellent result ($D = 0.601$), as did the specific index for each of the four factors (0.515, 0.646, 0.654, and 0.519, respectively). After testing the appropriateness of the internal consistency and the average discrimination index, we concluded that we could carry out the factor structure study.

Bivariate Normality Assumption
By having 40 items, a total of 780 correlations were obtained (40 × 39/2). The results showed that a bivariate normality

assumption considering χ^2 was accepted in 95.3% of the instances (743 correlations) ($p = 0.05/780 = 0.00006$ using the Bonferroni correction). Additionally, the RMSEA values were lower than 0.1 in 98.5% of occasions (768 correlations). These results support the use of the matrix of polychoric correlations as the basis for the factor analyses.

Standardized Factor Loadings
The standardized factor loadings (lambda) in the CFA (**Table 3**) were appropriate (over 0.3) for all of the items. The gamma values were high for the four factors (0.89, 0.81, 0.82, and 0.86, respectively).

Model Fit
Although the χ^2 test was significant, probably due to the large sample size effect, $\chi^2_{(730)} = 3807.37$, $p < 0.001$, the other fit indexes showed the adequacy of the second-order-four-factor model: $CAIC = 2460.93$ (saturated $CAIC = 5347.60$; independent $CAIC = 39115.98$); $RMSEA = 0.079$, 90% CI [0.075, 0.084]; $GFI = 0.97$; $AGFI = 0.97$; $CFI = 0.97$; $NFI = 0.95$; and $NNFI = 0.97$.

DIF
Thirty-two items (80% of the total) did not present any DIF; the 8 items that presented some DIF were items 2, 3, 5, 6, 18, 23, 27, and 29. For the variable gender, the items with DIF were items 3 ($\chi^2 = 4.38$, $p = 0.03$), 6 ($\chi^2 = 8.3$, $p < 0.001$), 23 ($\chi^2 = 6.44$, $p = 0.01$), 27 ($\chi^2 = 4.73$, $p = 0.02$), and 29 ($\chi^2 = 6.95$, $p < 0.001$). In items 3, 6 and 23, women scored higher than men, while in items 27 and 29, men scored higher than women. In seniority at work, the workers with more years in the company scored higher in items 2 ($\chi^2 = 4.28$, $p = 0.03$) and 5 ($\chi^2 = 6.82$, $p < 0.001$), while workers with fewer years of experience working scored higher in item 18 ($\chi^2 = 7.27$, $p < 0.001$). For the variable age, older workers scored higher in item 5 ($\chi^2 = 6.07$, $p = 0.01$) and item 6 ($\chi^2 = 5.74$, $p = 0.01$), while younger people scored higher in item 18 ($\chi^2 = 4.47$, $p = 0.03$). Finally, for type of employment relationship, workers with permanent contracts scored higher in items 5 ($\chi^2 = 4.94$, $p = 0.02$) and 6 ($\chi^2 = 6.78$, $p < .001$), while workers with eventual contracts scored higher in item 18 ($\chi^2 = 6.19$, $p = 0.01$).

Testing Assumptions before the Criterion Validity Correlation
The normality assumption was accepted on the global scale ($Z = 0.607$, $p = 0.855$), in F1 ($Z = 1.329$; $p = 0.058$) and in F2 ($Z = 0.693$, $p = 0.722$), while it was rejected in F3 ($Z = 1.729$, $p = 0.005$), F4 ($Z = 1.409$, $p = 0.038$), and the criterion ($Z = 3.047$, $p < 0.001$). The linearity assumption was accepted in all instances when relating with Y, X [$F_{(1,105)} = 187.569$, $p \leq 0.001$]; F1 [$F_{(1,168)} = 88.717$, $p \leq 0.001$]; F2 [$F_{(1,138)} = 146.957$, $p < 0.001$]; F3 [$F_{(1,177)} = 141,197$, $p \leq 0.001$]; and F4 [$F_{(1,177)} = 41,335$, $p < 0.001$]. Finally, the independence of errors was accepted in all occasions for the relationship of Y with X ($d = 1.608$), F1 ($d = 1.654$), F2 ($d = 1.634$), F3 ($d = 1.792$), and F4 ($d = 1.761$). As a conclusion, although most assumptions were accepted, because there was no normal

TABLE 3 | Standardized factors loadings obtained in the CFA.

Item	F1	F2	F3	F4
1. We take pride in our work	0.69			
2. We seek to understand the needs of our clients	0.64			
3. We readily adapt to new circumstances	0.51			
4. We strive to achieve successful outcomes	0.71			
5. We have experience to do our work well	0.40			
6. Our workday is adequate to develop our work	0.68			
7. Good relations with the other services	0.87			
8. Relevance of the job of each member	0.73			
9. Our work is important	0.73			
10. We develop our skills and knowledge	0.69			
11. Our work group is known for quality work		0.64		
12. We have a common purpose		0.75		
13. We have the necessary infrastructure		0.75		
14. We receive the necessary training		0.74		
15. The characteristics of our service appropriate		0.76		
16. Our service works correctly		0.76		
17. Our work group is known for its productivity		0.64		
18. We feel motivated doing our work		0.71		
19. The merit of our good job is recognized		0.73		
20. Our colleagues value our profession		0.77		
21. We are appreciated for the work we do		0.87		
22. Our specialization is recognized		0.73		
23. Our expectations have been fulfilled		0.79		
24. Our type of patient fits with the service		0.65		
25. We know our patients' characteristics		0.79		
26. We coordinate with the other hospital services		0.55		
27. We are recognized for our individual contributions		0.81		
28. We have a plan that guides our activities		0.72		
29. We participate in the decisions of our group		0.82		
30. We know what is expected in our work		0.71		
31. We have good communication within the group			0.86	
32. Good relationship between members			0.81	
33. I feel comfortable with my work group			0.87	
34. I have good personal relationships			0.84	
35. We work in a good work group climate			0.90	
36. We understand each other's capabilities				0.64
37. I know my professional shortcomings				0.64
38. We know the functions of the members				0.67
39. Our patients fit the specialty of our service				0.68
40. We know our shortcomings as a group				0.55

Some items are presented in short format. The full version is available in Supplementary Data 4. F1, Factor 1 Work satisfaction; F2, Factor 2 Productivity/achievement of aims; F3, Factor 3 Interpersonal relationships; F4, Factor 4 Performance at work.

distribution in Y, the non-parametric test (ρ) was used to study the relationship between Y and X, and Y and the sum of the different factors.

Validity Evidence Based on a Test-Criterion Relationship

Validity evidence based on a test-criterion relationship for the whole scale presented an adequate result, $\rho_{XY} = 0.68$, $p < 0.001$.

The relationships between the sum of scores in the different factors and Y also yielded adequate results with F1 ($\rho = 0.560$, $p < 0.001$), F2 ($\rho = 0.632$, $p < 0.001$), F3 ($\rho = 0.658$, $p < 0.001$), and F4 ($\rho = 0.430$, $p < 0.001$).

DISCUSSION

Based on the non-consensus of the main components that form the construct *work climate* and the lack of an instrument to measure it, by being applied to EHS specifically and with adequate psychometric properties, we elaborated a scale consisting of four dimensions (work satisfaction, productivity/achievement of aims, interpersonal relationships, and performance at work) and 40 items, which presented evidence of reliability as well as validity based on test content, internal structure and relations to other variables. Supplementary Data 4, 5 present the final version of the scale ready to be used by those who are interested in the English and Spanish versions, respectively.

Thus, as a result, Study 1 (a mixed-method approach) produced 56 items grouped into the four previously mentioned components. Study 2 (validity evidence based on test content through experts' opinion) permitted refinement of the scale in accordance with the representativeness, utility, and feasibility of its items (16 were removed as a result), and six were reorganized into different factors. Finally, Study 3 (psychometric properties) presented a second-order factor model with the four previously mentioned components with adequate values for reliability and validity evidence based on internal structure and test-criterion relationships.

The four factors obtained have certain similarities to those obtained in previous studies: (a) Work satisfaction (F1) is one of the factors that form the construct work climate in models by García et al. (2010), Rojas et al. (2011), and Zacher and Yang (2016); (b) Productivity/achievement of aims (F2) appears in Payne and Mansfield (1978) with the label of efficiency and in Brown and Calnan (2016); (c) Interpersonal relationships (F3) is also a factor in the proposals given by Fernández-Argüelles et al. (2015) and Litwin and Stringer (1968); finally, (d) Performance at work (F4) is a factor also found in Likert (1967) and Brown and Calnan (2016). On the other hand, some elements that are not in previous studies have been found to be relevant to measuring the work climate in EHS, such as items to evaluate personal and group shortcomings and possible quarrels between different professions in terms of delimitation of functions according to their specialty and recognition.

We obtained four factors when compared to the four-factor model by Payne and Mansfield (1978)—authority, efficiency, innovation, and adaptation—and we find one in common: productivity/achievement of aims, which is understood as efficiency. Although the other three factors are different, some of the contents of the items of our proposal are related, e.g., item 29, *We participate in the decisions of our work group,* refers to authority; item 13, *We have the necessary infrastructure to carry*

out our work, and item 14, *We receive the necessary training to carry out our tasks* are related to innovation; and finally, item 3, *We readily adapt to new circumstances,* is related to adaptation.

Additionally, given that the partial starting point of this work was 18 items proposed by Perry et al. (2005) and that 15 have been included in the last version of the proposed scale presented in Supplementary Data 4 (specifically, items 1–4, 8–12, 17, 27–30, and 36), we now compare the psychometric properties of both instruments. The reliability values were adequate in both cases: $\alpha = 0.96$ in the final version of the scale proposed and $\alpha = 0.87$ in Perry et al. (2005). Both instruments obtained evidence based on the internal structure, although the factors found differ substantively, where three (clarity, support, and challenge) were found in Perry et al. (2005). Both instruments obtained evidence based on the relation test—criterion: $\rho_{XY} = 0.68$, $p < 0.001$ when correlating the scale obtained and an omnibus item and $r_{XY} = 0.93$, $p < 0.001$ in Perry et al. (2005) after correlating their instrument with the one proposed by Stringer (2002). In this sense, because both proposals presented adequate psychometric properties, the most important contribution of the scale proposed in this work is its specification in setting; Perry et al. (2005) based their study on public health organizations, and the present work is specifically for EHS, a context with an extremely specific idiosyncrasy (Hunt et al., 2006; Carret et al., 2007; Arora et al., 2013; Johnston et al., 2016; Vedaa et al., 2016; Gill et al., 2017).

Taking into account that any process to obtain evidence is affected by the characteristics of the intervention contexts in this particular case, given the sample characteristics, the instrument was designed with the aim of measuring the global construct *quality of the work climate* (as opposed to proposals centered on a specific aspect within work climate, e.g., Fletcher and Nusbaum, 2010; Mazzetti et al., 2016) for all professionals (as opposed to works based on specific professions, e.g., Rama-Maceiras et al., 2012; Popoola, 2016; Olsen et al., 2017) who work in EHS (as opposed to other specific contexts, e.g., Ehrhart et al., 2014).

To check empirically that the items worked in an unbiased way across different groups, we analyzed their DIF. Thirty-two items (80%) did not present any DIF. The fact that 8 items (20%) presented DIF does not necessarily indicate the weakness of the instrument given that substantive reasons explain the differences (American Educational Research Association et al., 2014). Item 2, *We seek to understand the needs of our clients,* was scored higher by workers with more experience because there is evidence that confirms a direct relationship between experience and implication with the patients (Ballester-Arnal et al., 2016) and the experience and perception of patients' needs in a holistic way instead of only targeting the isolated symptom (Zamanzadeh et al., 2015). Item 3, *We readily adapt to new circumstances,* and 23, *Our expectations when we entered the working group have been fulfilled,* were scored higher by women because females have a greater ability to adapt to new situations (Catalyst, 2007) and are more conformist (Aspiazu, 2016) in work. Item 5, *We have the necessary experience to do our work well*, was scored higher by workers with more experience, those who were older and those who had permanent contracts due to a real difference in the level of the aspect measured in this item (experience). Item 6, *Our workday is adequate to develop our work,* was scored higher

by women, older workers, and those with permanent contracts. Differences in gender can be explained by the fact that women tend to assign greater value and spend time most productively in the workplace since they are also involved in other types of family activities (Artazcoz et al., 2004; Eagly and Carli, 2007); and workers with more experience and permanent contracts spend less time carrying out their functions than those who are less experienced and with eventual contracts, so it is logical that the first group believes more strongly than the second one that the time they have is sufficient to perform their work. Item 18, *We feel motivated when doing our work*, was more valued by younger workers with eventual contracts because studies conclude that there is an inverse relationship between age and job stability and motivation at work (Fernández et al., 2015; Akkermans et al., 2016; Dawson et al., 2017). Finally, items 27, *We are recognized for our individual contributions,* and 29, *We participate in the decisions of our work group,* were scored higher by men because they tend to develop dominant behaviors such as autonomy, independence and decision-making and seeking individual social recognition (Eagly et al., 2004; Godoy and Mladinic, 2009).

The proposed tool can be used in EHS to measure the work climate. Apart from a global value for each worker obtained by summing the values for the 40 items, the tool can be used to detect the global average work climate in EHS, factor by factor, or by studying each item to detect weaknesses so as to implement actions (with the origin in the work colleagues or the boss) to improve them; e.g., if a worker scored item 5 low (*We have the necessary experience to do our work well*), a workmate with more experience could start acting as a temporary mentor, or if the average of a work group in item 20 (*Our colleagues value our profession*) is low, then the boss of the group could implement strategies to improve the multidisciplinary work group conditions. In addition, by measuring the values with the scale before and after the actions implemented to improve the work climate, we can see the effectiveness of such actions (the use of inferential statistics for repeated measures would give evidence about the significance of the change).

In summary, the proposed scale can be used as a tool to diagnose the climate given in the work place and, based on this information, implement action to attempt to improve the situation (Perry et al., 2005). This can involve improvements at different levels since increasing the quality of the work climate will probably influence workers' satisfaction (Hooper et al., 2010), productivity (Brown and Leigh, 1996), interpersonal relationships (Lozano-Lozano et al., 2013), and performance (Engelen et al., 2016). Furthermore, it may act as a protective factor against alcoholism (Carreño et al., 2006), stress (Laposa et al., 2003), fatigue (Adriaenssens et al., 2011), or absenteeism (Mendoza-Llanos, 2015) in workers. Therefore, it will probably increase patients' satisfaction with the service (Hooper et al., 2010; Ajeigbe et al., 2013).

One limitation to highlight is the length of the resultant scale. Although 40 items to measure four factors is apparently not excessive, it is important to note that, currently, workers in EHS receive more patients than they can manage, so it would be difficult for them to find enough time to complete such a long instrument. Another possible limitation is the use of one item as

a criterion. In this sense, it is necessary to carry out research that further develops the instrument obtained in this study.

To cover the first limitation (excessive length of the scale), we will shorten the scale without excluding relevant information to measure the work climate in EHS. First, based on the Spearman-Brown prediction (or prophecy) formula that relates reliability and length of the test (Brown, 1910; Spearman, 1910), we will estimate the number of items that we can remove in each factor without implying a decrease in the reliability coefficient lower than 0.8, a cut-off point to be considered a good result following the criteria established by George and Mallery (2003). Second, we will develop a Delphi method (Dalkey and Helmer, 1963) to determine, with a sample of at least 20 judges (50% experts in psychometrics and 50% EHS workers), the most redundant items to be removed. In further applications of the resulting short scale, we will check whether, as expected, the reliability coefficient maintains good reliability for coefficients in all the factors. To cover the second limitation (the use of one item as a criterion), we will use another scale as a criterion, specifically the scale proposed by Biggs et al. (2016), which presents partial validity evidence (based on relationships with other variables).

Additionally, we will delve into validity based on relations to other variables that present convergent and discriminant evidence of the resultant scale following the multitrait-multimethod matrix proposed by Campbell and Fiske (1959). We will request completion of three different instruments from ∼100 EHS workers: the scale we proposed; another scale with partial validity evidence to measure the work climate in EHS (the one proposed by Biggs et al., 2016); and the Safety Attitudes Questionnaire (Sexton et al., 2016), an instrument that measures safety climate, a construct similar to work climate although not exactly the same (work climate is considered wider). We expect to find high reliability coefficients in the three instruments, a high correlation between the two instruments that measure work climate in EHS (convergent evidence) and a low correlation (at least lower than that obtained in the convergent evidence study) between our instrument and the one that measures a different construct (discriminant evidence).

Finally, we will test its factorial invariance to check whether the use of the scale obtained can be generalized to different hospitals, countries, genders, roles, and professions. First, we will gather a massive number of answers to the scale from participants of different hospitals in one country, from different countries (initially, Spain and Chile), different genders, different roles (coordinator/boss or without being in charge of other people), and different professions (nurses, nursing assistants, doctors, orderlies, and administrative officers). Second, using the software LISREL, we will test the invariance of the structural model and the measurement model across groups (Byrne, 1998).

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the "Declaration on bioethics and human rights, UNESCO, 2005" with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the "Ethics Committee, Universidad Autónoma de Chile."

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00010/full#supplementary-material

REFERENCES

Adriaenssens, J., De Gucht, V., Van Der Doef, M., and Maes, S. (2011). Exploring the burden of emergency care: predictors of stress-health outcomes in emergency nurses. *J. Adv. Nurs.* 67, 1317–1328. doi: 10.1111/j.1365-2648.2010.05599.x

Ajeigbe, D. O., McNeese-Smith, D., Leach, L. S., and Phillips, L. R. (2013). Nurse-physician teamwork in the emergency department: impact on perceptions of job environment, autonomy, and control over practice. *J. Nurs. Adm.* 43, 142–148. doi: 10.1097/NNA.0b013e318283dc23

Akkermans, J., de Lange, A. H., van der Heijden, B. I. J. M., Kooij, D. T. A. M., Jansen, P. G. W., and Dikkers, J. S. E. (2016). What about time? Examining

chronological and subjective age and their relation to work motivation. *Career Dev. Int.* 21, 419–439. doi: 10.1108/CDI-04-2016-0063

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association.

Arora, M., Asha, S., Chinnappa, J., and Diwan, A. D. (2013). Burnout in emergency medicine physicians. *Emerg. Med. Australas.* 25, 491–495. doi: 10.1111/1742-6723.12135

Artazcoz, L., Benach, J., Borrell, C., and Cortès, I. (2004). Unemployment and mental health: Understanding the interactions among gender, family roles, and social class. *Am. J. Public Health* 94, 82–88. doi: 10.2105/AJPH.94.1.82

Aspiazu, E. L. (2016). Heterogeneidad y desigualdades de género en el sector salud: entre las estadísticas y las percepciones sobre las condiciones de trabajo [Heterogeneity and gender inequalities in the health sector: between statistics and perceptions about working conditions]. *Revista Pilquen* 19, 55–66. Available online at: http://www.scielo.org.ar/pdf/spilquen/v19n1/v19n1a05.pdf

Bagozzi, R. P., and Yi, Y. (2012). Specification, evaluation, and interpretation of structural equation models. *J. Acad. Market. Sci.* 40, 8–34. doi: 10.1007/s11747-011-0278-x

Ballester-Arnal, R., Gómez-Martínez, S., Gil-Juliá, B., Ferrándiz-Sellés, M. D., and Collado-Boira, E. J. (2016). Burnout y factores estresantes en profesionales sanitarios de las unidades de cuidados intensivos [Burnout and stressors in health workers in intensive care units]. *Revista Psicopatol. Psicol. Clín.* 21, 129–136. doi: 10.5944/rppc.vol.21.num.2.2016.16146

Bandalos, D. L. (1993). Factors influencing cross-validation of confirmatory factor analysis models. *Multivariate Behav. Res.* 28, 351–374. doi: 10.1207/s15327906mbr2803_3

Barbero-García, M. I. (1993). *Psicometría II. Métodos de Elaboración de Escalas [Psychometrics II. Methods to Develop Scales].* Madrid: UNED.

Barbero-García, M. I., Vila-Abad, E., and Holgado-Tello, F. P. (2008). Tests Adaptation in cross-cultural comparative studies. *Acción Psicol.* 5, 7–16. doi: 10.5944/ap.5.2.454

Bendayan, R., Blanca, M. J., Fernández-Baena, J. F., Escobar, M., and Trianes, M. V. (2013). New empirical evidence on the validity of the Satisfaction with Life Scale in early adolescents. *Eur. J. Psychol. Assess.* 29, 36–43. doi: 10.1027/1015-5759/a000118

Biggs, D. M., Swailes, S., and Baker, S. (2016). The measurement of worker relations: the development of a three-component scale. *Leadersh. Organ. Dev. J.* 37, 2–12. doi: 10.1108/LODJ-08-2012-0098

Billings, J., and Raven, M. C. (2013). Dispelling an urban legend: frequent emergency department users have substantial burden of disease. *Health Aff.* 32, 2099–2108. doi: 10.1377/hlthaff.2012.1276

Borg Xuereb, C., Shaw, R. L., and Lane, D. A. (2016). Patients' and physicians' experiences of atrial fibrillation consultations and anticoagulation decision-making: a multi-perspective IPA design. *Psychol. Health* 31, 436–455. doi: 10.1080/08870446.2015.1116534

Bragard, I., Dupuis, G., and Fleet, R. (2015). Quality of work life, burnout, and stress in emergency department physicians: a qualitative review. *Eur. J. Emerg. Med.* 22, 227–234. doi: 10.1097/MEJ.0000000000000194

Brown, P. R., and Calnan, M. W. (2016). Chains of (dis)trust: exploring the underpinnings of knowledge sharing and quality care across mental health services. *Sociol. Health Illn.* 38, 286–305. doi: 10.1111/1467-9566.12369

Brown, S. P., and Leigh, T. W. (1996). A new look at psychological climate and its relationship to job involvement, effort, and performance. *J. Appl. Psychol.* 81:358. doi: 10.1037/0021-9010.81.4.358

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *Br. J. Psychol.* 3, 296–322.

Browne, M. W., Cudeck, R., Bollen, K. A., and Long, J. S. (1993). Alternative ways of assessing model fit. *Sage Focus Editions* 154, 136–136.

Byrne, B. M. (1998). *Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications, and Programming.* Mahwah, NJ: Lawrence Erlbaum Associates.

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56, 81–105. doi: 10.1037/h0046016

Carreño, S., Medina-Mora, M. A., Martínez, N., Juárez, F., and Vázquez, L. (2006). Características organizacionales, estrés y consumo de alcohol en trabajadores de una empresa textil mexicana [Organizational characteristics, stress and alcohol consumption in workers from a Mexican textile company]. *Salud Mental* 29, 63–70. Available online at: http://www.medigraphic.com/pdfs/salmen/sam-2006/sam064i.pdf

Carret, M. L., Fassa, A. G., and Kawachi, I. (2007). Demand for emergency health service: factors associated with inappropriate use. *BMC Health Serv. Res.* 7:131. doi: 10.1186/1472-6963-7-131

Castaneda, G. A., and Scanlan, J. M. (2015). Job satisfaction in nursing: A concept analysis. *Nurs. Forum* 9, 130–138. doi: 10.1111/nuf.12056

Catalyst (2007). *The Double-Bind Dilemma for Women in Leaderships: Damned If You Do, Doomed If You Don't.* New York, NY: Author.

Chacón-Moscoso, S., Sanduvete-Chaves, S., and Sánchez-Martín, M. (2016). The development of a checklist to enhance methodological quality in intervention programs. *Front. Psychol.* 7:1811. doi: 10.3389/fpsyg.2016.01811

Chakravarti, M., Laha, R. G., and Roy, J. (1967). *Handbook of Methods of Applied Statistics.* New York, NY: John Wiley & Sons.

Chiang, H. Y., Hsiao, Y. C., and Lee, H. F. (2017). Predictors of hospital nurses' safety practices: work environment, workload, job satisfaction, and error reporting. *J. Nurs. Care Qual.* 32, 359–368. doi: 10.1097/NCQ.0000000000000240

Dalkey, N., and Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Manage. Sci.* 9, 458–467.

Davenport, D. L., Henderson, W. G., Mosca, C. L., Khuri, S. F., and Mentzer, R. M. (2007). Risk-adjusted morbidity in teaching hospitals correlates with reported levels of communication and collaboration on surgical teams but not with scale measures of teamwork climate, safety climate, or working conditions. *J. Am. Coll. Surg.* 205, 778–784. doi: 10.1016/j.jamcollsurg.2007.07.039

Dawson, C., Veliziotis, M., and Hopkins, B. (2017). Temporary employment, job satisfaction and subjective well-being. *Econ. Ind. Democracy* 38, 69–98. doi: 10.1177/0143831X14559781

Delgado-Sánchez, A., Bellón, J. A., Martinez, M. T., Luna, J. D., López, L. A., and Lardelli, P. (2006). Las dimensiones del clima organizacional percibidas por los médicos de familia [Dimensions of organizational climate perceived by family physicians]. *Aten. Primaria* 37, 489–497. doi: 10.1157/13089089

Eagly, A. H., and Carli, L. L. (2007). Women and the labyrinth of leadership. *Harv. Bus. Rev.* 62–71. Available online at: https://www.researchgate.net/profile/Linda_Carli/publication/5957753_Women_and_the_labyrinth_of_leadership/links/0046351eeab13a8f9b000000.pdf

Eagly, A. H., Wood, W., and Johannesen-Schmidt, M. C. (2004). "Social roles theory of sex differences and similarities: implications for the partner preferences of women and men," in *The Psychology of Gender*, eds A. H. Eagly, A. Beall, and R. J. Sternberg (New York, NY: Guilford Press), 269–295.

Ehrhart, M. G., Aarons, G. A., and Farahnak, L. R. (2014). Assessing the organizational context for EBP implementation: the development and validity testing of the Implementation Climate Scale (ICS). *Implement. Sci.* 9:157. doi: 10.1186/s13012-014-0157-1

Eliseo, L. J., Murray, K. A., White, L. F., Dyer, S., Mitchell, P. A., and Fernandez, W. G. (2012). EMS providers' perceptions of safety climate and adherence to safe work practices. *Prehospital Emerg. Care* 16, 53–58. doi: 10.3109/10903127.2011.621043

Elmi, S., Hassankhani, H., Abdollahzadeh, F., Abadi, M. A., Scott, J., and Nahamin, M. (2017). Validity and reliability of the Persian practice environment scale of nursing work index. *Iran. J. Nurs. Midwifery Res.* 22, 106–111. doi: 10.4103/1735-9066.205953

Engelen, L., Dhillon, H. M., Chau, J. Y., Hespe, D., and Bauman, A. E. (2016). Do active design buildings change health behaviour and workplace perceptions? *Occup. Med.* 66, 408–411. doi: 10.1093/occmed/kqv213

Estryn-Behar, M., Doppia, M. A., Guetarni, K., Fry, C., Machet, G., Pelloux, P., et al. (2011). Emergency physicians accumulate more stress factors than other physicians–results from the French SESMAT study. *Emerg. Med. J.* 28, 397–410. doi: 10.1136/emj.2009.082594

Fernández, A. L., Rodríguez, D. S., Aranda, M. J. G., Paiz, A. L., and Saavedra, L. H. (2015). Estudio sobre las relaciones del síndrome de burnout con algunos factores psicosociales [Study on the relationships of burnout syndrome with some psychosocial factors]. *ReiDoCrea* 4, 94–99. Available online at: http://digibug.ugr.es/bitstream/10481/37009/1/ReiDoCrea-Vol.4-Art.13.pdf

Fernández-Argüelles, R., Cobos, P. A., and Figueroa, M. R. (2015). Evaluation of the organizational climate in a center for rehabilitation and special education. *Revista Cubana de Salud Pública* 41, 593–602.

Field, A. P. (2000). *Discovering Statistics Using SPSS for Windows: Advances Techniques for the Beginner*. London: Sage.

Fletcher, T. D., and Nusbaum, D. N. (2010). Development of the competitive work environment scale: a multidimensional climate construct. *Educ. Psychol. Meas.* 70, 105–124. doi: 10.1177/0013164409344492

García, A., Moro-Tejedor, M. N., and Medina-Torres, M. (2010). Evaluation and dimensions that define the labor environment and job satisfaction in nursing staff. *Revista de Calidad Asistencial* 25, 207–214. doi: 10.1016/j.cali.2010.02.003

George, D., and Mallery, P. (2003). *SPSS forWindows Step by Step: A Simple Guide and Reference*. Boston, MA: Allyn & Bacon.

Gill, P. J., Saunders, N., Gandhi, S., Gonzalez, A., Kurdyak, P., Vigod, S., et al. (2017). Emergency department as a first contact for mental health problems in children and youth. *J. Am. Acad. Child Adolesc. Psychiatry* 56, 475–482. doi: 10.1016/j.jaac.2017.03.012

Glück, J., König, S., Naschenweng, K., Redzanowski, U., Dorner, L., Straßer, I., et al. (2013). How to measure wisdom: Content, reliability, and validity of five measures. *Front. Psychol.* 4:405. doi: 10.3389/fpsyg.2013.00405

Godoi, V. C. G., Ganassin, G. S., Inoue, K. C., and de Moraes, N. L. (2016). Reception with risk classification: Characteristics of the demand in an emergency unit. *Cogitare Enferm* 21, 1–8. Available online at: http://www.saude.ufpr.br/portal/revistacogitare/wp-content/uploads/sites/28/2016/12/44664-186755-1-PB.pdf

Godoy, L., and Mladinic, A. (2009). Estereotipos y roles de género en la evaluación laboral y personal de hombres y mujeres en cargos de dirección [Gender stereotypes and roles in the work and personal evaluation of men and women in management positions]. *Psykhe* 18, 51–64. doi: 10.4067/S0718-22282009000200004

Gómez, C. A. (2004). Dise-o, construcción y validación de un instrumento que evalúa clima organizacional en empresas colombianas, desde la teoría de respuesta a los ítems [Design, construction and validation of an instrument that evaluates organizational climate in Colombian companies, from the item response theory]. *Acta Colomb. Psicol.* 11, 97–113. Available online at: http://editorial.ucatolica.edu.co/ojsucatolica/revistas_ucatolica/index.php/acta-colombiana-psicologia/article/view/482/479

González, A., Padilla, J. L., Hidalgo, M. D., Gómez-Benito, J., and Benítez, I. (2011). EASY-DIF software for analyzing differential item functioning using the Mantel-Haenszel and standardization procedures. *Appl. Psychol. Meas.* 35, 483–484. doi: 10.1177/0146621610381489

Hoe, S. L. (2008). Issues and procedures in adopting structural equation modeling technique. *J. Appl. Quant. Methods* 3, 76–83. Available online at: http://www.jaqm.ro/issues/volume-3,issue-1/pdfs/jaqm_vol3_issue1.pdf#page=81

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, M. I., and Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Qual. Quant.* 44, 153–166. doi: 10.1007/s11135-008-9190-y

Holgado-Tello, F. P., Chacón-Moscoso, S., Vila-Abad, E., Delgado, B., Sanduvete-Chaves, S., and Barbero-García, M. I. (2015). Evaluación de programas de formación continua en contextos no estandarizados: complementariedad entre Análisis Factorial y Multinivel para la obtención de evidencias de validez de constructo [Training program evaluation in non-standardized context: complementarity across factorial and multilevel analysis to obtain construct validity evidence]. *Anal. Psicol.* 31, 725–732. doi: 10.6018/analesps.31.2.172501

Hooper, C., Craig, J., Janvrin, D. R., Wetsel, M. A., and Reimels, E. (2010). Compassion satisfaction, burnout, and compassion fatigue among emergency nurses compared with nurses in other selected inpatient specialties. *J. Emerg. Nurs.* 36, 420–427. doi: 10.1016/j.jen.2009.11.027

Hooper, D., Coughlan, J., and Mullen, M. (2008). Structural equation modelling: guidelines for determining model fit. *Electron. J. Bus. Res. Methods* 6, 56–60. Available online at: http://www.sciencedirect.com/science/article/pii/S0099176709005534?via%3Dihub

Hunt, K. A., Weber, E. J., Showstack, J. A., Colby, D. C., and Callaham, M. L. (2006). Characteristics of frequent users of emergency departments. *Ann. Emerg. Med.* 48, 1–8. doi: 10.1016/j.annemergmed.2005.12.030

Insel, P., and Moos, R. H. (1974). Psychological environments: Expanding the scope of human ecology. *Am. Psychol.* 29, 179–188. doi: 10.1037/h0035994

International Test Commission (2005). *ITC Guidelines for Translating and Adapting Tests*. Available online at: https://www.intestcom.org/files/guideline_test_adaptation.pdf

Johnston, A., Abraham, L., Greenslade, J., Thom, O., Carlstrom, E., Wallis, M., et al. (2016). Review article: staff perception of the emergency department working environment: integrative review of the literature. *Emerg. Med. Australas.* 28, 7–26. doi: 10.1111/1742-6723.12522

Jöreskog, K. G. (2003). *Factor Analysis by MINRES. To the Memory of Harry Harman and Henry Kaiser*. Available online at: http://www.ssicentral.com/lisrel/techdocs/minres.pdf

Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W., and Butler, J. L. (2004). *Applied Linear Statistical Models*. Boston, MA: McGraw-Hill/Irwin.

Laposa, J. M., Alden, L. E., and Fullerton, L. M. (2003). Work stress and posttraumatic stress disorder in ED nurses/personnel (CE). *J. Emerg. Nurs.* 29, 23–28. doi: 10.1067/men.2003.7

Lee, S., Davis, K. D., Neuendorf, C., Grandey, A., Lam, C. B., and Almeida, D. M. (2016). Individual-and organization-level work-to-family spillover are uniquely associated with hotel managers' work exhaustion and satisfaction. *Front. Psychol.* 7:1180. doi: 10.3389/fpsyg.2016.01180

Leung, C. C., and Gomersall, C. D. (2016). "Intensive care unit triage," in *Quality Management in Intensive Care: A practical Guide*, eds H. Flaatten, B. Guidet, and A. Valentin (Cambridge, UK: Cambridge University Press), 77–84.

Li, C., Wu, Y., and Kee, Y. H. (2016). Validation of the volunteer motivation scale and its relations with work climate and intention among chinese volunteers. *Asian J. Soc. Psychol.* 19, 124–133. doi: 10.1111/ajsp.12127

Likert, R. (1967). *The Human Organization: Its Management and Value*. New York, NY: McGraw-Hill.

Litwin, G., and Stringer, R. (1968). *Motivation and Organizational Climate*. Boston, MA: Harvard Business School Press.

Lozano-Lozano, J. A., Chacón-Moscoso, S., Sanduvete-Chaves, S., and Pérez-Gil, J. A. (2013). Main components of organizational climate in an emergency medical service: a qualitative approach. *Acción Psicol.* 10, 101–114. doi: 10.5944/ap.10.2.12213

Mazzetti, G., Schaufeli, W. B., Guglielmi, D., and Depolo, M. (2016). Overwork climate scale: psychometric properties and relationships with working hard. *J. Manage. Psychol.* 31, 880–896. doi: 10.1108/JMP-03-2014-0100

Mendoza-Llanos, R. (2015). Job satisfaction and organizational culture as predictors of absenteeism. *Rev. Méd. Chil.* 143, 1028–1033. doi: 10.4067/S0034-98872015000800010

Meneghel, I., Borgogni, L., Miraglia, M., Salanova, M., and Martínez, I. M. (2016). From social context and resilience to performance through job satisfaction: a multilevel study over time. *Hum. Relations* 69, 2047–2067. doi: 10.1177/0018726716631808

Moos, R. H., and Insel, P. M. (1974). *The Work Environment Scale*. Palo Alto, CA: Consulting Psychologists Press.

Morata-Ramírez, M. A., and Holgado-Tello, F. P. (2013). Construct validity of Likert scales through confirmatory factor analysis: a simulation study comparing different methods of estimation based on Pearson and polychoric correlations. *Int. J. Soc. Sci. Stud.* 1, 54–61. doi: 10.11114/ijsss.v1i1.27

Morata-Ramírez, M. A., Holgado-Tello, F. P., Barbero-García, M. I., and Méndez, G. (2015). Análisis factorial confirmatorio: recomendaciones sobre mínimos cuadrados no ponderados en función del error Tipo I de Ji-Cuadrado y RMSEA [Confirmatory factor analysis: unweighted least squares recommendations as a function of Type I Chi-Squared error and RMSEA]. *Acción Psicol.* 12, 79–90. doi: 10.5944/ap.12.1.14362

Olsen, E., Bjaalid, G., and Mikkelsen, A. (2017). Work climate and the mediating role of workplace bullying related to job performance, job satisfaction, and work ability: a study among hospital nurses. *J. Adv. Nurs.* 31, 880–896. doi: 10.1111/jan.13337

Osterlind, S. J. (1998). *Constructing Test Items*. Boston, MA: Kluwer Academic Publishers.

Patterson, P. D., Huang, D. T., Fairbanks, R. J., and Wang, H. E. (2010). The emergency medical services safety attitudes questionnaire. *Am. J. Med. Qual.* 25, 109–115. doi: 10.1177/1062860609352106

Payne, R. L., and Mansfield, R. (1978). Correlates of individual perceptions of organizational climate. *J. Occup. Psychol.* 51, 209–218. doi: 10.1111/j.2044-8325.1978.tb00417.x

Pérez-Gil, J. A., Chacón-Moscoso, S., and Moreno, R. (2000). Validez de constructo: el uso del análisis factorial exploratorio - confirmatorio para obtener evidencias de validez [Construct validity: The use of exploratory - confirmatory factor analysis to obtain validity evidence]. *Psicothema* 2, 442–446. Available online at: http://www.psicothema.es/pdf/601.pdf

Perry, C., LeMay, N., Rodway, G., Tracy, A., and Galer, J. (2005). Validating a work group climate assessment tool for improving the performance of public health organizations. *Hum. Resour. Health* 3, 1–8. doi: 10.1186/1478-4491-3-10

Popoola, S. O. (2016). Personal factors, psychological work climate, and role related factors as precursors of organizational commitment of records management employees in state civil services of Northwest, Nigeria. *Afr. J. Bus. Manage.* 10, 318–327. Available online at: http://www.academicjournals.org/journal/AJBM/article-full-text/D5F099159857

Prieto, G., and Muñiz, J. (2000). Un modelo para evaluar la calidad de los test utilizados en Espa-a [A model to evaluate the quality of the tests used in Spain]. *Papeles Psicól.* 77, 65–75. Available online at: http://www.redalyc.org/pdf/778/77807709.pdf

Rama-Maceiras, P., Parente, S., and Kranke, P. (2012). Job satisfaction, stress and burnout in anaesthesia: relevant topics for anaesthesiologists and health care managers? *Eur. J. Anaesthesiol.* 29, 311–319. doi: 10.1097/EJA.0b013e328352816d

Rojas, M., Tirado, L. L., Pacheco, R. L., Escamilla, R., and López, M. (2011). Work climate in Mexican hemodialysis units. A cross sectional study. *Neurología* 21, 76–83. doi: 10.3265/Nefrologia.pre2010

Sabri, S. (2013). Item analysis of student comprehensive test for research in teaching beginner string ensemble using model based teaching among music students in public universities. *Int. J. Educ. Res.* 1, 1–14. Available online at: http://www.ijern.com/journal/December-2013/28.pdf

Safi, M. H., Torkaman-Nejad, S., and Arshi, S. (2016). Job satisfaction and its influencing factors among Shomal health center of Tehran personnel in 2014. *Commun. Health* 2, 64–71. Available online at: http://journals.sbmu.ac.ir/en-ch/article/view/9126

Sanduvete-Chaves, S., Chacón-Moscoso, S., Sánchez-Martín, M., and Pérez-Gil, J. A. (2013). The revised Osterlind index. A comparative analysis in content validity studies. *Acción Psicol.* 10, 3–10. doi: 10.5944/ap.10.2.11821

Schneider, B., Ehrhart, M. G., and Macey, W. H. (2013). Organizational climate and culture. *Annu. Rev. Psychol.* 64, 361–388. doi: 10.1146/annurev-psych-113011-143809

Schulz, H., Zacher, H., and Lippke, S. (2017). The importance of team health climate for health-related outcomes of white-collar workers. *Front. Psychol.* 8:74. doi: 10.3389/fpsyg.2017.00074

Sexton, J. B., Schwartz, S. P., Chadwick, W. A., Rehder, K. J., Bae, J., Bokovoy, J., et al. (2016). The associations between work–life balance behaviours, teamwork climate and safety climate: cross-sectional survey introducing the work–life climate scale, psychometric properties, benchmarking data and future directions. *BMJ Qual. Saf.* 6:44. doi: 10.1136/bmjqs-2016-006032

Spearman, C. C. (1910). Correlation calculated from faulty data. *Br. J. Psychol.* 3, 271–295. doi: 10.1111/j.2044-8295.1910.tb00206.x

Strauss, A., and Corbin, J. M. (1998). *Basic of Qualitative Research: Techniques and Procedures for Developing Grounded*. Thousand Oaks, CA: Sage.

Stringer, R. (2002). *Leadership and Organizational Climate*. Upper Saddle River, NJ: Prentice Hall.

Swiger, P. A., Raju, D., Breckenridge-Sproat, S., and Patrician, P. A. (2017). Adaptation of the practice environment scale for military nurses: a psychometric analysis. *J. Adv. Nurs.* 73, 2219–2236. doi: 10.1111/jan.13276

Tavakol, M., and Dennick, R. (2011). Making sense of Cronbach's alpha. *Int. J. Med. Educ.* 2, 53–55. doi: 10.5116/ijme.4dfb.8dfd

Taylor, J. A., Dominici, F., Agnew, J., Gerwin, D., Morlock, L., and Miller, M. R. (2011). Do nurse and patient injuries share common antecedents? An analysis of associations with safety climate and working conditions. *BMJ Qual. Saf.* 21, 101–111. doi: 10.1136/bmjqs-2011-000082

Torres, E., and Zegarra, S. J. (2015). Organizational climate and work performance in the Bolivarian educational institutions of the city Puno - 2014. *Comuni@cción* 6, 5–14. Available online at: http://www.redalyc.org/articulo.oa?id=449844870001

Uribe-Prado, J.-F., Patlan, J., and García, A. (2015). Psychosomatic manifestations, commitment and burnout as consequents from climate and organization culture: a path analysis. *Consultoría Administración* 60, 447–467. Available online at: http://www.cya.unam.mx/index.php/cya/article/view/720

Vedaa, Ø., Harris, A., Bjorvatn, B., Waage, S., Sivertsen, B., Tucker, P., et al. (2016). Systematic review of the relationship between quick returns in rotating shift work and health-related outcomes. *Ergonomics* 59, 1–14. doi: 10.1080/00140139.2015.1052020

Wankhade, P., and Mackway-Jones, K. (2015). *Ambulance Services: Leadership and Management Perspectives*. Cham: Springer.

Yela, M. (1957). *La Técnica del Análisis Factorial: Un Método de Investigación en Psicología y Pedagogía [The Factor Analysis Technique: A Research Method in Psychology and Pedagogy]*. Madrid: Biblioteca Nueva.

Zacher, H., and Yang, J. (2016). Organizational climate for successful aging. *Front. Psychol.* 7:1007. doi: 10.3389/fpsyg.2016.01007

Zadow, A. J., Dollard, M. F., McLinton, S. S., Lawrence, P., and Tuckey, M. R. (2017). Psychosocial safety climate, emotional exhaustion, and work injuries in healthcare workplaces. *Stress Health* 33, 558–569. doi: 10.1002/smi.2740

Zamanzadeh, V., Jasemi, M., Valizadeh, L., Keogh, B., and Taleghani, F. (2015). Effective factors in providing holistic care: a qualitative study. *Indian J. Palliat. Care* 21, 214–224. doi: 10.4103/0973-1075.156506

Zenteno, A. C., and Durán, C. A. (2016). Factores y prácticas de alto desempe-o que influyen en el clima laboral: análisis de un caso [High performance factors and practices that influence the work climate: a case analysis]. *Innovar* 26, 119–136. doi: 10.15446/innovar.v26n59.54367

Observation of Interactions in Adolescent Group Therapy: A Mixed Methods Study

Eulàlia Arias-Pujol[1]* and M. Teresa Anguera[2]

[1] FPCEE Blanquerna, Ramon Llull University, Barcelona, Spain, [2] Faculty of Psychology, University of Barcelona, Barcelona, Spain

Group psychotherapy is a useful clinical practice for adolescents with mental health issues. Groups typically consist of young people of similar ages but with different personalities, and this results in a complex communication network. The goal of group psychoanalytic psychotherapy is to improve participants' mentalization abilities, facilitating interactions between peers and their therapist in a safe, containing environment. The main aim of this study was to analyze conversation turn-taking between a lead therapist, a co-therapist, and six adolescents over the course of 24 treatment sessions divided into four blocks over 8 months. We employed a mixed-methods design based on systematic observation, which we consider to be a mixed method itself, as the qualitative data collected in the initial observation phase is transformed into quantitative data and subsequently interpreted qualitatively with the aid of clinical vignettes. The observational methodology design was nomothetic, follow-up, and multidimensional. The choice of methodology is justified as we used an *ad-hoc* observation instrument combining a field format and a category system. Interobserver agreement was analyzed quantitatively by Cohen's kappa using the free QSEQ5 software program. Once we had confirmed the reliability of the data, these were analyzed by polar coordinate analysis, which is a powerful data reduction technique that provides a vector representation of relationships between categories. The results show significant relationships between the therapist and (1) the activation of turn-taking by the participants and the co-therapist and silence and (2) conversation-facilitating interventions and interventions designed to improve mentalization abilities. Detailed analysis of questions demonstrating interest in others showed how the communication changed from radial interactions stemming from the therapist at the beginning of therapy to circular interactions half way through. Repetition was found to be a powerful conversation facilitator. The results also illustrate the role of the therapist, who (1) did not facilitate interventions by all participants equally, (2) encouraged turn-taking from more inhibited members of the group, (3) stimulated conversation from the early stages of therapy, and (4) favored mentalization toward the end. Despite its complexity, polar coordinate analysis produces easy-to-interpret results in the form of vector maps.

Keywords: group therapy, adolescent interactions, mixed-method, polar coordinates analysis, mentalization

INTRODUCTION

Peer groups are a natural setting for young people (Erikson, 1968). In the social context, Malekoff (2014) and Tellerman (2001) consider group work to be a protective factor for teenagers, pre-teenagers, and their families. In the field of public health, group psychotherapy is a useful clinical practice for adolescents with varying mental health issues (Reid and Kolvin, 1993; Cramer-Azima, 2002). Adolescent mental health disorders have increased over the last three decades (Nuffield Foundation, 2013) and today's teenagers have higher rates of anxiety, behavioral problems, and mood disorders (Merikangas et al., 2010).

Little has been published on group therapy in children or adolescents. Most of the studies conducted to date have reported on brief cognitive-behavioral interventions with specified diagnostic populations (Pollock and Kymissis, 2001). There has also been research into group counseling and psychotherapy with children and adolescents that indicates that the peer feedback that occurs in such settings is a key part of the process of change (Shechtman, 2007). The theoretical orientation behind this study was a combination of interpersonal and psychodynamic theories. Pingitore (2016) validated the benefits of interpersonal group therapy, an approach originally proposed by Yalom (2005), by quantitatively analyzing audio recordings of interventions by eight adolescents who took part in a process-oriented psychotherapy group for 3 months. Within a Kleinian psychoanalytic framework and following the contributions of Devi and Fenn (2012) published a systematic thematic analysis of a latency-aged children's group. Through clinical extracts, they showed how the children shifted from paranoid-schizoid functioning to depressive functioning over the course of therapy. They concluded that psychotherapy was beneficial in latency-aged children, as it provided them with the opportunity to observe and try to attach meaning to the interactions of other people, to respond to these interactions, to initiate contact and to help and be helped in a safe environment. Such experiences improve individuals' ability to recognize and observe mental states in both themselves and others and to develop empathy.

More research has been conducted in adults. A recent meta-analysis of group psychotherapy for social anxiety disorders concluded that group interventions were as effective as individual psychotherapy or pharmacotherapy (Barkowski et al., 2016). Group therapy is also beneficial for adults with moderate or severe depression (Pylvänäinen et al., 2015) or eating disorders (Simpson et al., 2010) and it has been shown to reduce symptoms of anxiety, depression, and avoidance in adults with personality disorders (Skewes et al., 2015). Schwartze et al. (2016) recently published a meta-analysis that showed that cognitive behavioral therapy was effective for patients with obsessive-compulsive disorder. Another randomized controlled study that compared the outcomes of short- and long-term psychodynamic psychotherapy (90-min weekly sessions for 20 or 80 weeks) in 167 adult outpatients with mood, anxiety, and personality disorders found that patients in both groups made significant gains, and concluded that short- and long-term therapy seemed equally effective for typical outpatients

seeking group psychotherapy, with the exception of symptomatic distress, for which a more favorable treatment effect was found for the long-term therapy (Lorentzen et al., 2013). A recent open prospective controlled study showed the efficacy of short-term dynamic group psychotherapy (37–39 sessions lasting 75 min over 9 months) in primary care patients with depressive symptoms (Bros et al., 2016).

In pyschotherapy research, there is growing concern for integrating qualitative methods, which provide a more holistic view of the person, and quantitative methods, which seek to provide a more objective view (Lutz and Hill, 2009). Despite the dearth of publications in the last decade, there are encouraging signs of a growing interest in the use of mixed-methods research in psychology (Roberts and Povee, 2014). By integrating complementary perspectives derived from quantitative and qualitative methods and analyses, mixed-methods research offers both rigor and flexibility and is likely to see an increase in future years (Anguera and Hernández-Mendo, 2016).

In this article, we describe the results of a study based on systematic observation, which we consider to be a mixed method in itself (Anguera and Hernández-Mendo, 2016). The study consisted of systematically observing video-recordings of adolescent group therapy sessions over a period of several months. The observation produced a large set of qualitative conversational data, subsequently analyzed quantitatively via polar coordinate analysis to detect changes in behaviors over the course of therapy.

The aim of the group therapy analyzed was to promote autonomy and maturity through interactions between peers and their therapist in a safe, containing environment (Torras de Beà, 2013). Group sessions of this type produce complex communication networks. Participants are typically young people of similar ages with different personalities who have difficulty relating to others and often perform poorly at school.

Psychodynamic interventions have been described as "conversation therapies," as the relationship between the person seeking treatment and the therapist forms the basis of the therapy (Malmberg and Fenton, 2008). We studied group communication as a conversation in which we analyzed turn-taking (who) and content (what).

Foulkes (1986) described two roles for group analysis leaders, or conductors: a role as dynamic-administrator and a role as analyst-interpreter. The function of the first is to set up the group, establish norms and boundaries, and create a safe, supportive, and containing environment designed to increase participation, expressiveness, and interaction and communication. The function of the second, by contrast, is related to mental activity, and consists of observation, listening, and understanding, and the ability to put into words everything they are understanding.

In the group studied, interventions by a therapist largely seek to (a) facilitate conversation and (b) promote mentalization, i.e., stimulate thought, reflection, and understanding about oneself and one's relationships with others.

In the *ad-hoc* observation instrument used in the study, we labeled this first group of interventions DYN, as they have a dynamic, stimulating function. They are interventions in

the form os a request or question in which the emitter (the therapist or participants) show interest in the life of the receiver. Demonstrating interest in others by asking questions, allowing them to intervene, and showing curiosity in their answers is considered to be a specific benefit of group therapy as opposed to individual therapy (Yalom, 2005; Torras de Beà, 2013). In previous studies, we saw that DYN interventions were very common in all sessions and that over the course of therapy, their use increased among participants and decreased among therapists, forming significant behavioral patterns (Arias and Anguera, 2004, 2005; Arias, 2011).

The second group of interventions in the observation instrument was called MNT to reflect the concept of mentalization described by Fonagy et al. (Fonagy, 1991; Fonagy et al., 1995), which is understood as the ability to explain and give meaning to one's own behaviors and those of others within a process of mental representation, thoughts, desires, and expectations. This ability is not innate: it needs to be developed within a safe, affective environment, which in psychoanalytic group psychotherapy is achieved by maintaining a stable internal and external setting while containing anxieties. MNT interventions are part of the therapist's role (Bateman and Fonagy, 2012), while DYN interventions correspond to either the therapist or the participants over the course of the sessions.

At the beginning of these group sessions, communication is generally radial, i.e., it diverges outwards toward the participants from the formal leader of the group, the therapist. With time, it becomes circular, with participants spontaneously intervening and demonstrating interest in each other. This shift in the direction of communication is an indicator of the group process, and our aim was to objectively analyze this process by studying the therapist's interventions.

The main aim of this study was to apply polar coordinate analysis to analyze conversation turn-taking and DYN and MNT interventions in a group therapy program involving a lead therapist, a co-therapist, and six adolescents. The program consisted of 24 group sessions, divided into four blocks, held over a period of 8 months.

MATERIALS AND METHODS

Design

In this mixed-methods study, we applied systematic observation, which meets the rigorous standards of scientific inquiry while at the same time offers the flexibility needed in real-life settings. Observational methodology permits the capture of spontaneous behaviors as they occur in a natural environment (Sackett, 1978; Anguera, 1979, 2003; Bakeman and Quera, 1995b, 2011; Portell et al., 2015a,b). It is thus an ideal methodology for studying communication in group therapy, and has proven to be suitable for studying the changes that occur over the course of therapy (Pascual-Leone et al., 2009).

There are eight possible study designs in observational methodology (Blanco-Villaseñor et al., 2003; Sánchez-Algarra and Anguera, 2013). The design used in this study was N/F/M (nomothetic/follow-up/multidimensional). It was *nomothetic* because we conducted a parallel analysis of the therapist,

the co-therapist, and six adolescents, *follow-up* because we performed both intersessional analyses (24 successive sessions) and intrasessional analyses (sequential recording of all behaviors from the start to finish of each session), and multidimensional because the *ad-hoc* observation instrument contained various dimensions selected on the basis of the theoretical framework and our experience.

The systematic observation was non-participative and the behaviors were highly perceivable.

Participants

There were eight participants: the therapist (T), the co-therapist (coT), and six adolescents (G, D, JM, F, L, M). The adolescents (four boys and two girls) had requested support at the Center for Child and Adolescent Mental Health of the *Eulàlia Torras de Beà* Foundation in Barcelona, Spain. They all had difficulties relating to others and difficulties learning at school; they had normal or normal-low intelligence according to the Weschler Intelligence Scale for Children–Fourth Edition (WISC-IV, Weschler, 2006). Two had a mild behavioral disorder, three had anxiety problems, and one tended to disconnect (**Table 1**, codes ICD-9-CM, Ministerio de Sanidad, Servicios Sociales e Igualdad, 2014).

The inclusion criteria were (a) an age of 12–15 years and (b) recommendation for group therapy following diagnostic evaluation at the Mental Health Center. The exclusion criteria were (a) anticipated difficulty attending all the therapy sessions and (b) contraindication for group therapy.

The group was led by an expert therapist, assisted by a co-therapist who participated as an observer. Both were clinical psychologists trained in group psychoanalytic psychotherapy.

In accordance with the principles of the Declaration of Helsinki and the Ethical Code of the General Council of the Official College of Psychologists of Spain, the participants were informed that they were being filmed. They were shown the location of the video cameras, which were positioned discretely to minimize reactivity bias. Informed consent was also obtained from the parents of the minors.

Instruments

In systematic observation (Anguera, 2003; Sánchez-Algarra and Anguera, 2013), a distinction is made between recording instruments (i.e., those used to record or code data) and observation instruments (purposed-designed instruments to analyze a given subject).

TABLE 1 | Patient characteristics.

Pseudonym	Age	Gender	Total IQ	Diagnosis (ICD-9-CM CODES)
Gabriel	14	Male	111	313.83/315.02/313.81.1
Danny	14	Male	110	309.23/313.0
John M.	14	Male	92	309.23/300.00.1/301.4.01
Fred	13	Male	90	309.23/297.3
Lucy	15	Female	84	309.23/315.5/313.81.1
Meg	13	Female	110	309.23/300.21/300.2

Pseudonyms have been used to protect confidentiality.

TABLE 2 | Dimensions and category systems in the observation instrument for therapists and patients.

Dimensions	Category systems	Description
DYN Facilitating of conversation	DYN = {FF, FO, RP, RT, QA, QC, QV}	*Facilitating of conversation*. Suitable questions or requests to start or enhance dialogue; routines such as greetings and other conversational rituals; requests for clarification; verification questions; full or partial repetitions of a previous intervention in the form of a statement or a question; vocalizations indicating that the communication channel is still open. FF = Phatic function. Vocalization indicating that the communication channel is still open. It indicates continued attention and cooperation, without the addition of new information. Typical vocalizations are "hmmn," "hum," or "aha." FO = Conversational routines or rituals, such as greetings or expressions of gratitude. RP = Total or partial reproduction of a previous utterance in the form of a statement not a question. This could be an answer to a request for clarification or it could have a phatic function, such as, for example, when the speaker simply echoes what a person has just said, indirectly encouraging them to continue. RT = Bringing back a topic of conversation. Intervention in which a participant brings back a subject previously brought up by another participant after a change of subject (CT) or interruption, thus making sure it is not forgotten. QA = Expressive question. Request, question, or series of adequate questions to start or promote dialogue and keep the main topic of conversation flowing. The person gives the turn to another person and shows interest in them. QC = Clarifying question. Question asking for clarification about what is happening. The person intervenes to clarify their own confusion and/or surprise in the form of a question. The speaker asks about a particular topic, doubt, or puzzlement, or about expressions, gestures, noises, or laughter he/she has not understood. It is a strategy used by the therapist when the adolescents are "doing their own thing." QV = Repetition of a previous statement in the form of a question. It is used to confirm what has just been said. It has a phatic function, as the speaker is conveying that the communication channel is still open. It can also be a strategy to emphasize a particular word or intervention.
MNT	MNT = {MNT}	*Reflective function*. Interventions focused on promoting thought, reflection, and understanding of oneself and one's relationships with others. They seek to stimulate the ability to understand what is happening in the minds of others. They are used by the therapist and can be directed at an individual or at the group as a whole. They include emphatic interventions, which put words to other participants' feelings.
EXP	EXP = {RA EC CD RB}	*Expressivity*. Interventions and answers manifesting the thoughts and/or feelings of the person speaking, adding new content to the conversation; short answers; sequences of words that continue the main subject of conversation; interventions that revisit subjects already dealt with.
DEF	DEF = {RD_N_P CT}	*Defensive expressions*. Interventions in which the participant avoids answering a previous question; verbalizations expressing the opposite of what has been said or done; projection of conflicts onto others; changing of subject.
DIS	DIS = {ED PD}	*Dislike*. Interventions expressing dislike, disagreement, distaste, or defiance.
ORD	ORD ={ORD}	*Ordering*. Prescriptive verbalizations, authoritarian demands (including exclamations).
HUM	HUM = {R EO}	*Humor*. Interventions with a clearly ironic/wry intention, jokes, jibes, and laughter.
CFR	CFR = {CFR}	*Confrontation*. Verbal interventions used by participants to express what they feel is happening in the group or see in some of their peers. They mirror the behavior of another. Peer feedback.
EX	EX = {EX}	*Exclamation*. Onomatopoeic word or words indicating a strong emotion of surprise, joy, or sadness.
S4	S4	*Degradation of vocal behavior*. Failed spontaneous interventions, interventions that progressively become weaker, abandoning of turn.
WHI	WHI = {S5}	*Whispering*. Talking in a low voice, with the intention of being heard by only a few people, establishing complicity. It leads to confused murmuring.
TO	TO = {TO}	*Touching*. Intentional physical contact with another person.
NOI	NOI = {MO S2 S3}	*Noise*. Noise or noises produced by a person, through their body (e.g., sneezing, burping, clapping), interaction with an object (e.g., chair, table, wall), or movement.
S1	S1	*Surrounding noise*. Sounds from outside the therapy room that are loud enough to be clearly heard.
Q	Q = {Q}	*Silence*. No words. Indicates no behavior.

Recording Instrument

The group sessions were recorded using two video cameras, two microphones, two video units, and two screens comprising a closed-circuit television system. The dataset was built in the software program GSEQ5, v.5.1 (Bakeman and Quera, 2011) using an initial transcription of the video content. In accordance with the principles of the Declaration of Helsinki and the Ethical Code of the General Council of the Official College of Psychologists of Spain, the participants were informed that they were being filmed. They were shown the location of the video cameras, which were positioned discretely to minimize reactivity bias.

According to the terminology proposed by Bakeman (1978), the data recorded were type II data, i.e., they were concurrent (as we considered various dimensions and each behavior needs to be coded using a specific code) and event-based (as the behaviors were coded as they occurred, thereby providing information on order and sequence, two essential factors for our study). It is also possible to record duration, but this was not relevant to the purpose of our study. Once annotated, each behavior generates a co-occurrence of codes (corresponding to the different dimensions) and is methodologcally considered to be a multievent (Bakeman, 1978). A total of 30,436 multievents were coded in our study.

Observation Instrument

The *ad-hoc* observation instrument used in the study combined a field format and category systems. It is a flexible instrument in which the different dimensions considered can be broken down into different categories according to the theoretical framework and experience. Considering the specific goals of the study and based on previous experiences (Arias and Anguera, 2004, 2005), the observation instrument was redesigned to include 15 forms of communication. These forms, or dimensions of communication, were derived from the work of Torras de Beà (2013) on group psychotherapy and of Tusón (1995) and Calsamiglia and Tusón (1999) on conversation analysis.

The 15 dimensions included in the observation instrument are Facilitating of conversation, Reflective function, Expressivity, Defensive expressions, Dislike, Ordering, Humor, Confrontation, Exclamation, Degradation of vocal behavior, Whispering, Touching, Noise, Surrounding noise, and Silence (**Table 2**). Each of these dimensions was broken down to build a category system that fulfilled the requirements of exhaustivity and mutual exclusivity (Anguera, 2003).

It should be noted that some dimensions gave rise to a single category, but given their conceptual relevance, we considered it important to include them as dimensions in the instrument. The dimensions and categories are shown in **Table 2**.

Procedure

The parents of the six adolescents were notified that their children had been proposed for group therapy after a diagnostic evaluation period. In addition, they all agreed to participate in a parallel group led by another therapist.

All the sessions were video-recorded and transcribed in full. Thirty sessions were held but due to technical difficulties with the recording, six were discarded because of poor audio. Therefore, 24 sessions were included in the final analysis. Each of the sessions lasted an hour. The sessions were grouped into four periods spanning an 8-month period.

Data Quality Control Analysis: Inter-observer Agreement

For the data quality control analysis, two observers analyzed and coded four of the therapy sessions. They had been previously

TABLE 3 | Polar coordinate analysis results corresponding to interventions by the therapist (T) as the focal behavior and interventions by the participants (G D JM F L M), interventions by the co-therapist (CT), and silence as conditional behaviors.

Category	Quadrant	Prospective perspective	Retrospective perspective	Ratio	Radius	Angle
BLOCK 1						
CT	I	1.96	3.64	0.88	4.13 (*)	61.78
G	III	−3.59	−5.52	−0.84	6.58 (*)	236.98
D	I	4.03	2.57	0.54	4.78 (*)	32.52
JM	I	1.27	2.86	0.91	3.13 (*)	66.01
F	I	3.69	6.02	0.85	7.06 (*)	58.45
L	III	−4.54	−5.7	−0.78	7.28 (*)	231.49
M	III	−4.51	−2.93	−0.54	5.38 (*)	212.99
Q	I	3.01	3.48	0.76	4.61 (*)	49.13
BLOCK 2						
CT	I	4.81	4	0.64	6.26 (*)	39.74
G	III	−7.33	−5.58	−0.61	9.21 (*)	217.31
D	III	−6.62	−5.93	−0.67	8.89 (*)	221.85
JM	IV	0.07	−0.21	−0.95	0.22	288.21
F	I	3.8	0.86	0.22	3.89 (*)	12.74
L	IV	0.81	−1.06	−0.8	1.34	307.19
M	I	5.17	5.39	0.72	7.47 (*)	46.16
Q	I	5.15	5.4	0.72	7.46 (*)	46.39
BLOCK 3						
CT	I	8.63	7.43	0.65	11.38 (*)	40.71
G	III	−16.95	−16.15	−0.69	23.41 (*)	223.62
D	III	−19.21	−15.73	−0.63	24.83 (*)	219.32
JM	I	7.42	8.75	0.76	11.48 (*)	49.71
F	I	13.93	11.89	0.65	18.32 (*)	40.46
L	III	−2.08	−5.61	−0.94	5.99 (*)	249.64
M	III	−1.83	−3.04	−0.86	3.55 (*)	239.03
Q	I	6.48	6.39	0.7	9.1 (*)	44.59
BLOCK 4						
CT	I	8.36	7.87	0.69	11.48 (*)	43.26
G	III	−5.44	−4.19	−0.61	6.86 (*)	217.62
D	III	−11.09	−9.16	−0.64	14.38 (*)	219.56
JM	III	−2.66	−2.93	−0.74	3.96 (*)	227.78
F	I	6.57	3.78	0.5	7.58 (*)	29.92
L	III	−11.6	−14.96	−0.79	18.93 (*)	232.22
M	III	−5.97	−7.98	−0.8	9.97 (*)	233.19
Q	I	3.98	10.16	0.93	10.91 (*)	68.59

Significant relationships (p < 0.05) between the focal behavior and conditional behaviors.

trained using the approach described by Anguera (2003). Agreement was assessed quantitatively using Cohen's kappa statistic (Cohen, 1960, 1968) in GSEQ5 (version 5.1) following the recommendations of Bakeman and Quera (1995a,b, 2001, 2011). According to the criteria of Landis and Koch (1977), the level of agreement was "almost perfect", with kappa values ranging between 0.86 and 0.93 for all the sessions.

Data Analysis

Polar coordinate analysis was used to analyze DYN and MNT interventions in accordance with the study objective. Polar coordinate analysis is a commonly used quantitative analytical method in observational methodology that identifies the statistical relationship between a behavior of interest (referred to in polar coordinate analysis as the *focal behavior*) and other behaviors (referred to as *conditional behaviors*). Associations between pairs of behaviors are represented graphically by vectors. Polar coordinate analysis requires a prior stage consisting of lag sequential analysis (Bakeman, 1978, 1991), a technique used to reveal behavioral patterns based on occurrence of behaviors after (prospective) or before (retrospective) a given behavior

(as the focal behavior is known in lag sequential analysis). The technique is based on calculating conditional and unconditional probabilities (based, respectively, on matched frequencies and simple frequencies) for each of the time lags considered, which may be positive or negative.

Lag sequential analysis produces large volumes of data, which are subsequently reduced through a powerful data reduction algorithm based on the $Z_{sum} = \frac{\Sigma z}{\sqrt{n}}$ parameter proposed by Cochran (1954), where z is the standard value corresponding to each lag for each of the conditional behaviors (known as *target* behaviors) and n is the number of lags considered. The Z_{sum} is calculated for each target behavior for both positive lags (prospective Z_{sum}) and negative lags (retrospective Z_{sum}). The technique thus yields a statistical relationship between the given behavior and each of the target behaviors, which is reflected by a prospective and a retrospective Z_{sum} value, as proposed by Sackett (1980, 1987). To optimize the procedure, Anguera (1997) proposed a modification to the original technique (1980, 1987) based on the concept of genuine retrospectivity. This modified technique has been used on multiple occasions in the past two decades and was employed in the current study.
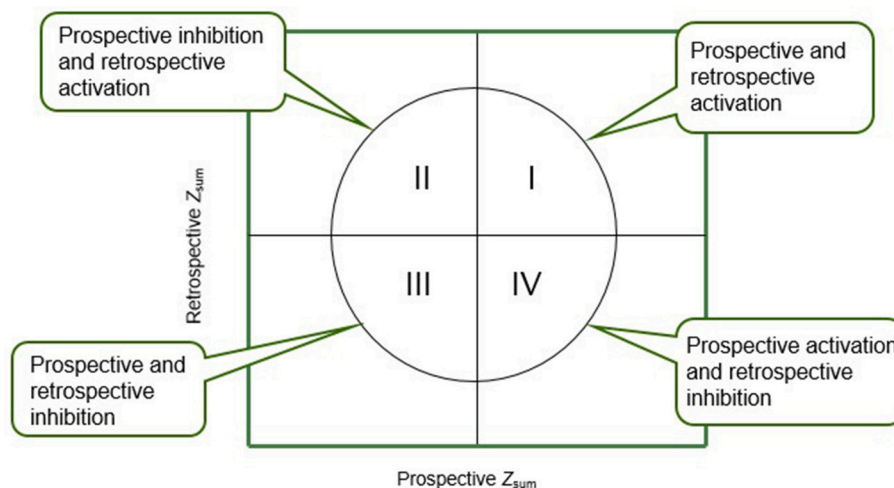
FIGURE 1 | Characteristics of the quadrants in which the vectors are located according to the activation (+) or inhibition(−) sign carried by the Prospective and Retrospective Zsum values.
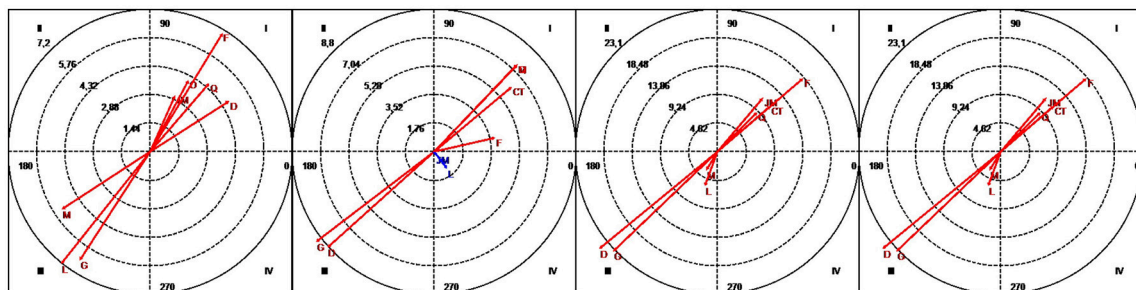
FIGURE 2 | Vectors corresponding to interventions by the therapist (T) as the focal behavior and interventions by the participants (G, D, JM, F L, and M), interventions by the co-therapist (coT), and silence (Q) as conditional behaviors. Session blocks 1-2-3-4 (from left to right).

Polar coordinate analysis integrates the prospective and retrospective perspectives with the help of a vectorial map that contains four quadrants in which the prospective and retrospective Z_{sum} values are plotted along the X and Y axis, respectively. Each target behavior analysis thus can be located in one of the four quadrants depending on the combination of negative/positive signs (**Table 3**).

Polar coordinate analysis uses the prospective and retrospective Z_{sum} values for each conditional behavior to calculate the length and angle of the corresponding vector, thus allowing these to be graphically represented. The length of the vector is $\sqrt{(Z^2_{sum}Prospective + Z^2_{sum}Retrospective)}$, and is considered to be statistically significant ($p < 0.05$) when it exceeds 1.96 The angle of the vector is calculated as follows: $\varphi = \text{arc sen} \frac{Z_{sum}Retrospective}{Length}$ and it is then adjusted according to the quadrant in which it is located: quadrant I ($0 < \varphi < 90$) $= \varphi$; quadrant II ($90 < \varphi < 180$) $= 180 - \varphi$; quadrant III ($180 < \varphi < 270$) $= 180 + \varphi$; quadrant IV ($270° < \varphi < 360°$) $= 360° - \varphi$.

The meanings of the different quadrants are shown in **Figure 1**.

Quadrants I and III are symmetrical in terms of the relationship they depict between the focal behavior and the different conditional behaviors they contain. Quadrant I (++) indicates mutual activation while quadrant III (−−) indicates mutual inhibition. Quadrants II and IV, in turn, depict asymmetrical relationships. Quadrant II (−+) indicates that the focal behavior inhibits but at the same time is activated by the conditional behaviors, while quadrant IV (+−) indicates the opposite (i.e., the focal behavior activates and is inhibited by the corresponding conditional behaviors).

The polar coordinate analysis for this study was performed in HOISAN v. 1.6.3.2 (Hernández-Mendo et al., 2012), which contains all the necessary modules and also produces partial results for adjusted residuals and z values in addition to analytical parameters and polar coordinate maps. The analysis was conducted by exporting the data file from GSEQ5 to HOISAN.

Polar coordinate analysis has been used in certain areas of clinical psychology, such as groups of children with autistic siblings (Venturella, 2016). It has also been widely applied in sports (Perea et al., 2012; Robles-Prieto et al., 2014; Echeazarra et al., 2015; López-López et al., 2015; Morillo-Baro et al., 2015; Sousa et al., 2015; Castañer et al., 2016, 2017; Aragón et al., 2017) and school settings (Herrero Nivela, 2000; Anguera et al., 2003; López et al., 2016; Santoyo et al., 2017). As a final note of interest, when Sackett (1980) first presented polar coordinate analysis, he used it to study turn-taking in conversation.

RESULTS AND DISCUSSION

In the sections below, we describe the relationships detected between interventions by the therapist and the group participants using polar coordinate analysis.

TABLE 4 | Polar coordinate analysis results with interventions by the therapist (T) as the focal behavior and DYN categories (broken down) and MNT as conditional behaviors.

Category	Quadrant	Prospective perspective	Retrospective perspective	Ratio	Radius	Angle
BLOCK 1						
QA	IV	2.36	−3.87	−0.85	4.54 (*)	301.4
QC	III	−4.61	−4.61	−0.71	6.52 (*)	224.98
FF	II	−0.42	1.94	0.98	1.99 (*)	102.32
FO	IV	0.11	−1.99	−1	2 (*)	273.02
RP	I	1.28	5.16	0.97	5.32 (*)	76.07
QV	III	−1.53	−1.59	−0.72	2.2 (*)	226.01
RT	II	−0.07	0.31	0.98	0.32	102.7
MNT	III	−3.91	−1.89	−0.44	4.34 (*)	205.83
BLOCK 2						
QA	IV	4.9	−1.81	−0.35	5.22 (*)	339.72
QC	III	−3.54	−2.49	−0.58	4.33 (*)	215.2
FF	I	1.36	3.06	0.91	3.35 (*)	66.01
FO	I	0.74	2.15	0.95	2.28 (*)	71.07
RP	I	4.83	6.78	0.81	8.33 (*)	54.5
QV	I	1.46	1.07	0.59	1.81	36.25
RT	II	−2.56	0.47	0.18	2.61 (*)	169.63
MNT	II	−2.88	1.96	0.56	3.48 (*)	145.78
BLOCK 3						
QA	L	14.08	6.32	0.41	15.43 (*)	24.17
QC	III	−1.11	−3.01	−0.94	3.21 (*)	249.85
FF	I	3.49	5.31	0.84	6.35 (*)	56.66
FO	I	3.61	4.11	0.75	5.47 (*)	48.65
RP	I	3.72	4.87	0.79	6.13 (*)	52.64
QV	I	5.39	3.69	0.56	6.54 (*)	34.38
RT	II	−2.11	3.14	0.83	3.78 (*)	123.86
MNT	II	−0.32	5.49	1	5.5 (*)	93.33
BLOCK 4						
QA	I	10.12	4.3	0.39	10.99 (*)	23.01
QC	IV	0.38	−1.51	−0.97	1.56	284.25
FF	I	4.96	6.75	0.81	8.38 (*)	53.72
FO	I	4.7	3.78	0.63	6.03 (*)	38.83
RP	I	4.59	5.54	0.77	7.2 (*)	50.39
QV	I	4	1.46	0.34	4.26 (*)	20.04
RT	II	−1	2.74	0.94	2.91 (*)	110.1
MNT	I	3.86	9.43	0.93	10.19 (*)	67.76

*Significant relationships (p < 0.05) between the focal behavior and conditional behaviors.

Relationships between Turn-Taking by the Therapist, Turn-Taking by the Participants and the Co-therapist, and Silence

The focal behavior was intervention by the therapist (T) and the conditional behaviors were interventions by the participants (G, D, JM, F L, and M), interventions by the co-therapist (coT), and silence (Q) in the four blocks of sessions spanning 8 weeks.

As shown in **Table 3**, the majority of results were significant.

The graphs in **Figure 2** show the vectors representing turn-taking by the participants and the co-therapist and silence. In the case of the adolescents, some of the vectors are located
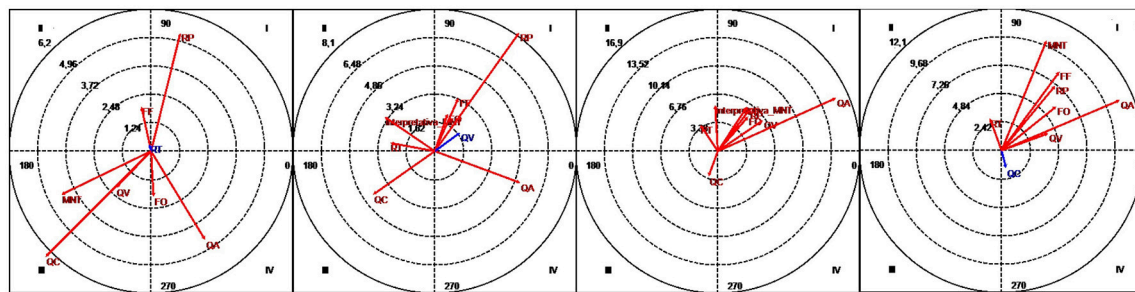
FIGURE 3 | Vectors corresponding to interventions by the therapist (T) as the focal behavior and conversationfacilitating DYN categories (FF, FO, RP, RT, QA, QC, QV) and the mentalizing or reflective function MNT category as conditional behaviors. Session blocks 1-2-3-4 (from left to right).

in the mutual inhibition quadrant (quadrant III) while others are located in the mutual activation quadrant (quadrant I). On analyzing the four blocks of sessions grouped by time, it can clearly be seen that the turn-taking behavior by D, L, and M changed over the course of therapy, that of the co-therapist and silence remained stable.

Relationship between the Therapist and DYN and MNT Interventions

Again, the focal behavior was intervention by the therapist (T) and the conditional behaviors were the DYN categories FF, FO, RP, RT, QA, QC, and QV and the MNT category.

The majority of results in this case were also significant (**Table 4**).

The graphs in **Figure 3** show the vectors for the different relationships distributed among the four quadrants. On examining the figures by blocks of time, it can be seen that the vectors tend to form clusters, with the majority located in the mutual activation quadrant (quadrant I) by the end of the therapy. Note that the length of the radius for repetition (RP) and the quadrant in which it was located (quadrant I) remained stable over the four periods.

Below we discuss the significance of the relationships detected by polar coordinate analysis in five sections. We also illustrate our findings with clinical vignettes containing coded transcripts of the interventions.

Turn-Taking by the Therapist and the Adolescents

All the significant results are located in two opposing quadrants, indicating two clearly differentiated types of relationship: mutual activation and mutual inhibition. The therapist always facilitates intervention by Fred, the participant with the greatest difficulty relating to others, and in the early phases of therapy, she also encourages interaction from Danny, John M, and Meg. Her interventions never activate those of the two impulsive participants, Gabriel and Lucy. This does not mean that she excludes these participants, simply that they intervene on their own initiative. The changes detected in Danny, John M, and Meg are an indication of the progress they make over the therapy.

TABLE 5 | Clinical vignette 1.

Vignette 1 (Block 1). Danny has been on a trip to a museum with his school.

T – It's a different museum, right? [QA]

D – Yes, it was an industry. [RA]

T – It was an industry; is it located in an old factory? [RP] [QA]

D – Yes, in a factory, they used an industry from the 1960s. [RA]

T – Hmmm... And you said that you had to do an assignment? [FF] [PA]

D – They gave us a sheet of paper and we had to fill it in. [RA]

T – With the things you were seeing and the explanations they were giving you? [QA]

D – Yes. [RB]

Block 1 is characterized by radial communication between the therapist and all the participants. Vignette 1 shows an example of an interaction between the therapist and Danny (**Table 5**).

However, not all interactions are the same. Gabriel and Lucy, for example, spontaneously take turns in these early sessions (**Table 6**).

Lucy raises conflicts about herself that interest everyone (**Table 7**).

John M is a reserved person with anxiety problems. He has difficulty intervening and when he does, he often mumbles, says very little, and adheres to what has just been said (**Table 8**).

Haen and Weil (2010) have highlighted the difficulties that adolescents have engaging during this initial stage of therapy. In our study, as the therapy progresses, the adolescents start to communicate much more naturally and spontaneously and bring up issues that concern them, such as going out, the end of the school year, and their expectations for the coming year. Vignette 5, which contains an excerpt from this last block, shows how Danny, John M, Lucy, and Meg chat freely amongst themselves, without encouragement from the therapist. Amidst jokes, exclamations, gestures, and laughter, they talk about meeting outside the group and about their fears of traveling alone on the train or underground for the first time (**Table 9**).

Turn-Taking by the Therapist and the Co-therapist

The co-therapist and the therapist was mutually activated (quadrant 1). The co-therapist's interventions reflect

TABLE 6 | Clinical vignette 2.

Vignette 2 (Block 1). The topic of conversation is about getting down to studying and passing and failing subjects

G – Yes, at the beginning you see it as far off, Well... that's what I think, and you do nothing. [RA]

T – Hmmm. [FF]

G – But then, when you see that you are getting bad marks, and that if you don't get your act together, well, they will fail you, then you study. [EC]

T – Is that the same with all of you? [QA]

L – For me it's the opposite. [RA]

T – Aha. [FF]

L – In the first, in the first term, well that was it, I had to study, and because I spent the summer studying..., I mean, I don't care, the truth is that it doesn't matter if it's at the beginning of the year or at the end [EC]

G – That's the bad thing, like she says, yes, because if you have to study in September, yuck! In my school, they do courses in July, right there, and I spend a month at school. They give you minimum goals and at the end of the course, they test you, you can do at least three...[EC]

L – Yeah, well imagine if you've got seven subjects left for the summer, for September. [EC]

TABLE 7 | Clinical vignette 3.

Vignette 3 (Block 1).

Lucy has just explained that she has been to different schools:

T – And now, how are you? (current school) [QA]

L – Fine, but I don't like it, I don't like any of the girls in my class. [RA]

T – What do you mean? What don't you like about the girls in your class? [QC] [QA]

L – That they're always saying I'm very childish because I don't wear make-up or show my thighs, I don't like that! [RA]

T – Hmmm [FF]

L – And they say I'm very childish because I'm 15 but I don't like wearing make-up or going off into corners kissng guys. I'm not into that, but that's what they appear to do. [EC]

T – Hmmm. [FF]

L – And when they ask me if I'm coming with them, I don't go. I'm not into that [EC]

T – Hmmm. What do the rest of you think about what Lucy is saying? [FF] [QA]

M – Good. [RB]

T – Good. What do you mean? [RP] [QC]

M – That... She will end up better than them, they're the ones going astray. [RA]

TABLE 8 | Clinical vignette 4.

Vignette 4 (Block 1).

The topic of conversation is about marks and exams. They have all explained how they are assessed. John M says nothing until the therapist asks him directly.

T – And what about you, John M? How are you assessed? [QA]

JM – Like her. [RA]

I suppose you're referring to Lucy, who has just spoken.

T – Exactly exactly like her? [QA]

JM (in a low voice)- Yes [RB]

TABLE 9 | Clinical vignette 5.

Vignette 5 (Block 4).

Lucy is explaining that she's going to be in a play in a village near the Mental Health Center. Meg asks her directly:

M (addressing L) – And you don't feel embarrassed? [QA]

L – Yes, and they say that they're going to throw eggs at us. [RA]

D – Jeez. [EE]

JM – Count me in. [EO]

D – You know what I mean, yahoo! One by one! (gestures of throwing eggs) [EO] [EE] [EO]

JM – Haha. [R]

L – I hope they're joking, because if not, they'll get in trouble. [CFR]

M (addressing L) – Can you get there by train? [QA]

L – Yes. [RB]

M (in a low voice) – Darn. [EE]

L – If you can get there by train? [RT]

D – I'll bring some hens, hahaha. [EO] [R]

JM – Let's go, yay! [EO] [EE]

M – You get there by train? [PV]

D – Yes! [RB]

L – Or you can go by car or... [EC]

D – There are tracks and a station, hahaha. [EO] [R]

M – Bah! I'm not going by train. [EE] [EC]

JM – Hee hee. [R]

JM – Hee hee hee. How are you going to go, on foot? Haha. [R] [EO] [R]

JM – Haha. [R]

M – Haha. No. [R] [RB]

M – No, because of what happens to her with the underground (referring to being afraid to ride alone) [EXP]

JM and D in unison – The same things happens to you with the train. [CFR]

M – No, because the first time I go on a train alone, well ...[DEF]

D – You'll get lost... [CFR]

M – No...[DEF]

her role of interfering as little as possible in the group dynamics. They complement those of the main therapist. Together, they form a team and create and maintain a safe environment (Shechtman 2007; Torras de Beà, 2013; Malekoff, 2014).

The Therapist and Silence

The therapist generates silence but also breaks it (quadrant 1).

The examples below show how the adolescents fall silent when faced with difficult issues, such as verbalizing why they are in the group or talking about their relationship with their parents or their concerns about sexuality (**Tables 10**–**12**).

The Therapist and DYN Interventions

The different strategies for facilitating conversation (FF, FO, RP, RT, QA, QC, and QV) showed varying patterns of change over the course of therapy but converged at the end.

Repetition (RP) was the most powerful strategy, as it activated conversation from the start of the therapy program. The next most powerful strategies were phatic function (FF) and greetings (FO). The transcripts of the sessions show that in the early sessions, it was the therapist who verbally greeted the adolescents (by saying hello and goodbye). However, few of them responded

TABLE 10 | Clinical vignette 6.

Vignette 6 (block 2). The therapist challenges the participants with questions, she takes them to a level of mentalization that they are not ready for yet and they become inhibited.

T – Why are we coming to the group? And why? We are all coming for something, aren't we? [MNT]

Silence. [Q]

T – Why do you think you are coming? How are we are trying to help you here? [MNT]

Silence. [Q]

T – Maybe we have to go over this again... [EXP]

TABLE 11 | Clinical vignette 7.

Vignette 7 (block 3). At another moment, silence allows the adolescents to express themselves with sincerity:

T – How would you like your parents to treat you? What do you expect? [MNT]

Silence. [Q]g

D – Them not to use such tough punishments [EXP]

T – Not to use such tough punishments [EXP]

G – They always use the worst possible punishments [EXP]

T – The worst? What does that mean, what you like most? [QV] [MNT]

G – Yes, they punish you with the things you like most. [EC]

T – And what happens then? How do you feel? [MNT]

G (in a very low voice) – Crap... [EXP]

Silence. [Q]

T – How do you all feel? Do you get discouraged? Do you feel that they are disheartening you? [MNT]

Silence. [Q]

TABLE 12 | Clinical vignette 8.

Vignette 8 (Block 3)

This silence expresses the difficulty talking about sexuality.

T – Maybe you talk about condoms at school, do you? Amongst ourselves too, right? [QA]

Silence. [Q]

T – No? [PV]

JM – Haha. [R]

G – Haha. [R]

T – Jokingly, jokingly, it makes you laugh. I think that it is, that it's something that's talked about at school, about their use, right? [MNT] [QA]

Silence. [Q]

T – You're all a little quiet, aren't you? Eh? What do you think about condoms? Do you know anything? Do you talk about them with each? [MNT] [A]

Pause. [Q]

T – Before you were talking about AIDS, somebody said this word like with a lot of disgust, about the risk of infection ...[MNT]

G burps and covers his mouth, mumbles something to D that I don't understand. [NOI] [S4]

D – Brrr. [EE]

G – But blood doesn't have to come out to get an infection. [EXP]

and the others returned the greeting or made a non-verbal gesture. This behavior changes after the first block, indicating an increase in reciprocity between the therapist and the participants.

The appearance of QA (questions directed at others) in the second half of the therapy is, in our opinion, a highly significant indicator of the group process. It tells us that the communication is no longer radial and that the adolescents have achieved one of the most important benefits of group therapy, which is showing interest in others (Yalom, 2005) in the presence of the therapist (Torras de Beà, 2013).

It is also interesting to see how QV (repetition of a previous utterance in the form of a question) changes from being mutually inhibitory to being mutually activating. We think that this strategy initially surprised the adolescents but was then gradually adopted by them. The same was not observed for QC (clarifying questions), which were used only by the therapist when the adolescents were "doing their own thing" and she was "excluded" from the group. Examples of what she said were: "I'm not quite following you now...maybe I'm being a bit dense, can you help me understand what's going on?" This strategy is similar to the attitude of respectful curiosity shown by therapists in the Adolescent Mentalization-Based Integrative Treatment (AMBIT) approach (mentalizing stance), which is designed to help put a halt to non-mentalization mental states (Benvington et al., 2012; Dangerfield, 2016).

Bringing back a central topic of conversation (RT) and suggesting looking at this in greater depth was only done by the therapist.

At the end of therapy, all the categories in the DYN dimension except RT are located in the mutually activating quadrant. This supports the idea that the communication strategies used by the therapist were adopted by the participants, enabling them to talk more autonomously and facilitating their personal growth (Yalom, 2005; Torras de Beà, 2013).

The Therapist and MNT Interventions

The changes observed in the MNT category, which corresponds to interventions aimed at improving the adolescents' mentalization abilities, also reflect interesting aspects of the group process. The MNT category changed from inhibitory (quadrant III) to partially inhibitory (quadrant II) and finally to mutually activating (quadrant I). The changes also show that the therapist's role changed over time, as mentalization strategies were only used by her. We can deduce that the participants gradually overcame their early inhibitions and dependence and acquired more sophisticated mentalizing abilities, helping them to become more aware of themselves and of others. This result is consistent with the concept known as the interpretative function of the therapist within the theories of Foulkes (1986) and Torras de Beà (2013)

CONCLUSIONS

Polar coordinate analysis provides a new approach for gaining insights into dialogue in group pyschotherapy. The results show that the technique provides a novel means of analyzing the role of the therapist and describing her conversational style. The therapist proved to be an expert in creating a communicative environment that allowed the adolescents to

grow. She employed four core strategies: (1) she did not facilitate communication equally for all participants, (2) she encouraged turn-taking by the more inhibited members of the group, (3) she stimulated conversation from the early stages of therapy, and (4) she promoted mentalization toward the end of therapy.

We were particularly pleased to see that the use of repetition (RP) facilitated communication flows from the beginning. The positive results indicate that rather than simply acting as an echo or a loudspeaker, this strategy produces a mirroring effect similar to that described in the social biofeedback theory of parental affect-mirring (Gergely and Watson, 1996), in which the person talking, apart from being listened to, is brought into a mirror-like interaction. This regulatory effect is a prerequisite for the mentalization process that facilitates the development of the self (Fonagy et al., 2002).

Observational methodology and polar coordinate analysis could prove to be of great value for detecting changes in psychotherapy models based on spoken conversation.

AUTHOR CONTRIBUTIONS

EA developed the project. MA performed the method section and polar coordinate analysis. Both authors have participated in the writing of the article.

REFERENCES

Anguera, M. T. (1979). Observational typology. *Qual. Quant.* 13, 449–484.

Anguera, M. T. (1997). *From Prospective Patterns in Behavior to Joint Analysis with a Retrospective Perspective. En Colloque sur Invitation Ân̂ Méthodologie D'analyse des Interactions SocialesÂž.* Paris: Université de la Sorbonne.

Anguera, M. T. (2003). "Observational methods (general)," in *Encyclopedia of Psychological Assessment, Vol. 2*, ed R. Fernández-Ballesteros (London: Sage), 632–637.

Anguera, M. T., and Hernández-Mendo, A. (2016). Avances en estudios observacionales de ciencias del deporte desde los mixed methods [Advances in mixed methods observational studies in sports sciences]. *Cuad. Psicol. Dep.* 16, 17–30.

Anguera, M. T., Santoyo, C., and Espinosa, M. C. (2003). "Evaluating links intensity in social networks in a school context through observational designs," in *Culture, Environmental Action and Sustainability*, eds R. García Mira, J. M. Sabucedo Cameselle, and J. Romay Martínez (Göttingen: Hogrefe and Huber), 286–298.

Aragón, S., Lapresa, D., Arana, J., Anguera, M. T., and Garzón, B. (2017). An example of the informative potential of polar coordinate analysis: sprint tactics in elite 1,500-m track events. *Meas. Phys. Educ. Exerc. Sci.* doi: 10.1080/1091367X.2016.1245192. [Epub ahead of print].

Arias, E. (2011). Metodología observacional e investigación del proceso grupal. [Observational Methodology and Group Process Research]. *Teoría Práct. Grupoanalítica* 1, 435–448.

Arias, E., and Anguera, M. T. (2004). Detección de patrones de conducta comunicativa en un grupo terapéutico de adolescentes [Detection of communicative behavior patterns in a therapeutic group of adolescents]. *Acción Psicol.* 3, 199–206. doi: 10.5944/ap.3.3.513

Arias, E., and Anguera, M. T. (2005). Análisis de la comunicación en un grupo terapéutico de adolescentes: estudio diacrónico [Analysis of Communication in a Therapeutic Group of Adolescents: A Diachronic Study]. *Rev. Psicopatol. Salud Ment. Niño Adolesc.* M1, 25–36.

Bakeman, R. (1978). "Untangling streams of behavior: Sequential analysis of observation data," in *Observing Behavior, Vol. 2: Data Collection and Analysis Methods*, ed G. P. Sackett (Baltimore, MD: University of Park Press), 63–78.

Bakeman, R. (1991). From lags to logs: advances in sequential analysis. *Mex. J. Behav. Anal.* 17, 65–83.

Bakeman, R., and Quera, V. (1995a). Log-lineal approaches to lag-sequential analysis when consecutive codes may cannot repeat. *Psychol. Bull.* 118, 272–284. doi: 10.1037/0033-2909.118.2.272

Bakeman, R., and Quera, V. (1995b). *Analyzing Interaction: Sequential Analysis with SDIS and GSEQ.* Cambridge: Cambridge University Press.

Bakeman, R., and Quera, V. (2001). Using GSEQ with SPSS. *Metod. C Comport.* 3, 195–214.

Bakeman, R., and Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences.* Cambridge: Cambridge University Press.

Barkowski, S., Schwartze, D., Strauss, B., Burlingame, G. M., Barth, J., and Rosendahl, J. (2016). Efficacy of group psychotherapy for social anxiety disorder: a meta-analysis of randomized-controlled trials. *J. Anxiety Disord.* 39, 44–64. doi: 10.1016/j.janxdis.2016.02.005

Bateman, A., and Fonagy, P. (ed.) (2012). *Handbook of Mentalizing in Mental Health Practice.* London: American Psychiatric Publishing.

Benvington, D., Fuggle, P., Fonagy, P., Target, M., and Asen, E. (2012). Innovations in practice: Adolescent Mentalization-Based Integrative Therapy (AMBIT) – a new integrated approach to working with the most hard to reach adolescents with severe complex mental health needs. *Child Adolesc. Ment. Health* 18, 46–51. doi: 10.1111/j.1475-3588.2012.00666.x

Blanco-Villaseñor, A., Losada, J. L., and Anguera, M. T. (2003). Analytic techniques in observational designs in environment-behavior relation. *Medio Ambient. Comport. Hum. Rev. Int. Psicol. Ambient.* 4, 111–126.

Bros, I., Notó, P., and Bulbena, A. (2016). Effectiveness of short-term dynamic group psychotherapy in primary care for patients with depressive symptoms. *Clin. Psychol. Psychother.* doi: 10.1002/cpp.2029. [Epub ahead of print].

Calsamiglia, H., and Tusón, A. (1999). *Las Cosas del Decir [Things about saying].* Barcelona: Ariel.

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Canton, A., and Hileno, R. (2016). Goal scoring in soccer: a polar coordinate analysis of motor skills used by lionel messi. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Fernandes, T., and Hileno, R. (2017). Mastery in goal scoring, T-pattern detection and polar coordinate analysis of motor skills used by lionel messi and cristiano ronaldo. *Front. Psychol.* 8:741. doi: 10.3389/fpsyg.2017.00741

Cochran, W. G. (1954). Some methods for streghning the common χ^2 test. *Biometrics* 10, 417–451. doi: 10.2307/3001616

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement of partial credit. *Psychol. Bull.* 70, 213–220. doi: 10.1037/h0026256

Cramer-Azima, F. J. (2002). "Group psychotherapy for children and adolescents," in *Child and Adolescent Psychiatry: A Comprehensive Textbook, 3rd Edn.*, ed M. Lewis (Philadelphia, PA: Lippincott Williams Wilkins), 1032–1036.

Dangerfield, M. (2016). "Sense un lloc a la teva ment, sense un lloc al món". Aspectes tècnics del treball amb adolescents negligits. ["No place in your mind, no place in the world". Technical aspects of working with neglected adolescents]. *Rev. Catalana Psicoanàl.* 33, 99–129.

Devi, A., and Fenn, E. (2012). 'Make sure you keep our house safe!' Thematic analysis of a children's psychotherapy group. *J. Child Psychother.* 38, 318–334. doi: 10.1080/0075417X.2012.720104

Echeazarra, I., Castellano, J., Usabiaga, O., and Hernández-Medo, A. (2015). Diferencias en el uso estratégico del espacio en categorías infantil y cadete de fútbol: una aplicación del análisis de coordenadas polares [Differences in the strategic use of space in under 14 and under 16 soccer: a polar coordinate analysis]. *Cuadern. Psicol. Deporte* 15, 169–180. doi: 10.4321/S1578-84232015000100017

Erikson, E. (1968). *Identity, Youth, and Crisis*. New York, NY: Norton.

Fonagy, P. (1991). Thinking about thinking: some clinical and theoretical considerations in the treatment of a borderline patient. *Int. J. Psychoanal.* 72, 639–656.

Fonagy, P., Gergely, G., Jurist, E., and Target, M. (2002). *Affect Regulation, Mentalization, and the Development of the Self*. New York, NY: Other Press.

Fonagy, P., Leigh, T., Kennedy, R., Mattoon, G., Steele, H., Target, M., et al. (1995). "Attachment, borderline states and the representations of emotions and cognitions in self and other," in *Rochester Symposium on Developmental Psychopathology: Cognition and Emotion*, eds D. Cicchetti and S. T. Rochester (New York, NY: University of Rochester Press), 371–414.

Foulkes, S. H. (1986). *Psicoterapia Grupo- Analítica. Métodos y Principios. [Group Analytic Psychotherapy: Methods and Principles]*. México: Gedisa.

Gergely, G., and Watson, J. (1996). The social biofeedback theory of parental affect-mirroring: the development of emotional self-awareness and self-control in infancy. *Int. J. Psychoanal.* 77, 1181–1212.

Haen, C., and Weil, M. (2010). Group therapy on the edge: adolescence, creativity, and group work. *Group* 34, 37–52. doi: 10.2307/41719263

Hernández-Mendo, A., López-López, J. A., Castellano, J., Morales-Sánchez, V., and Pastrana, J. L. (2012). Hoisan 1.2: programa informático para uso en metodología observacional [Hoisan 1.2: Software for observational methodology]. *Cuad. Psicol. Dep.* 12, 55–78. doi: 10.4321/S1578-84232012000100006

Herrero Nivela, M. L. (2000). Utilización de la técnica de coordenadas polares en el estudio de la interacción infantil en el marco escolar [Use of polar coordinate analysis to study children's interactions at school]. *Psicothema* 12, 292–297.

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174. doi: 10.2307/2529310

López, J., Valero, A., Anguera, M. T., and Díaz, A. (2016). Disruptive behavior among elementary students in physical education. *SpringerPlus* 5, 1154. doi: 10.1186/s40064-016-2764-6

López-López, J. A., Menescardi, C., Estevan, I., Falcó, C., and Hernández-Mendo, A. (2015). Análisis técnico-táctico en Taekwondo con coordenadas polares a través del software HOISAN [Technical-tactical analysis in Taekwondo with polar coordinates through software HOISAN]. *Cuad. Psicol. Dep.* 15, 131–142. doi: 10.4321/S1578-84232015000100013

Lorentzen, S., Ruud, T., Fjeldstad, A., and Høglend, P. (2013). Comparison of short- and long-term dynamic group psychotherapy: randomised clinical trial. *Br. J. Psychiatry* 203, 280–287. doi: 10.1192/bjp.bp.112.113688

Lutz, W., and Hill, C. E. (2009). Quantitative and qualitative methods for psychotherapy research: introduction to special section. *Psychother. Res.* 19, 369–373. doi: 10.1080/10503300902948053

Malekoff, A. (2014). *Group Work with Adolescents. Principles and Practice, 3rd Edn.* New York, NY: Guildford Press.

Malmberg, L., and Fenton, M. (2008). "Psicoterapia psicodinámica individual y psicoanálisis para la esquizofrenia y enfermedades mentales graves," [Individual psychodynamic psychotherapy and psychoanalysis for schizophrenia and serious mental disorders] (Revisión Cochrane traducida), in *La Biblioteca Cochrane Plus* (Oxford; Chichester: Update Software Ltd.; The Cochrane Library; John Wiley & Sons, Ltd.). Available online at: http://www.update-software.com

Merikangas, K. R., He, J., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., et al. (2010). Lifetime prevalence of mental disorders in US adolescents: results from the National Comorbidity Survey Replication-Adolescent Supplement (NCS-A). *J. Am. Acad. Child Adolesc. Psychiatry* 49, 980–989. doi: 10.1016/j.jaac.2010.05.017

Ministerio de Sanidad, Servicios Sociales e Igualdad (2014). *Clasificación Internacional de enfermedades 9ª Revisión, Modificación Clínica (CIE-9-MC)* [International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)] Available online at: www.msssi.gob.es

Morillo-Baro, J. P., Reigal, R. E., and Hernández-Mendo, A. (2015). Análisis del ataque posicional de balonmano playa masculino y femenino mediante coordenadas polares [Analysis of positional attack in beach handball male and female with polar coordinates]. *RICYDE Rev. Int. Cienc. Deporte* XI, 226–244. doi: 10.5232/ricyde

Nuffield Foundation (2013). *Social Trends and Mental Health: Introducing the Main Findings*. London: Nuffield Foundation.

Pascual-Leone, A., Greenberg, L. S., and Pascual-Leone, J. (2009). Developments in task analysis: new methods to study change. *Psychother. Res.* 19, 527–542. doi: 10.1080/10503300902897797

Perea, A., Castellano, J., Alday, L., and Hernández Mendo, A. (2012). Analysis of behaviour in sports through Polar Coordinate Analysis with MATLAB. *Qual. Quant.* 46, 1249–1260. doi: 10.1007/s11135-011-9435-z

Pingitore, F. (2016). The "voice" and perspectives of adolescents participating in a short-term psychotherapy group. *Int. J. Group Psychother.* 67, 1–23. doi: 10.1080/00207284.2016.1260460

Pollock, K., and Kymissis, P. (2001). The future of adolescent group therapy: an analysis of historical trends and current momentum. *J. Child Adolesc. Group Ther.* 11, 3–11. doi: 10.1023/A:1016689625271

Portell, M., Anguera, M. T., Chacón, S., and Sanduvete, S. (2015a). Guidelines for reporting evaluations based on observational methodology (GREOM). *Psicothema* 27, 283–289. doi: 10.7334/psicothema2014.276

Portell, M., Anguera, M. T., Hernández-Mendo, A., and Jonsson, G. K. (2015b). Quantifying biopsychosocial aspects in everyday contexts: an integrative methodological approach from the behavioral sciences. *Psychol. Res. Behav. Manag.* 8, 153–160. doi: 10.2147/PRBM.S82417

Pylvänäinen, P. M., Muotka, J. S., and Lappalainen, R. (2015). A dance movement therapy group for depressed adult patients in a psychiatric outpatient clinic: effects of the treatment. *Front. Psychol.* 6:980. doi: 10.3389/fpsyg.2015.00980

Reid, S., and Kolvin, I. (1993). Group psychotherapy for children and adolescentes. *Arch. Dis. Child.* 69, 244–250 doi: 10.1136/adc.69.2.244

Roberts, L. D., and Povee, K. (2014). A brief measure of attitudes toward mixed methods research in psychology. *Front. Psychol.* 5:1312. doi: 10.3389/fpsyg.2014.01312

Robles-Prieto, F. J., Castellano, J., and Perea, A. E. (2014). Diferencias del juego entre la Selección Española de Fútbol y sus rivales [Differences of the game betweeb the Football Spanish Team and its rivals]. *RICCAFD Rev. Iberoameric. Cienc. Activ. Física Deporte* 3, 1–8.

Sackett, G. P. (ed.) (1978). *Observing Behavior (Vol. 2): Data Collection and Analysis Methods*. Baltimore, MD: University of Park Press.

Sackett, G. P. (1980). "Lag Sequential Analysis as a data reduction technique in social interaction research," in *Exceptional Infant. Psychosocial Risks in Infant-Environment Transactions*, eds D. B. Sawin, R. C. Hawkins, L. O.Walker, and J. H. Penticuff (New York, NY: Brunner/Mazel), 300–340.

Sackett, G. P. (1987). "Analysis of sequential social interaction data: some issues. recent developments, and a causal inference model," in *Handbook of Infant Development*, ed J. D. Osofsky (New York, NY: Wiley), 855–878.

Sánchez-Algarra, P., and Anguera, M. T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: impact of recording and coding predominating perspectives. *Qual. Quant.* 47, 1237–1257. doi: 10.1007/s11135-012-9764-6

Santoyo, C., Jonsson, G. K., Anguera, M. T., and López-López, J. A. (2017). *Observational Analysis of the Organization of On-Task Behavior in the Classroom Using Complementary Data Analyses.* An. Psicol.- Spain.

Schwartze, D., Barkowski, S., Burlingame, G. M., Strauss, B., and Rosendahl, J. (2016). Efficacy of group psychotherapy for obsessive-compulsive disorder: a meta-analysis of randomized controlled trials. *J. Obsessive Compuls. Relat. Disord.* 10, 49–61. doi: 10.1016/j.jocrd.2016.05.001

Shechtman, Z. (2007). *Group Counseling and Psychotherapy with Children and Adolescents: Theory, Research and Practice.* New York, NY: Routledge

Simpson, S. G., Morrow, E., van Vreeswijk, M. F., and Reid, C. (2010). Group schema therapy for eating disorders: a pilot study. *Front. Psychol.* 1:182. doi: 10.3389/fpsyg.2010.00182

Skewes, S. A., Samson, R. A., Simpson, S. G., and van Vreeswijk, M. (2015). Shortterm group schema therapy for mixed personality disorders: a pilot study. *Front. Psychol.* 5:1592. doi: 10.3389/fpsyg.2014.01592

Sousa, D. J., Prudente, J. N., Sequeira, P., López-López, J. A., and Hernández-Mendo, A. (2015). Análisis de las situaciones de juego 2vs2 en el campeonato europeo masculino de balonmano 2012: aplicación de la técnica de coordenadas polares [Analysis of 2vs2 situations in men's european handball championship 2012: application of polar coordinates technique]. *Cuadern. Psicol. Deporte* 15, 181–194. doi: 10.4321/S1578-84232015000100018

Tellerman, J. S. (2001). The Solutions Unlimited Now-SUN(sm) program: psychodynamic/cognitive structured groups for teens, pre-teens, and families. *J. Child Adolesc. Group Ther.* 11, 117–134. doi: 10.1023/A:1014758406520

Torras de Beà, E. (2013). *Psicoterapia de Grupo [Group Psychotherapy]*. Barcelona: Octaedro.

Tusón, A. (1995). *Anàlisi de la Conversa [Conversation analysis]*. Barcelona: Empúries.

Venturella, M. (2016). *Autismo, un Hermano, Otros Hermanos, un Grupo. Estudio Observacional del Proceso de un Grupo de Niños con Hermanos Diagnosticados de Trastorno de Espectro Autista (TEA) [Autism, a Sibling, Other Siblings, a Group. Observational Study of the Process within a Group of Children with Siblings with Autism Spectrum Disorder]* Unpublished Doctoral Dissertation, Universidad Ramon Llull, Barcelona.

Weschler, D. (2006). *Weschler Intelligence Scale for Children, 4th Edn (WISC-IV)*. Madrid: TEA Ediciones.

Yalom, I. D. (2005). *The Theory and Practice of Group Psychotherapy, 5th Edn.* New York, NY: BasicBooks.

Parental and Infant Gender Factors in Parent–Infant Interaction: State-Space Dynamic Analysis

M. Angeles Cerezo[1]*, Purificación Sierra-García[2], Gemma Pons-Salvador[1] and Rosa M. Trenado[1]

[1] Department of Psychology, University of Valencia, Valencia, Spain, [2] Department of Developmental Psychology, National Distance Education University, Madrid, Spain

This study aimed to investigate the influence of parental gender on their interaction with their infants, considering, as well, the role of the infant's gender. The State Space Grid (SSG) method, a graphical tool based on the non-linear dynamic system (NDS) approach was used to analyze the interaction, in Free-Play setting, of 52 infants, aged 6 to 10 months, divided into two groups: half of the infants interacted with their fathers and half with their mothers. There were 50% boys in each group. MANOVA results showed no differential parenting of boys and girls. Additionally, mothers and fathers showed no differences in the Diversity of behavioral dyadic states nor in Predictability. However, differences associated with parent's gender were found in that the paternal dyads were more "active" than the maternal dyads: they were faster in the rates per second of behavioral events and transitions or change of state. In contrast, maternal dyads were more repetitive because, once they visited a certain dyadic state, they tend to be involved in more events. Results showed a significant discriminant function on the parental groups, fathers and mothers. Specifically, the content analyses carried out for the three NDS variables, that previously showed differences between groups, showed particular dyadic behavioral states associated with the rate of Transitions and the Events per Visit ratio. Thus, the transitions involving 'in–out' of 'Child Social Approach neutral – Sensitive Approach neutral' state and the repetitions of events in the dyadic state 'Child Play-Sensitive Approach neutral' distinguished fathers from mothers. The classification of dyads (with fathers and mothers) based on this discriminant function identified 73.10% (19/26) of the father–infant dyads and 88.5% (23/26) of the mother–infant dyads. The study of father-infant interaction using the SSG approach offers interesting possibilities because it characterizes and quantifies the actual moment-to-moment flow of parent–infant interactive dynamics. Our findings showed how observational methods applied to natural contexts offer new facets in father vs. mother interactive behavior with their infants that can inform further developments in this field.

Keywords: father–infant interaction, mother–infant interaction, parental gender, infant gender, state-space grid (SSG), dynamic systems

INTRODUCTION

In the last three decades, greater recognition has been given to the role of the father in child development (NIDCH Early Child Care Research Network, 2000; Tamis-LeMonda et al., 2004; Lamb, 2010). Moreover, father involvement predicts the quality of family interactions from the earliest stages of a child's life (Simonelli et al., 2016). Wider involvement of fathers in the rearing and caring for their infants leads to increased opportunities for early interactions (Pleck, 2010). In this respect, the focus should be on, not only the amount of time the father spends with his infant but also, more importantly, how he uses that time and the quality of the relationship, using objective measures (Yago et al., 2014).

Studying father–infant interactions goes beyond the interest in the specific area itself because the quality of interaction has a strong relationship with the child's development and attachment (Shaw et al., 2005; Kochanska et al., 2008). In the area of early interaction and attachment, researchers have been traditionally focused on the mother as primary caregiver, to which the use of the Maternal Sensitivity construct has been central (Ainsworth et al., 1974). However, according to Bowlby (1969/1982), the child's choice of attachment figure depends on who cares for him, or her, and on the composition of the family. Therefore, other people can fulfill the role of primary caregiver. Moreover, the child's relationships with other figures who share the role of caregiver, along with the primary caregiver, were an area of concern previously noted by Mary Ainsworth herself (Ainsworth et al., 1978).

Father–Infant Interaction vs. Mother–Infant Interaction

Research findings involving mothers and fathers interacting with babies shows a mixed picture. Some studies, involving children aged from 9 to 24 months, find differences between mothers and fathers. Thus, in contrast with mothers, fathers observed in free-play, were less sensitive and often intrusive, for example, introducing questions or requiring information that may interfere with play (Leaper, 2000; Volling et al., 2002; Lovas, 2005; Kwon et al., 2012; Hallers-Haalboom et al., 2014; Fuertes et al., 2016). Likewise, Feldman and Klein (2003) found that fathers interacting more through contact and physical play, were usually less positive, more unpredictable, and characterized by sudden peaks of emotional intensity. Fathers used more stimulation and exploratory play and less emotional support behaviors (Grossmann et al., 2008).

Other studies, involving a wider range of child ages, from 0 to 36 months, find no differences between fathers and mothers in the quality of interactions with their children (e.g., Goossens and van IJzendoorn, 1990; Braungart-Rieker et al., 2001; Lewis and Lamb, 2003; John et al., 2012; Yago et al., 2014) or in the intensity of their negative affect (Ekas et al., 2011).

These studies involved infants of different ages from very young infants to toddlers (Braungart-Rieker et al., 2001; Yago et al., 2014) and different measures, from event-based schemes

to rating scales (Volling et al., 2002; Ekas et al., 2011). Some of these studies compared mother and father with the same child (e.g., Yago et al., 2014) which could be clouding possible differences, as it is plausible that a partner influences the quality of parent–infant interaction. In fact, some research indicates that partners can become similar through the process of marital life (Easterbrooks and Goldberg, 1984; Osnat and Bonnie, 1995; Lundy, 2003). Especially during infancy, parents can rely on each other in searching and implementing successful strategies of interaction with their infant, leading to bidirectional modeling (Braungart-Rieker et al., 1998; Schoppe-Sullivan et al., 2007). Additionally, Barnett et al. (2008) found that perceived high level of marital quality was associated with interdependence of sensitive parenting behaviors in mother–infant and father–infant interactions. This could explain the high correlations between the scores obtained by couples of fathers and mothers in sensitivity and intrusiveness (Braungart-Rieker et al., 1998; Volling et al., 2002; Tamis-LeMonda et al., 2004; Hallers-Haalboom et al., 2014). Finally, most of the studies did not consider infant gender as a factor (Braungart-Rieker et al., 2001; Kwon et al., 2012).

Gender of Parents and Infants: Early Interaction

Two theoretical frameworks describe some mechanisms regarding differential parenting of boys and girls. The biosocial theory proposes that the parents use gender-differentiated parenting as a means of gender-role socialization (Eagly and Wood, 2002; Wood and Eagly, 2012) and gender schema theories (Bem, 1981; Markus et al., 1982) proposes that parenting would be affected by parents' gender-role stereotypes.

Considering these two theories, Endendijk et al. (2016), conducted a meta-analysis of 126 observational studies, involving 15,034 families to examine parental differences with their sons and daughters. They used 'autonomy-supportive strategy' in parental behavior, that is, child-centered responding and promoting autonomy through support, conceptually similar to the construct of parental sensitivity as formulated within Attachment Theory (Bowlby, 1969/1982; Ainsworth et al., 1978) and controlling strategies, similar to parenting practices described within Coercion Theory (Patterson, 1982; Eddy et al., 2001). Contrary to their expectations, no overall gender-differentiated effect was found in autonomy-supportive strategies, and they found very small effects ($d = 0.08$) of child gender on parents' use of control after excluding outlying effect sizes by which parents used more controlling strategies with boys than with girls.

Endendijk et al. (2016) in their meta-analytic study included boys and girls from 0 to 18 years. Although biosocial theory does not explicitly consider child age, it is plausible to expect some gender-specific parenting related with developmental level. With older children expressing their demands more clearly, parents would be more effective in adjusting their behavior to their demands. Hallers-Haalboom et al. (2014) reported that mothers responded in a more sensitive and non-intrusive way to their older children (between 2.5 and 3.5 years) compared to younger ones (12 months) without being influenced by infant gender.

Likewise, in the area of parental control some meta-analytic evidence supports this, for example, the findings reported by Leaper et al. (1998) by which gender differences in the mother's directive speech was more evident with older children than with younger ones. In contrast, other studies found that parental control decreases with the child's age in favor of child self-control (Lytton and Romney, 1991). The combined effect size reported by Endendijk et al. (2016) for the differences in parental controlling: more with boys than with girls was largest in the youngest age group (0–2 years: $d = 0.16$). The findings for that age group were coming mainly from studies involving toddlers, because in the pool of 126 studies, 16.67% (21 studies) included children averaging in age from 1 to 2 years and only one study (Huber, 2012) included children whose average age was under 12 months.

Research suggests that parent–infant interaction can be affected, not only by the gender of children but also the parent's gender. However, Endendijk et al. (2016) testing only differential controlling of boys and girls in those studies which included fathers and mothers, found no effect of parental gender on the extent of their differential treatment of boys and girls.

In summary, there is no consensus about the extent to which parents treat their sons and daughters differently. Moreover, there are other factors, like the setting of the interaction, which may interact with the gender factor. Thus, in meta-analytic studies it has been found that differences of gender on interaction are often lower in relatively unstructured settings, such as free-play, than in structured tasks such as problem-solving (Leaper et al., 1998; Endendijk et al., 2016).

Therefore, the heterogeneity of measures, age, and settings, across studies in this field can prevent potential gender-differences being detected. However, even the existence of parental differences in the treatment of their children may be reflecting differences in parental practices which may be due to factors other than gender, like birth order (Lovas, 2005; Hallers-Haalboom et al., 2014). Some studies showed that fathers and mothers are more sensitive to the first child than to later ones. These differences were especially pronounced when the second born was the same gender as the firstborn, and fathers were more likely to show differential treatment than mothers (Van IJzendoorn et al., 2000; Furman and Lanthier, 2002; Hallers-Haalboom et al., 2014).

Although studies on dyadic interaction show that the gender of parents and children may affect parental behavior, the direction of those influences is not yet conclusive (Hallers-Haalboom et al., 2014). Nor is it conclusive that the level of parental sensitivity depends, exclusively at least, on the combinations of parent–child gender (Lovas, 2005; Schoppe-Sullivan et al., 2007; Hallers-Haalboom et al., 2014; Endendijk et al., 2016).

Early Interaction and Measurement

The area of caregiver–infant interaction has been influenced by the Sensitivity construct, referred to as 'Maternal Sensitivity' because it has been regarded as one of the most important mediators of attachment patterns (Ainsworth et al., 1978). This construct consists of awareness of the child's cues and demands, his/her appropriate interpretation, and the ability to respond quickly and accurately (Ainsworth et al., 1978). This

central feature of maternal behavior was originally assessed with a global rating scale: Ainsworth's Maternal Sensitivity Rating Scale (Ainsworth et al., 1974). This strategy has been the most influential and common in this field (for a review of the Global Interaction Scales, see Leclère et al., 2014).

One of the most important features of rating-scale approaches is that they do not capture the temporal dimension of the interaction. Research has highlighted that infants develop an early procedural representation of the world before they develop symbolic forms of representation (Beebe et al., 2010). In addition, the infant's procedural form of representation is based on his perception of contingency and the predictability of events: infants develop ongoing expectations of sequences of events, within the self, within the 'other' and between the two (Tarabulsy et al., 1996; Beebe and Lachmann, 2002; Gergely, 2004). This unfolds during the process of interaction over time. Therefore, to examine these central aspects of parent–infant interaction requires a real time sequential coding approach.

In general, rating-scale and temporal sequential approaches, referred to as macro- and micro-analytic approaches, respectively, have tended to favor different contexts. The macro-analytic one has traditionally used naturalistic situations, such as free-play and the most frequent setting for micro-analytic studies has been face-to-face interaction with the mother on a chair, facing the infant, who is secured in a baby seat. In the latter context, the coding of mother–infant interaction states is done in minor units, for example, units of 1 s. These constraints on the mother–infant interaction and the fragmentation of the analysis have been criticized (Mesman, 2010). In fact, free-play offers greater ecological validity because the mother, or caregiver, has no restrictions on their behavior with their child.

In this context, there is a third way: the observational strategies of sequential coding in real time in a free-play situation (Cerezo et al., 2008). Indeed, the interaction during free-play can be sequentially coded as it unfolds, with mutually exclusive and exhaustive defined categories for infant and mother. Thus, the recorded data can be read as a sort of abbreviated text, reflecting the stream of behavior. The analyses can provide important information, not only about "what" the parent responds to, but also "with what" and "when," in that stream of social exchange (Cerezo et al., 2012, 2016). Therefore, micro-analytic approaches, that is to say, approaches including the temporal dimension, may be a further step in the understanding of sensitivity and "appropriateness" of parental responses because they look at parental matching/contingent behavior to the child's behavior, which fosters synchrony and mutual emotional regulation in the interaction (Sroufe, 1995; Feldman, 2007; Woodhouse, 2010).

Non-linear Dynamic Systems Approach to Interaction

Consideration of the temporal dimension in the measurement of parent–infant interaction allows for the examination of the dynamic process in dyadic interaction. In this context, the non-linear dynamic systems (NDS) framework and its principles that account for properties of dynamic, complex, adaptive, open systems, offer an instrument to examine these processes

(Prigogine and Stengers, 1984; Thelen and Smith, 1994, 1998). This approach has led to a paradigm shift in multiple fields (Fogel, 2011), including that of dyadic interaction. Indeed, by accepting the dyad as a dynamic system, then the NDS principles can account for dyadic behavioral patterns that emerge and stabilize through the system's internal feedback processes (Hollenstein et al., 2004; Hollenstein, 2007; Dishion, 2012). The State Space Grid (SSG) analysis, a graphical tool based on a dynamic system approach (Hollenstein, 2013) allows visualizing the content, temporal and affective flow of the interactions (Sravish et al., 2013) and relevant structures and dimensions of that interaction (Feldman, 2007; Beebe et al., 2010).

Few studies have used NSD and SSG indices in early parent–child interaction. Sravish et al. (2013) have used this paradigm to study the dynamic regulation behaviors between the child and his caregiver through the face-to-face Still-Face paradigm. Cerezo et al. (2012) used the SSG for the study of dyadic flexibility, in the context of interactions. In that study, dyadic flexibility was an index of sensitivity, a precursor of attachment. However, all these studies have focused on mother–infant dyads.

Antecedents and Purpose of this Study

The present study is part of a research program focusing on detecting precursors of attachment in which the central character has been the mother (Cerezo et al., 2006, 2008, 2012, 2016). The present development of the research program addresses father–infant interaction, compared with mother–infant dyads, using the same observational methodology, the NDS approach and tools of previous studies and, additionally, it considers the factor of the infant's gender in the interactive process. The general purpose was to progress the understanding of paternal behavior using systematic observation, so the study of precursors of attachment can include fathers when they are caring for their infants, including the potential infant gender effects on that parent–infant interaction.

Given the lack of studies, using this approach for this specific topic, the overall purpose of the current study was exploratory. Specifically, the purpose was twofold: on the one hand, to compare the interactive profile of dyadic temporal organization of fathers with their babies and mothers with theirs and, on the other hand, to examine the effect of the baby's gender on the interaction.

MATERIALS AND METHODS

Participants

The participants in this study were 52 infants: 26 children interacting with their mothers and 26 children interacting with their fathers. Boys and girls were equally represented (50%) in both groups.

The parents came from the general population who joined a community-based program provided on a universal basis to support parenting during the first 2 years of life. The Program comprised six trimestral visits over a period of a year and a half. As part of the service, parent and infant are videotaped in a free-play situation, with parental consent, to analyze their interaction

and provide parents with individual guidance (for a description: Cerezo and Pons-Salvador, 1999; for a summary of evaluation studies: Pons-Salvador et al., 2014). The first time they attend, parents do not receive any information after they have finished the free-play and for more specific and personalized feedback, they need to wait until their second visit when their free-play has been analyzed.

The criteria to select the cases were:

For the parents: to be (a) the biological mother or father, (b) involved with childcare, (c) to be the first visit to the Program; so no previous intervention received and this visit should be before the child was 12 months old. For the infants: (a) absence of congenital anomalies or neurological diseases, all the children' development was appropriate for their chronological age, assessed by the Developmental Scales of Knobloch et al. (1980) and (b) gender with 50% of girls in each group.

The selection procedure was the following: First, the infants interacting with their fathers were selected. Although the program is offered to both parents, about 15% of fathers attended at least one visit (out of the six Program visits). The initial pool of data comprised 980 families. There were 145 cases of infants interacting with their fathers at least once. From these there were thirty-five who met the criterion of having their first visit to the program on their own and, of those, 29 dyads met the criterion of the child's age. There were 13 girls in that group; to balance the gender factor thirteen out of the sixteen cases involving interaction with a male infant were randomly selected for the final group. Secondly, to select the group of infants interacting with their mothers, 13 girls were randomly selected from those cases in which mother attended the first visit, so had no prior intervention and then 13 boys to comprise a similar group to the one of infants with their fathers.

In both groups, mothers' and fathers' interaction was assessed when their children were, on average, 36.47 weeks of age ($SD = 2.85$), ranging from 26 to 44 weeks. No significant differences in infants' age ($t = 3.37$, $df = 48$, $p = 0.71$) with mothers: $M_{age} = 36.62$, $SD_{age} = 2.46$ and with fathers: $M_{age} = 36.32$, $SD_{age} = 3.22$. The second half of the first year shows important advances for the child's social and emotional development. From the neuro-relational approach, regarding adjustment and interaction, Lillas and Turnbull (2009) point to the age range 6 to 10 months as being the one where children display bi-directional intentional communication (child–adult) interaction.

The mean age of the 26 mothers in the study was 28.35 years ($SD = 6.09$), ranging from 17 to 41 years. As for the 26 fathers, the mean age was 32.48 years ($SD = 6.76$), with a range between 21 and 48 years. The two groups had a similar average number of children: 1.58 ($SD = 1.03$) in the group of mothers and in the group of fathers, 1.73 ($SD = 1.18$). The birth order for infants was similar in both groups. Thus, the majority of the infants were first or second born: 84.61% and 88.46%, in father and mother groups, respectively, 11.53 and 7.69% were third or fourth born, and only 3.84% in both groups were children in fifth or sixth position. The comparison between mother and father groups in the number of dyads with first child vs. second child vs. third child plus, showed no significant differences. In the father–infant

group, all the fathers, except one, came from two-parent families and there were three single-mothers in the mother–infant group.

The participants were all resident in Ireland. In the fathers group 84.61% were Irish or from other European countries, like Poland, and the rest, 15.39%, were from the United States, the Philippines, Libya, and Turkey. With regard to the mothers, 76.92% were Irish or from other European countries. There were 7.69% mothers from South-American countries, 11.53% from African countries and 3.8% from India.

Regarding educational levels, 69.23% mothers and 42.3% fathers had only Secondary School education (Intermediate/Junior Cert level: 26.92% of mothers and 7.69% of fathers; Leaving Certificate level 42.31% mothers and 34.61% fathers). Studies at third level: 23.10% of mothers and 34.61% of fathers, finally 7.69% of mothers and 11.53% of fathers reported post-graduate studies. One father reported only having Primary school studies (3.84%) and two fathers did not provide this information (7.69%).

Regarding occupations, there were 11.53% mothers working full time at home. In the two groups, 53.84% of mothers and 46.15% of fathers reported being unemployed, 11.53% mothers and 7.69% of fathers worked in unskilled occupations, while 15.38% of mothers and 30.76% fathers were in semi-skilled jobs, and, finally, 11.53% in each group were in qualified occupations.

Procedures

In the context of the visit to the Program, the professional left the parent with the infant in a room for the free-play. This session took place in a room with a table and a chair and there were some toys appropriate for the child's age. The parent was told to play with his/her child the way she/he normally did and if she/he wanted, she/he could use the toys. The systematic observation was carried out on 4–5 min free-play session. An average of 5 min play is sufficient, according to studies in this area by Kemppinen et al. (2005). On the table there was a matt and in all cases parents played with their child on the table or on their lap. The session was videotaped for coding as part of the routine of the Program. Consequently, the protocol for the staff was that the parent and child play between 4 and 5 min, with flexibility. The final duration did not have to do with any parental/infant characteristic. Sometimes, the play was closer to 4 min and sometimes a bit longer, depending on the circumstances (i.e., staff may have been momentarily busy when the timer rang went off).

The parents were informed consent to be videotaped. Additionally, they gave written consent for the anonymous use of their data for research purposes. This study was approved by the Research Ethics Committee of the University of Valencia.

Measures
Parent–Infant Interaction and Early Mother–Child Interaction Coding System-Revised (Códigos para la Interacción Temprana Materno-Infantil: CITMI-R (Trenado and Cerezo, 2007, Unpublished)

The structure and coding rules of CITMI-R are based on SOC III (Cerezo, 2000), being a parallel version for young children. The CITMI-R categories, having been defined in a
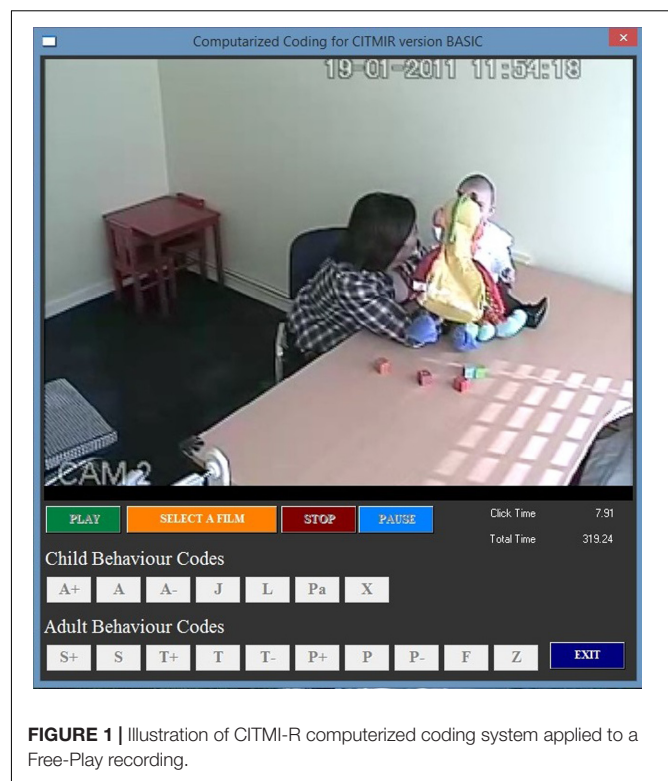
FIGURE 1 | Illustration of CITMI-R computerized coding system applied to a Free-Play recording.

mutually and exhaustive way, put the stream of mother–child interaction into observational data that can be analyzed. The computerized coding, specially devised for CITMI-R coding, allows the observer to code in real time without interruption during the period of observation (**Figure 1**).

CITMI has shown good standards of psychometric properties in content validity and criterion validity with dyads from Spain, Brazil, and Ireland (Alvarenga and Cerezo, 2013; Trenado et al., 2014).

The CITMI-R Observational Categories

It includes four categories for the parent's behavior, three interactive and two non-interactive (**Table 1**). For the child there are four categories, one interactive and three non-interactive. All interactive behaviors, according to affect, can either be positive, neutral, or negative, except for "Sensitive" parental behavior, which, by definition and nature, can only be either neutral or positive affect. Therefore, there were 54 possible dyadic states, "parent–infant variables," used for SSG: nine maternal/paternal codes $\times 6$ infant's codes. **Table 1** shows a descriptive summary of the CITMI-R.

NDS and State Space Grid Measures

The unit was the dyad. The codes for the infant (x-axis) and the parent (y-axis) were represented on a quasi-ordinal scale from the most positive to the most negative (**Figure 2**). The state-space grid for this study consisted of 54 (cells) potentially possible dyadic states: any combination of parent–child behavior. Each dyad "danced" around the state-space grid during the Free-Play session, which was considered as an individual trajectory. Each

Child categories

__Interactive__

Social approach (A): Social approach, verbal, or non-verbal, to the parent as a response to her/him or as a child's initiative. It has three affect or valences: positive, neutral, and negative.

__Non-Interactive__

Solitary play (J): The child is involved in his/her own game with or without a toy; she/he is clearly demonstrating interest in the exploration (his own hands, clothes, objects, etc.).

Solitary crying and/or whining (L): The child express general discomfort usually related to being tired, sleepy or hungry.

Passive/disinterested/apathetic behavior (Pa): The child shows a bored or non-attentive facial expression. If the child ever catches hold of something there is not looking at it or exploration.

Parental categories

__Interactive__

Sensitive (S): Social approach, verbal or non-verbal, that meets the demands of the situation and is appropriate for the age, abilities and interests of children. This approach DOES NOT interrupt child's ongoing activity, or intrude in child's space. It includes proposals of toys or games to the child in a way that the child has a choice to accept it or not. It has two affect or valences: positive and neutral.

Intrusive (T): Social approach, verbal or non-verbal, that interrupts on-going activities of the child and/or invades his space and it is not meeting child's needs. It includes proposals of toys or games to the child in a way that the child has no choice or they are above child's skills or reach, like putting a toy in his hand, or too far away. It has three affect or valences: positive, neutral, and negative.

Protective (P): Social approach, verbal or non-verbal that interrupts child's ongoing activity, or intrudes in child's space with the aim of protection or help (wiping child's nose, changing child's position, etc.). It has three affect or valences: positive, neutral and negative.

__Non-Interactive__

Indifferent/non-response (F): Lack of interaction with the child showing lack of attentiveness and lack of facial expression, or the parent looks away not responding to child's approach.
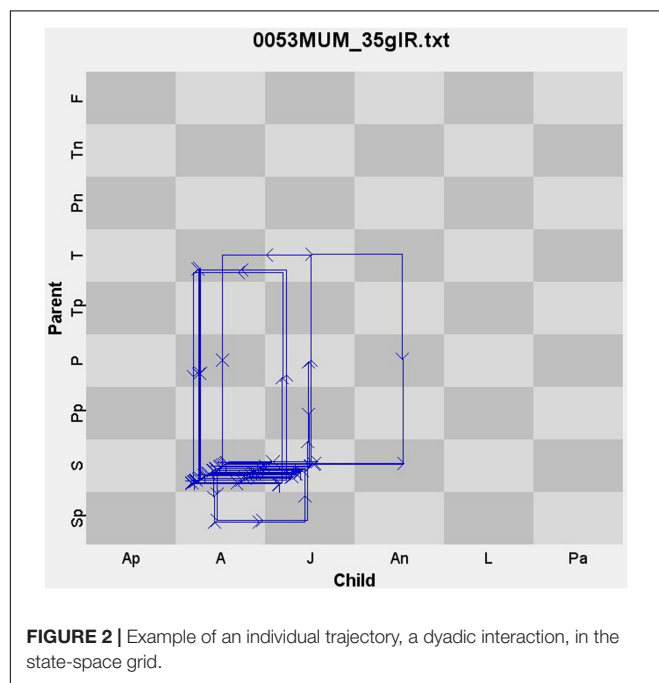
FIGURE 2 | Example of an individual trajectory, a dyadic interaction, in the state-space grid.

trajectory begins in one state (cell) and, as time progresses, tends to visit other cells on the grid (**Figure 2**).

The variables derived from the SSG data were:

(a) Different states that the dyad visit in the total state-space grid. A greater number of states (cells) visited means a greater range of content. The variable operationalized was the number of different cells occupied: 'Diversity' the value of which could go from one to 54 possible states.

(b) "Dispersion" in the state-space grid: a relative value, the sum of the squared proportional duration across all the cells. The values range from 0 (no dispersion at all - all behavior in one cell, therefore highest predictability) to 1 (maximum dispersion, less predictability). "Dispersion" strongly correlates with duration entropy (Hollenstein, 2013). The value is created by the formula: $1 - \left(\left(n \sum \left(d_i / D \right)^2 \right) - 1 \right) \Big/ (n - 1)$. Where D is the total duration, d_i is the duration in the cell i, and n the total number of cells.

(c) Dyadic event corresponded to a single node in a particular cell. Dyadic events are the smallest unit of states displayed on the grid and can have different durations. It was operationalized as the 'Frequency of Events' registered.

(d) Changes from one state to another, transitions or movements between cells on the grid done by the dyad. This is a content independent and dynamic measure of variability because transitions can take place between any numbers of cells. A cell visit beginning upon a trajectory's entry into the cell and ending upon its exit. There can be one or more consecutive events occurring within a single visit. The variable operationalized was 'Transitions,' number of visits, minus 1, because the first event in the first cell is counted as a visit.

'Events' and 'Transitions' are frequency-based measures. Therefore, they are time dependent, the longer the observation the more opportunities. Consequently, it is recommended that the variables of frequency of Events and

frequency of Transitions be divided by the total duration of the trajectory because the duration of each trajectory – time observed in each dyad - can be slightly different, Thus the variables were converted into rates per second: 'Transitions rate' and 'Events rate.'

(e) Events per Visits Ratio. When there are no repeating events, the 'Events per Visits' ratio is 1, the higher the ratio, the more repetitive events take place.

Reliability for Interactive Measures

The quality of the data coded from the free-play was validated by having a second independent coder, who was unaware of the study's purpose, coding one third of the total number of free-play episodes. These 16 dyads were randomly selected, half from the father–infant group and half from the mother–infant group. For the purpose of the reliability analyses, the second coding had the same length as the first, to avoid differences between the outputs in the length of the coded observation. The average length of the observation across the 16 dyads was 5.2 min ($SD = 1.3$).

Specifically, for the reliability analysis of the measurements obtained by CITMI-R, three approaches were used. Firstly, Alignment Kappa (Bakeman and Quera, 2011) was computed to calculate the agreement between coders. This method identifies commission-omission errors and is based on an algorithm that determines the optimal global alignment between two single code event sequences. The mean Kappa statistic for the 16 episodes analyzed, and for all categories, was 0.68 ($SD = 0.06$). The values from 0.61 to 0.80 are considered good and observer accuracies of 90% or better result in alignment kappa of 0.60 or better (Quera et al., 2007). Secondly, the SSG was used and the main variables derived were considered to test the reliability in terms of agreement between the two observers on the NDS variables under study. The Pearson correlations were computed for "Diversity," $r_{xx} = 0.74$, mean scores for coders 1 and 2: 8.19 ($SD = 2.85$), and 8.56 ($SD = 2.63$); "Duration per cell," $r_{xx} = 0.84$, mean scores for coders 1 and 2: 44.37 ($SD = 23.05$), and 45.21 ($SD = 25.33$); "Number of Events," $r_{xx} = 0.91$, mean scores for coders 1 and 2: 126.94 ($SD = 29.58$), and 112.50 ($SD = 25.32$) and, finally for "Number of Visits," $r_{xx} = 0.71$, mean scores for coders 1 and 2: 50.06 ($SD = 16.02$), and 42.13 ($SD = 15.42$). All the correlation values were shown to be statistically significant ($p < 0.002$). Thirdly, given that Intraclass Correlation Coefficient (ICC) is also recommended (Fleiss and Cohen, 1973; Shoukri, 2004), this was also computed for the same SSG variables. The following values were obtained: "Diversity," ICC = 0.85; "Duration per cell," ICC = 0.91; "Number of Events," ICC = 0.94; "Number of Visits," ICC = 0.83. All values were shown to be statistically significant ($p < 0.001$) and can be interpreted as excellent, as all were above 0.80.

Data Analyses

The SSG analysis (Lewis et al., 1999; Hollenstein, 2013) allows visualizing the content, temporal, and affective flow of the interactions (Sravish et al., 2013) and relevant structures and dimensions of the interaction (Peck, 2003; Feldman, 2007; Beebe et al., 2010).

The parental and infant gender factors on parent–infant interaction were examined by conducting a two-way MANOVA (Multivariate analysis of variance) for the data obtained with the SSG regarding 'Diversity,' 'Dispersion,' 'Event rate,' 'Transition rate,' and 'Events per Visit' ratio. The assumption of homogeneity was tested by computing variance Levene's Test.

To study the magnitude of the relationships between the variables analyzed significance level value of 0.05 was considered, and the effect size statistic (η^2) was computed. The statistical package SPSS v.21 for Windows was used for the analyses.

As part of the analysis plan, if the SSG variables showed differences between groups, it was planned to analyze the 54 possible potential dyadic states for those particular variables, to predict father–infant dyads vs. mother–infant dyads. For this purpose a Linear Discriminant Analysis was chosen with a stepwise variable selection method, applying the Wilks Lambda method, and the verification criteria associated with the F values by default, in SPSS program.

RESULTS

After a preliminary analyses section, the results section regarding parental and infant gender on parent infant interaction will address the two goals of the study. Firstly, the study of the NDS and State-Space measures considering parental and infant gender and secondly, to examine the NDS and State-Space measures considering the behavioral dyadic states to predict parental gender membership of the dyads.

Preliminary Analyses

The duration of the Free-play ranged from 240.05 to 348.60 s. As a preliminary step, the observation time was analyzed in relation to parental and infant gender using ANOVA. The dependent variable was the duration of the free-play.

The results showed that there were statistically significant differences in the observation time between the groups of dyads. The free-play from the father–infant dyads was shorter ($M = 267.58$, $SD = 13.69$) than the one involving mothers and infants ($M = 321.07$, $SD = 13.69$), ($F(1,51) = 7.83$, $p = 0.008$, $\eta^2 = 0.14$). No differences were found for the infant's gender ($F(1,51) = 0.70$, $p = 0.40$, $\eta^2 = 0.015$), or parental gender by infant gender ($F(3,153) = 0.02$, $p = 0.90$, $\eta^2 = 0.00$). Consequently, for subsequent analyses the number of events and transitions, based on the number of visits, were divided into the observation time in seconds for each dyad, using the rate per second.

Parental and Infant Gender on the Parent–Infant Interaction: NDS and State-Space Variables

The two-way MANOVA to test parental gender and infant gender on the parent–infant interaction measures showed no interaction effect between the two factors: parental and infant gender on the combined dependent variables. There was a multivariate effect for parental gender on the parent–infant interaction measures ($F(5,44) = 6.52$, $p = 0.00$, Wilk's λ = 0.563, $\eta^2 = 0.43$) (**Figure 3**).
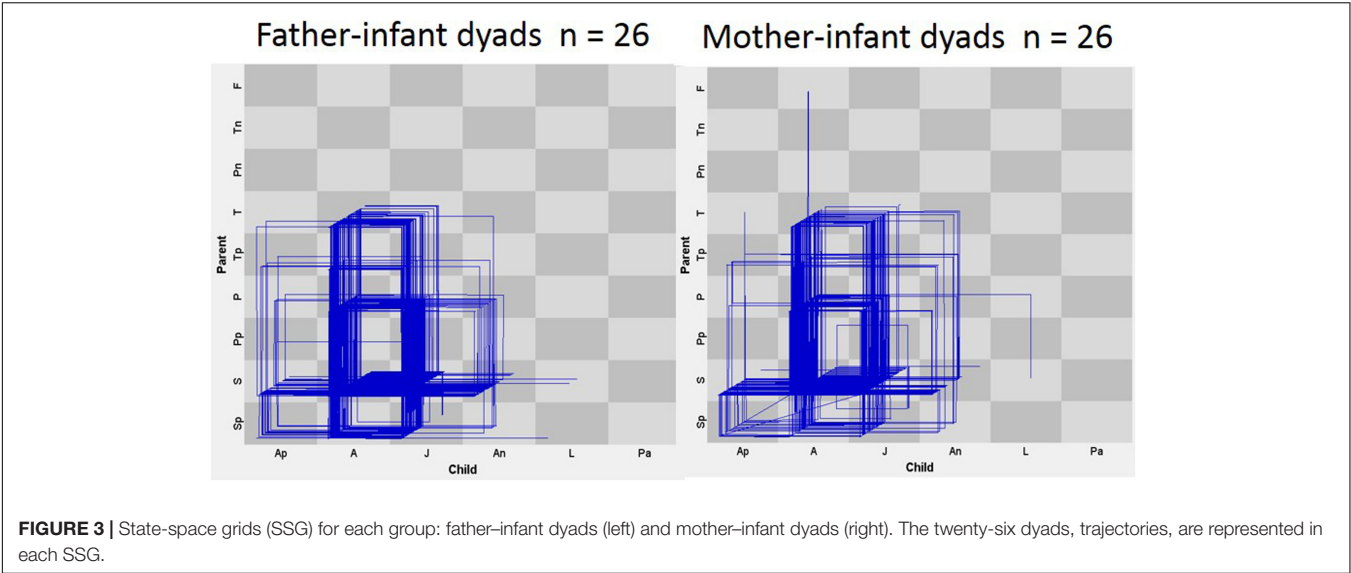
FIGURE 3 | State-space grids (SSG) for each group: father–infant dyads (left) and mother–infant dyads (right). The twenty-six dyads, trajectories, are represented in each SSG.

TABLE 2 | Descriptive statistics and between-group comparisons.

NDS variables and SSG measures	Father–infant dyads			Mother–infant dyads			Father–infants vs. Mother–infants groups		
	Total	Boy	Girl	Total	Boy	Girl			
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	F	p	η²
Diversity	9.38 (3.29)	9.38 (2.98)	9.38 (3.71)	8.19 (2.65)	7.31 (2.69)	9.07 (2.39)	2.07	0.15	0.04
Dispersion	0.68 (0.12)	0.68 (0.13)	0.68 (0.11)	0.63 (0.11)	0.59 (0.13)	0.66 (0.09)	2.89	0.10	0.06
Events rate (seconds)	0.49 (0.09)	0.51 (0.09)	0.47 (0.08)	0.40 (0.08)	0.41 (0.07)	0.40 (0.09)	12.34	0.00**	0.20
Transitions rate (seconds)	0.22 (0.05)	0.22 (0.06)	0.21 (0.04)	0.15 (0.04)	0.15 (0.03)	0.15 (0.05)	23.23	0.00**	0.33
Events per visits ratio	2.34 (0.56)	2.44 (0.73)	2.25 (0.35)	2.84 (0.76)	2.74 (0.40)	2.94 (1.0)	7.09	0.01*	0.13

*$p < 0.05$, **$p < 0.001$.

Specifically, significant differences were found in the Event rate ($F(1,39) = 12.34$, $p = 0.001$, $\eta^2 = 0.20$) and in the Transition rate ($F(1,39) = 23.23$, $p < 0.001$, $\eta^2 = 0.33$), in both of those the father–infant dyads showed a higher rate per second than the mother–infant dyads. Moreover, the two groups showed significant differences in the Events per Visits ratio ($F(1,39) = 7.09$, $p = 0.010$, $\eta^2 = 0.13$). The mother–infant dyads showed a higher value than the father–infant dyads (**Table 2**). Taken together the results indicated that dyads with fathers changed more frequently from state-to-state and engaged in more events per second, while mothers were more repetitive than fathers when interacting with their infants (**Figure 4**).

Regarding exploratory examination about the possible differences in interaction with infant boys and girls, no significant differences were found in relation to infant gender ($F(5,44) = 0.93$, $p = 0.47$, $\eta^2 = 0.09$). Therefore, in terms of the measures examined in this study, parent–infant interaction was similar with boys and girls.

In summary, both mother and father dyads showed similar diversity in their interaction, in terms of the number of different dyadic states they went through, and similar levels of dispersion. Additionally, paternal dyads were more 'active' than maternal dyads: they were faster in the rate of Events and in the rate

of Transitions (per second). In contrast, maternal dyads were more repetitive than paternal ones because they engaged in more events once they visited a particular dyadic state. There were no differences between girls and boys.

Profile of the Dyadic States Considering the Parental and the Infant Gender

Discriminant analyses were used to determine the linear combination of SSG variables that best classified the 52 dyads into each of the two groups: with fathers and with mothers. We established regarding previous probabilities that all groups were equal. Therefore, based on a discriminant analysis of 'Events rate,' 'Transitions rate,' and 'Events per Visit' ratio, functions were derived for the total grid. These variables were selected because they showed significant differences, and the purpose was to examine, in terms of content, what state or states (dyadic behaviors) could potentially distinguish the two type of dyads. The Eigenvalues, relative variance, canonical correlations and significance tests are shown in **Table 3**.

As **Table 3** shows, overall the Wilks Lambda value is moderately high (0.68), and the Lambda transformed value showed a statistically significant level [$\chi^2(2,N = 52) = 18.44$,

FIGURE 4 | State-space grids measures that showed statistically significant differences between father–infant and mother–infant groups. From the top left: Events rate, top right: Transitions rate, and bottom: Events per Visit ratio.
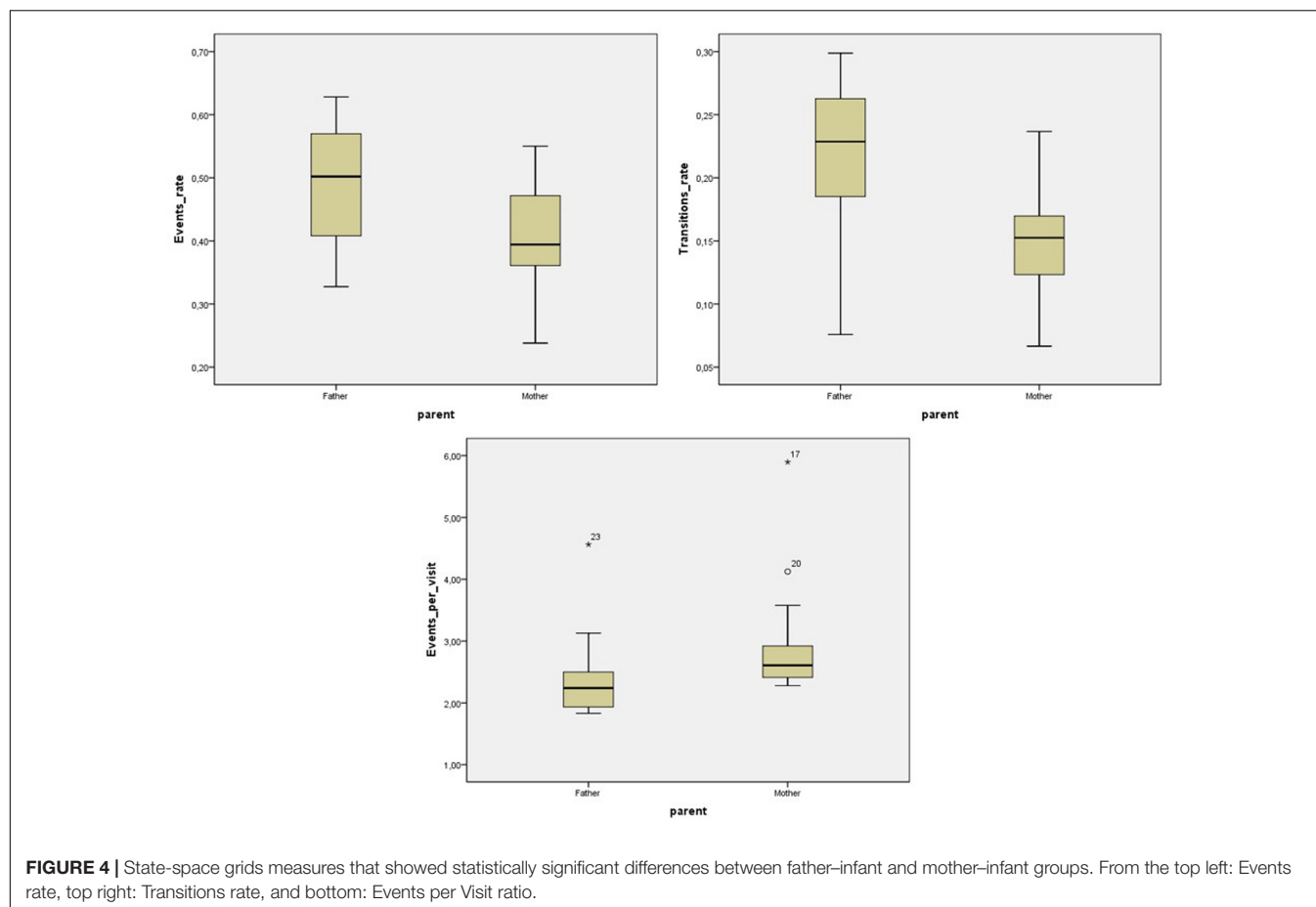
TABLE 3 | Linear Discriminant Analyses results, variables included in the stepwise discriminant analysis and summary of classification results.

Function	Eigenvalue	% variance	Canonical correlations	Wilks' lambda	Chi-square	df	p
1	0.46	100	0.56	0.68	18.44	8	0.000

Steps	Variables	Standardized coefficients	Wilks' lambda	F	df	p
1	Transitions rate AS	1.206	0.75	16.49	(1, 50)	0.000
2	Events per visits ratio JS	−0.637	0.68	11.19	(2, 49)	0.000

		Predicted group membership		Total
		Father	Mother	
Original group membership	Father	19 (73.1%)	7 (26.9%)	26
	Mother	3 (11.5%)	23 (88.5%)	26
		80.8% of original grouped cases correctly classified.		

$p = 0.000$], which supported the rejection of the null hypothesis. Therefore, the means of the father–infant dyads and mother–infant dyads on the discriminant function –the centroids, were significantly different. Likewise, the variance in the dependent variable accounted for by this model was 46%.

The stepwise discriminant analysis included two variables out of the three analyzed in the following order: 'Transitions rate' for the dyadic behavioral state 'Child Social Approach, neutral-Sensitive Approach neutral' (TR-AS) and 'Events

per Visit' ratio for 'Child Play-Sensitive Approach neutral' (ER-JS).

The standardized discriminant function coefficients indicated the relative importance of the independent variables in predicting group membership. This function was marked by a positive coefficient for TR-AS and negative weight for ER-JS. Thus, the lower the TR-AS and the higher the ER-JS the less likely it was that the dyad was from the father–infant group and more likely to belong to the mother–infant group (**Figure 5**).

FIGURE 5 | Scatter plot using the two discriminant dimensions: 'Transitions rate' for the dyadic behavioral state 'Child Social Approach, neutral-Sensitive Approach neutral" and 'Events per Visit' ratio for 'Child Play-Sensitive Approach neutral.'
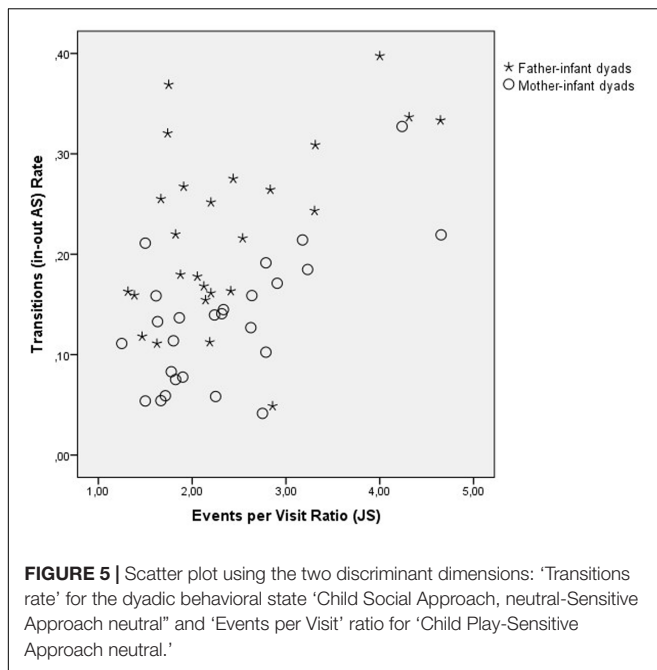
Classification Results Based on the Discriminant Function as Predictor

The classification of dyads (with fathers and mothers) based on this discriminant function for the two groups of parents, showed that 80.8% (42/52) of all the cases were correctly classified as compared with a chance classification of 50%. The function identified 73.10% (19/26) of the father–infant dyads and 88.5% (23/26) of the mother–infant dyads.

In summary, the content analyses carried out for the three NDS variables that previously showed differences between dyads involving fathers and mothers, indicated that there was no particular dyadic behaviors associated with the differences in terms of Events rate per second. However, there were particular states associated with the Transitions rate and the Events per Visit: The transitions involved the 'in–out' 'A-S' state ('Child Social Approach, neutral-Sensitive Approach neutral') and the repetitions of events involving 'J-S' ('Child Play-Sensitive approach neutral'). The combination of both in the discriminant function distinguished dyads with mothers from dyads with fathers.

DISCUSSION

Our results showed no differences between boys and girls in the parent–infant interaction, regardless of the parent's gender. This lack of differential parental interaction of boys and girls found in both groups of dyads, with fathers and with mothers, was partially congruent with the meta-analytic study results reported by Endendijk et al., 2016. They found no overall child gender-differentiated effect in parental autonomy-supportive strategies, conceptually similar to the Sensitivity construct. However, in those findings only the factor of the child's gender was considered.

Subsequently, those authors only selected the twenty-five studies that included mothers and fathers to test the parent's gender effect, and focused on the controlling strategies that had shown significant differences for the child's gender. They reported no parent gender effect in the extent of their differential treatment in controlling strategies with boys and girls. Regarding this parental control, some findings indicated that gender differences, in the use of parental control strategies are less relevant when the children are younger (Lytton and Romney, 1991; Leaper et al., 1998; Alink et al., 2006; Else-Quest et al., 2006). However, in Endendijk et al. (2016) meta-analysis the size effect was more relevant in the group aged 0–2. The infants in the present study ranged from 6 to 10 months. One possible reason to explain the discrepancy with our results could be that in the twenty-one studies that comprised their 0–2 age group, only Huber's (2012) study included children averaged under 12 months, the rest of them included toddlers, a developmental period for which parental controlling strategies are more relevant.

Other factors that need to be considered in the interpretation of our findings are the setting, i.e., free-play, and socio-economic status (SES). The free-play setting is relatively unstructured and in these settings gender differences in interaction are often lower than in structured tasks such as problem-solving (Endendijk et al., 2016). Additionally, our participants, in sociodemographic terms, were characterized by low-SES. According to biosocial theory (Eagly and Wood, 2002; Wood and Eagly, 2012) and gender schema theories (Bem, 1981; Markus et al., 1982), lower status would tend toward a more traditional division of roles which would result in a greater differentiation of gender roles that transmit into their parental practices. However, our findings did not support this. It could be that the division of gender roles has softened in the Western world (Cabrera et al., 2000; Lamb, 2010) and, as a result, has produced more egalitarian societies (Inglehart et al., 2003). In this regard, the date of publications has shown a significant association with their reported findings about differential parenting for boys and girls (Endendijk et al., 2016). The lack of child gender differences in interactions found in the present study may be due to the young age, under 10 months, of the infants and the free-play setting that is related with lower child gender differences in interaction and this was shown to be the case for both fathers and mothers.

The lack of consensus about the extent to which parents treat their sons and daughters differently can be partially explained by the wide range of child ages included in the studies, the variety of measures and observational strategies and settings. Future studies controlling these relevant factors will shed light on this particular issue.

In relation to mothers vs. fathers, the findings of the present study showed that infants interacting with their fathers and infants interacting with their mothers were involved in a similar number of different dyadic states and their interaction showed similar medium–high levels of predictability. The latter finding runs contrary to the findings of Feldman and Klein, (2003) who reported that fathers were more unpredictable than mothers were. However, the fact that the study was conducted with toddlers, in a compliance situation, could explain this discrepancy.

The differences associated with parent's gender showed that the dyads involving fathers, compared with their counterparts involving mothers, were having more back-and-forth per unit of time, i.e., Events rate and the discriminant analyses showed no particular type of event, in terms of behavioral content, distinguishing the two groups of dyads. The dyads with fathers were more active, as well, in changing from one type of dyadic state to another, per unit of time, i.e., rate of Transitions; analyses indicated that the behavioral dyadic state involved was 'A-S' ('Child Social Approach, neutral-Parent Sensitive Approach, neutral'). These findings seem to be in line with those that reported that fathers use more stimulation in terms of activating interaction (De Wolff and van IJzendoorn, 1997; Grossmann et al., 2008).

In contrast, dyads with mothers showed more repetition of the same dyadic event, once they moved into a particular state. Further analyses indicated that the visits to the behavioral state 'J-S' ('Child Play-Parent Sensitive Approach neutral') was the one where mothers were more likely to have more frequency of J-S exchanges, i.e., events, before they move to another state. Taken together the two factors, the Transitions rate of A-S and the Events per Visit ratio of JS, comprised the discriminant function that correctly classified 80.1% of the 52 dyads.

The field of father studies is receiving increasing attention in latter decades. However, the specific area of studies using observational measures of paternal interactive behavior with infants is still very limited. The focus is on comparing fathers with mothers to examine the influence of parental gender on their dyadic interaction and to considering, as well, the potential role of the infant's gender. However, progressing knowledge of paternal interactive features and sensitivity that may link with child attachment development is of particular interest for child development studies.

However, although the general sense is that gender of parents and children may affect parental behavior, the direction of those influences is not yet conclusive. Moreover, it is not conclusive, either, that the level of parental sensitivity depends, at least exclusively, on the combinations of parent–child gender (Lovas, 2005; Schoppe-Sullivan et al., 2007; Hallers-Haalboom et al., 2014; Endendijk et al., 2016). In this sense, future studies should specifically consider other factors like birth order, controlling interdependence effects between the parents.

The study of father–infant interaction using the SSG approach offers interesting possibilities because it characterizes and quantifies the actual moment-to-moment flow of infant–parent interactive dynamics. Our findings showed new facets in father vs. mother interactive behavior with their infants that can inform further developments in this field.

Limitations and Strengths

This study presents some limitations. Firstly, the groups who came from nonclinical populations included a proportion of individuals from diverse cultures; about 20% were from India, Turkey, Libya, or African countries. As caregiving behavior may be influenced by cultural factors in both fathers and mothers, some caution needs to be taken in generalizing the reported findings of the present study. Secondly, about half of the mothers and the fathers in this study reported being unemployed. No assessment of factors like depression or depressive mood, due to economic stress, was done, and this might have an effect on their interaction with their infants.

This study also presents some strength. It involved fathers and mothers, each interacting with their own infant, this controls for possible interdependence effects between the parents. The age of the infants was between 6 and 10 months, which reduces the possible interference of using a wide age range. Moreover, this age is very relevant to study interactive patterns that can be antecedents to the quality of child attachment. The number of girls and boys was the same in both groups of dyads with fathers and mothers. The participants came from a general population who joined a community-based program provided on a universal basis. Finally, the analytical approach using SSG methodology with dyadic variables allows for the characterization of different temporal features of father interaction with quantitative measures that can shed light on the paternal sensitivity construct.

AUTHOR CONTRIBUTIONS

MC, contributed to the conception of the work, acquisition of data, state-space grid analysis, interpretation of data and revising the work for intellectual content. PS-G, contributed to the design of the work, interpretation of data and revision. GP-S, contributed to the preparation of files and analyses testing the parental and infant gender factors on parent-interaction. RT, contributed to the coding data collection and analysis of the reliability of the observational measures. All authors contributed to drafting the work and gave their approval to the final version to be published. They also agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work was appropriately investigated and resolved.

FUNDING

REFERENCES

Ainsworth, M. D. S., Bell, S. M., and Stayton, D. J. (1974). "Infant–mother attachment and social development: socialization as a product of reciprocal responsiveness to signals," in *The Integration of a Child into a Social World*, ed. M. P. Richards (London: Cambridge University Press), 97–119.

Ainsworth, M. D. S., Blehar, M., Waters, E., and Wall, S. (1978). *Patterns of Attachment: A Psychological Study of the Strange Situation*. Hillsdale, NJ: LEA.

Alink, L. R. A., Mesman, J., Van Zeijl, J., Stolk, M. N., Juffer, F., Koot, H. M., et al. (2006). The early childhood aggression curve: development of physical aggression in 10- to 50-month-old children. *Child Dev.* 77, 954–966. doi: 10.1111/j.1467-8624.2006.00912.x

Alvarenga, P., and Cerezo, M. A. (2013). Interação mãe-criança: fidedignidade da versão brasileira do sistema observacional CITMI-R. *Aval. Psicol.* 12, 307–316.

Bakeman, R., and Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge: Cambridge University Press.

Barnett, M., Min, D., Mills-Koonce, R., Willoughby, M., and Cox, M. (2008). Interdependence of parenting of mothers and fathers of infants. *J. Fam. Psychol.* 22, 561–573. doi: 10.1037/0893-3200.22.3.561

Beebe, B., Jaffe, J., Markese, S., Buck, K., Chen, C., Cohen, P., et al. (2010). The origins of 12-month attachment: a microanalysis of 4-month mother–infant interaction [Monograph]. *Attach. Hum. Dev.* 12, 3–141. doi: 10.1080/14616730903338985

Beebe, B., and Lachmann, F. (2002). *Infant Research and Adult Treatment: Co-Constructing Interactions*. Hillsdale, NJ: The Analytic Press.

Bem, S. (1981). The BSRI and gender schema theory: a reply to Spence and Helmrich. *Psychol. Rev.* 88, 369–371. doi: 10.1037/0033-295X.88.4.369

Bowlby, J. (1969/1982). *Attachment and Loss*. New York, NY: Basic Books.

Braungart-Rieker, J., Garwood, M. M., Powers, B. P., and Notaro, P. C. (1998). Infant affect and affect regulation during the still-face paradigm with mothers and fathers: the role of infant characteristics and parental sensitivity. *Dev. Psychol.* 34, 1428–1437. doi: 10.1037/0012-1649.34.6.1428

Braungart-Rieker, J. M., Garwood, M. M., Powers, B. P., and Wang, X. (2001). Parental sensitivity, infant affect, and affect regulation: predictors of later attachment. *Child Dev.* 72, 252–270. doi: 10.1111/1467-8624.00277

Cabrera, N. J., Tamis-LeMonda, C. S., Bradley, R. H., Hoffert, S., and Lamb, M. E. (2000). Fatherhood in the twenty-first century. *Child Dev.* 71, 127–136. doi: 10.1111/1467-8624.00126

Cerezo, M. A. (2000). *Family Interaction. Observational Coding System Standardized Observation Codes - SOC III (English–Spanish Bilingual Edition)*. Valencia: University of Valencia Press.

Cerezo, M. A., and Pons-Salvador, G. (1999). Supporting appropriate parenting practices: A preventive approach of infant maltreatment in a community context. *Int. J. Child Fam. Welfare* 99, 42–61.

Cerezo, M. A., Pons-Salvador, G., and Trenado, M. R. (2008). Mother-infant interaction and children's socio-emotional development with high- and low-risk mothers. *Infant Behav. Dev.* 31, 578–589. doi: 10.1016/j.infbeh.2008.07.010

Cerezo, M. A., Pons-Salvador, G., Trenado, M. R., and Sierra, P. (2016). Mother-infant verbal and nonverbal interaction as a predictor of attachment: nonlinear dynamic analyses. *Nonlinear Dynamics Psychol. Life Sci.* 20, 485–508.

Cerezo, M. A., Trenado, R. M., and Pons-Salvador, G. (2006). Interacción temprana madre-hijo y factores que afectan negativamente a la parentalidad [Early mother-infant interaction and factors affecting parenting]. *Psicothema* 18, 544–550.

Cerezo, M. A., Trenado, R. M., and Pons-Salvador, G. (2012). Mother-infant interaction and quality of child's attachment: a nonlinear dynamical systems approach. *Nonlinear Dynamics Psychol. Life Sci.* 16, 237–389.

De Wolff, M., and van IJzendoorn, M. (1997). Sensitivity and attachment: a meta-analysis on parental antecedents of infant attachment. *Child Dev.* 68, 571–591. doi: 10.1111/j.1467-8624.1997.tb04218.x

Dishion, T. Y. (ed.) (2012). Developmental psychopathology: relationship dynamics [Special issue]. *Nonlinear Dynamics Psychol. Life Sci.* 16, 236–241.

Eagly, A. H., and Wood, W. (2002). A cross-cultural analysis of the behavior of women and men: Implications for the origins of sex differences. *Psychol. Bull.* 128, 699–727. doi: 10.1037//0033-2909.128.5.699

Easterbrooks, M. A., and Goldberg, W. A. (1984). Toddler development in the family: impact of father involvement and parenting characteristics. *Child Dev.* 55, 740–752. doi: 10.2307/1130126

Eddy, J. M., Leve, L. D., and Fagot, B. I. (2001). Coercive family processes: a replication and extension of Patterson's coercion model. *Aggress. Behav.* 27, 14–25. doi: 10.1002/1098-2337(20010101/31)27:1<14::AID-AB2>3.0.CO;2-2

Ekas, N., Braungart-Rieker, J., Lickenbrock, D., Zentall, S., and Maxwell, S. (2011). Toddler emotion regulation with mothers and fathers: temporal associations between negative affect and behavioral strategies. *Infancy* 16, 266–294. doi: 10.1111/j.1532-7078.2010.00042.x

Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., and Van Hulle, C. A. (2006). Gender differences in temperament: a meta-analysis. *Psychol. Bull.* 132, 33–72. doi: 10.1037/0033-2909.132.1.33

Endendijk, J., Groeneveld, M., Bakerman-Kranenburg, M., and Mesman, J. (2016). Gender-differentiated parenting revisited: meta-analysis reveals very few differences in parental control of boys and girls. *PLOS ONE* 11:e0159193. doi: 10.1371/journal.pone.0159193

Feldman, R. (2007). Parent–infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions. *J. Child Psychol. Psychiatry* 48, 329–354. doi: 10.1111/j.1469-7610.2006.01701.x

Feldman, R., and Klein, P. (2003). Toddlers' self-regulated compliance to mothers, caregivers, and fathers: implications for theories of socialization. *Dev. Psychol.* 39, 680–692. doi: 10.1037/0012-1649.39.4.680

Fleiss, J., and Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ. Psychol. Meas.* 33, 613–619. doi: 10.1177/001316447303300309

Fogel, A. (2011). Theoretical and applied dynamic systems research in developmental science. *Child Dev. Perspect.* 5, 267–272. doi: 10.1111/j.1750-8606.2011.00174.x

Fuertes, M., Faria, A., Beegghly, M., and Lopes-Dos-Santos, P. (2016). The effects of parental sensitivity and involvement in caregiving on mother-infant and father-infant attachment in a Portuguese sample. *J. Fam. Psychol.* 30, 147–156. doi: 10.1037/fam0000139

Furman, W., and Lanthier, R. (2002). "Parenting siblings," in *Handbook of Parenting*, ed. M. H. Bornstein (Mahwah, NJ: Erlbaum), 165–188.

Gergely, G. (2004). The role of contingency detection in early affect-regulative interactions and in the development of different types of attachment. *Soc. Dev.* 13, 468–478. doi: 10.1111/j.1467-9507.2004.00277.x

Goossens, F. A., and van IJzendoorn, M. H. (1990). Quality of infants' attachments to professional caregivers: relation to infant-parent attachment and day-care characteristics. *Child Dev.* 61, 832–837. doi: 10.2307/1130967

Grossmann, K., Grossmann, K. E., Kindler, H., and Zimmermann, P. (2008). "A wider view of attachment and exploration: the influence of mothers and fathers on the development of psychological security from infancy to Young adulthood," in *Handbook of Attachment: Theory, Research, and Clinical Applications*, eds J. Cassidy and P. R. Shaver (New York, NY: Guildford Press), 857–879.

Hallers-Haalboom, E. T., Mesman, J., Groeneveld, M. G., Endendijk, S. R., van der Pol, L. D., and Bakermans-Kranenburg, M. J. (2014). Mothers, fathers, sons and daughters: parental sensitivity in families with two children. *J. Fam. Psychol.* 28, 138–147. doi: 10.1037/a0036004

Hollenstein, T. (2007). State space grids: analyzing dynamics across development. *Int. J. Behav. Dev.* 31, 384–396. doi: 10.1177/0165025407077765

Hollenstein, T. (2013). *State Space Grids: Depicting Dynamics Across Development*. New York, NY: Springer.

Hollenstein, T., Granic, I., Stoolmiller, M., and Snyder, J. (2004). Rigidity in parent-child interactions and the development of externalizing and internalizing behavior in early childhood. *J. Abnormal Child Psychol.* 32, 595–607. doi: 10.1023/B:JACP.0000047209.37650.41

Huber, M. O. (2012). Interacciones tempranas y género infantil en familias monoparentales chilenas. *Rev. Latinoam. Psicol.* 44, 76–86.

Inglehart, R., Norris, P., and Rising, T. (2003). *Gender Equality and Cultural Change Around the World*. New York, NY: Cambridge University Press.

John, A., Halliburton, A., and Humphrey, J. (2012). Child-mother and child-father play interaction patterns witch preschoolers. *Early Child Dev. Care* 183, 483–497. doi: 10.1080/03004430.2012.711595

Kemppinen, K., Kumpulainen, K., Rasanen, E., Moilanen, I., Ebeling, H., Hiltunen, et al. (2005). Mother-child interaction on video compared with infant observation: is five minutes enough time for assessment? *Infant Ment. Health* 26, 69–81. doi: 10.1002/imhj.20031

Knobloch, H., Stevens, F., and Malone, A. F. (1980). *Manual of Developmental Diagnosis*. New York, NY: Harper and Row.

Kochanska, G., Aksan, N., Prisco, T. R., and Adams, E. (2008). Mother-child and father-child mutually responsive orientation in the first 2 years and children outcomes at preschool age: mechanisms of influence. *Child Dev.* 79, 30–44. doi: 10.1111/j.1467-8624.2007.01109.x/full

Kwon, K., Jeon, H., Lewsader, J., and Elicker, J. (2012). Mothers' and Fathers' parenting quality and toddlers' interactive behaviours in dyadic and triadic family contexts. *Infant Child Dev.* 21, 256–373. doi: 10.1002/icd.1746/full

Lamb, M. E. (2010). *The Role of the Father in Child Development*. Hoboken, NJ: Wiley.

Leaper, C. (2000). Gender, affiliation, assertion, and the interactive context of parent-child play. *Dev. Psychol.* 36, 381–393. doi: 10.1037/0012-1649.36.3.381

Leaper, C., Anderson, K. J., and Sanders, P. (1998). Moderators of gender effects on parents' talk to their children: a meta-analysis. *Dev. Psychol.* 34, 3–27. doi: 10.1037/0012-1649.34.1.3

Leclère, C., Viaux, S., Avril, M., Achard, C., Chetouani, M., Missonnier, S., et al. (2014). Why synchrony matters during mother-child interactions: a systematic review. *PLOS ONE* 9:e113571. doi: 10.1371/journal.pone.0113571

Lewis, C., and Lamb, M. (2003). Fathers' influences on children's development: the evidence from two-parent families. *Eur. J. Psychol. Educ.* 2, 211–228. doi: 10.1007/BF03173485

Lewis, M. D., Lamey, A. V., and Douglas, L. (1999). A new dynamic systems method for the analysis of early socioemotional development. *Dev. Sci.* 2, 458–476. doi: 10.1111/1467-7687.00090

Lillas, C., and Turnbull, J. (2009). *Infant/Child Mental Health, Early Intervention, and Relationship-Based Therapies: A Neurorelational Framework for Interdisciplinary Practice.* New York, NY: W.W. Norton.

Lovas, G. S. (2005). Gender and patterns of emotional availability in mother-toddler and father-toddler dyads. *Infant Ment. Health J.* 26, 327–353. doi: 10.1002/imhj.20056

Lundy, B. (2003). Father– and mother–infant face-to-face interaction: differences in mind-related comment and infant attachment? *Infant Behav. Dev.* 26, 200–212. doi: 10.1016/S0163-6383(03)00017-1

Lytton, H., and Romney, D. M. (1991). Parents' differential socialization of boys and girls. A meta-analysis. *Psychol. Bull.* 109, 267–296. doi: 10.1037/0033-2909.109.2.267

Markus, H., Crane, M., Berstein, S., and Siladi, M. (1982). Self-schemas and gender. *J. Pers. Soc. Psychol.* 42, 38–50. doi: 10.1037/0022-3514.42.1.38

Mesman, J. (2010). Maternal responsiveness to infants: comparing micro- and macro-level measures. *Attach. Hum. Dev.* 12, 143–149. doi: 10.1080/14616730903484763

NIDCH Early Child Care Research Network (2000). Factors associated with fathers' caregiving activities and sensitivity with young children. *J. Fam. Psychol.* 14, 200–219. doi: 10.1037/0893-3200.14.2.200

Osnat, E., and Bonnie, B. (1995). Interrelatedness of marital relations and parent-child relations: a meta-analytic review. *Psychol. Bull.* 118, 108–132. doi: 10.1037/0033-2909.118.1.108

Patterson, G. R. (1982). *Coercive Family Process: A Social Learning Approach.* Eugene, OR: Castalia.

Peck, C. (2003). Measuring sensitivity moment-by-moment: a micro-analytical look at the transmission of attachment. *Attach. Hum. Dev.* 5, 38–63. doi: 10.1080/1461673031000078661

Pleck, J. H. (2010). "Paternal involvement: Revised conceptualization and theoretical linkages with child outcomes," in *The Role of the Father in Child Development*, ed. M. E. Lamb (Hoboken, NJ: Wiley), 58–93.

Pons-Salvador, G., Cerezo, M. A., and Trenado, R. M. (2014). Dose-effect on the mothers and babies attending the programa de apoyo psicológico P/materno-infantil. *An. Psicol.* 30, 474–481. doi: 10.6018/analesps.30.2.141192

Prigogine, I., and Stengers, I. (1984). *Order out Chaos.* New York, NY: Bantam.

Quera, V., Bakeman, R., and Grisci, A. (2007). Observer agreement for event sequences: methods and software for sequence alignment and reliability estimates. *Behav. Res. Methods* 39, 39–49. doi: 10.3758/BF03192842

Schoppe-Sullivan, S., Mangelsdorf, S., Brown, G., and Sokolowski, M. S. (2007). Goodness-of-it in family context: infant temperament, marital quality, and early co-parenting behavior. *Infant Behav. Dev.* 30, 82–96. doi: 10.1016/j.infbeh.2006.11.008

Shaw, D., Lacourse, E., and Nagin, D. (2005). Developmental trajectories of conduct problems and hyperactivity from ages 2 to 10. *J. Child Psychol. Psychiatry* 46, 931–942. doi: 10.1111/j.1469-7610.2004.00390.x

Shoukri, M. (2004). *Measures of Inter-Observer Agreement.* Boca Raton, FL: Chapman &Hall/ CRC.

Simonelli, A., Parolin, M., Sacchi, C., De Palo, F., and Vieno, A. (2016). The role of father involvement and marital satisfaction in the development of family interactive abilities: a multilevel approach. *Front. Psychol.* 7:1725. doi: 10.3389/fpsyg.2016.01725

Sravish, A. V., Tronick, E., Hollenstein, T., and Beeghly, M. (2013). Dyadic flexibility during the face-to-face still-face paradigm: a dynamic systems analysis of its temporal organization. *Infant Behav. Dev.* 36, 432–437. doi: 10.1016/j.infbeh.2013.03.013

Sroufe, L. A. (1995). Emotional Development: The Organization of Emotional Life in the Early Years. New York, NY: Cambridge University Press.

Tamis-LeMonda, C., Shannon, J., Cabrera, N., and Lamb, M. (2004). Fathers and mothers at play with their 2- and 3-year-old: contribution to language and cognitive development. *Child Dev.* 75, 1806–1820. doi: 10.1111/j.1467-8624.2004.00818

Tarabulsy, G., Tessier, R., and Kappas, A. (1996). Contingency detection and the contingent organization of behavior interactions: implications for socioemotional development in infancy. *Psychol. Bull.* 120, 25–41. doi: 10.1037/0033-2909.120.1.25

Thelen, E., and Smith, L. B. (1994). *A Dynamic System Approach to the Development of Cognition and Action.* Cambridge, MA: The MIT Press.

Thelen, E., and Smith, L. B. (1998). "Dynamic systems theories," in *Handbook of Child Psychology: Theoretical Models of human development*, Vol. I, ed. W. Damon (Cambridge, MA: The MIT Press), 563–634.

Trenado, R. M., Pons-Salvador, G., and Cerezo, M. A. (2014). Interacción temprana: evaluación de la fiabilidad del sistema observacional CITMI-R, versión inglesa. *REMA Rev. Electrón. Metodol. Apl.* 19, 31–43.

Van IJzendoorn, M. H., Moran, G., Belsky, J., Pederson, D., Bakermans-Kranenburg, M. J., and Kneppers, K. (2000). The similarity of siblings' attachments to their mother. *Child Dev.* 71, 1086–1098. doi: 10.1111/1467-8624.00211

Volling, B. L., McElwain, N. L., Notaro, P. C., and Herrera, C. (2002). Parents' emotional availability and infant emotional competence: predictors of parent-infant attachment and emerging self-regulation. *J. Fam. Psychol.* 16, 447–465. doi: 10.1037/0893-3200.16.4.447

Wood, W., and Eagly, A. H. (2012). "Biosocial construction of sex differences and similarities in behavior," in *Advances in Experimental Social Psychology*, Vol. 46, eds J. M. Olson and M. P. Zanna (Burlington: Academic Press), 55–123.

Woodhouse, S. S. (2010). Dyadic interactions as precursors to attachment security: implications for intervention and research. *Attach. Hum. Dev.* 12, 151–157. doi: 10.1080/14616730903381514

Yago, S., Hirose, T., Okamitsu, M., Okabayashi, Y., Hiroi, K., Nakagawa, N., et al. (2014). Differences and similarities between father-infant interaction and mother-infant interaction. *J. Med. Dent. Sci.* 19, 7–16.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Behavioral Patterns in Special Education. Good Teaching Practices

Manuela Rodríguez-Dorta* and África Borges*

Departamento de Psicología Clínica, Psicobiología y Metodología, Universidad de La Laguna, Canary Islands, Spain

Providing quality education means to respond to the diversity in the classroom. The teacher is a key figure in responding to the various educational needs presented by students. Specifically, special education professionals are of great importance as they are the ones who lend their support to regular classroom teachers and offer specialized educational assistance to students who require it. Therefore, special education is different from what takes place in the regular classroom, demanding greater commitment by the teacher. There are certain behaviors, considered good teaching practices, which teachers have always been connected with to achieve good teaching and good learning. To ensure that these teachers are carrying out their educational work properly it is necessary to evaluate. This means having appropriate instruments. The Observational Protocol for Teaching Functions in Primary School and Special Education (PROFUNDO-EPE, v.3., in Spanish) allows to capture behaviors from these professionals and behavioral patterns that correspond to good teaching practices. This study evaluates the behavior of two special education teachers who work with students from different educational stages and educational needs. It reveals that the analyzed teachers adapt their behavior according the needs and characteristics of their students to the students responding more adequately to the needs presented by the students and showing good teaching practices. The patterns obtained indicate that they offer support, help and clear guidelines to perform the tasks. They motivate them toward learning by providing positive feedback and they check that students have properly assimilated the contents through questions or non-verbal supervision. Also, they provide a safe and reliable climate for learning.

Keywords: diversity, educational needs, special education, good teaching practices, behavioral patterns

INTRODUCTION

Diversity in education is a reality that appears before the term "attention to diversity" was coined. In every classroom, there are students who have diverse educational needs arising from different aspects. Covering them means including all educational resources to take advantage of the characteristics of these students and enhance their learning (Zavala and de la Torre, 2015).

In education, good teaching practices have always been related to what should be done to provide adequate education to achieve good learning (Del Valle Ballón, 2012). Some definitions relate good teaching practices to examples of conduct and successful procedures (Anna, 2003 is cited Cid-Sabucedo et al., 2009) or with activity that has been developed, has been evaluated and has been successful (Cid-Sabucedo et al., 2009).

Most of the extraordinary educational measures undertaken to address these schools imply an individualized programming, supports and personal and material resources. The teacher must be

supported by specialists such as teachers of Therapeutic Pedagogy (TP) (Gómez Montes, 2005). This is a discipline designed to correct development or learning dysfunctions. It is aimed at those students with temporary or permanent disability that makes more difficult for them to learn the lessons taught, or requires specialized attention to achieve the maximum level of education according to their possibilities (Decree 157/1986 of October 24 on the Management of Therapeutic Pedagogy in an Integrated System, 1986).

In the Canary Islands, students who require very specific attention and resources are cared for in the Specialized Unit (*Aulas Enclave,* in Spanish) which are located in regular centers of compulsory education (Alegre, 2000).

Teachers specialized in special education require specific competence. They must acquire this competence in their initial (Pegalajar-Palomino, 2014) and permanent education (Sykes et al., 2010; Conklin, 2012; Pegalajar-Palomino, 2014). To make their work in the classroom effective they must use, mix and adapt different strategies, resources and opportunities to the different characteristics, levels and needs of students (Ruiz Rodríguez, 2003; Martínez Geijo, 2007; Pegalajar-Palomino, 2011; González-Peiteado, 2013). Therefore, Special Education is different from the one that takes place in the regular classroom, demanding more dedication from the teacher.

The concept of good teaching practices is becoming more popular and starting to be used at a theoretical level (Bain, 2006), although it has not been defined or pointed which teaching conducts or behaviors represents this concept. The first research studies on regard this topic have been developed with university teachers (Díaz et al., 2015; Borges et al., 2016a,c). However, there is a large body of research about the effectiveness of teachers and teaching. They highlight specific aspects that are of interest, although most are not contextualized in Special Education.

Thus, teaching planning and the organization of the classroom seem to be one of the aspects of most influence the performance in different educational levels (Brophy and Good, 1986). Related to this aspect, several authors point out the importance and influence of an appropriate structure of the lessons to help students in their cognitive development and improve their performance (Mortimore et al., 1988; Renkl and Helmke, 1992; Brown, 2009; Hunt et al., 2009; Orlich et al., 2010).

With regard to the transmission of content by the teacher, research highlights the importance of clearly explaining the objectives of each lesson (Melton, 1978; Cotton, 1995). Other authors point out other aspects which are also relevant, such as the presentation of information in an organized way, pointing out transitions to new topics, using a variety of examples and frequent reminder of the essential principles (Maddox and Hoole, 1975; Smith and Cotton, 1980; Kallison, 1986; Mayer and Gallini, 1990; Hiebert et al., 1991).

Other studies indicate aspects are related with student's motivation. This is based on psychoeducational principles from which arise various strategies related to arouse and maintain attention, generate cognitive dissonance and positive expectations for learning (Hernández, 1991; Hernández and García, 1995 is cited Hernández-Jorge, 2005). Strategies such as individualized attention, reinforcement, frequent supervision,

public and private praises, etc. (Murillo et al., 2011) and the use of diverse activities to adapt to different moments, circumstances, students, etc. (Dalton, 2007; Hunt et al., 2009) are the best for the student to learn. Another important aspect is the use of activities that require the active participation of students. In this sense, the traditional "learning by doing" strategy is still effective (Muijs and Reynolds, 2001).

With respect to the climate generated in class and that is given by the interaction between teacher and students, it is important that the students feel comfortable to participate in the activities (Muijs and Reynolds, 2001). Simple aspects such as greeting or asking about general aspects of the students at the beginning of the class make students feel safe and comfortable (Hernández-Jorge, 2005).

Also, research refers to aspects related to the classroom's discipline. It has been found that if teachers apply an adequate control method to the unwanted behaviors of the students in the classroom, they will obtain the maximum benefit from teaching (Omoteso and Semudara, 2011). In this sense, the most successful teachers in the control of their class are those who, among other aspects, adequately analyze the different disturbing stimuli of the classroom, use simple rules and make explicit to students and use clear behavioral indicators both verbal and non-verbal (Hernández-Jorge, 2005).

Another aspect that is highlighted by research and that favors a quality education is the evaluation and continuous monitoring of the students (Daloz, 1986; Stronge et al., 2004; Killen, 2005; Brookhart, 2009; Orlich et al., 2010; Murillo et al., 2011). Evaluation is relevant as a way to identify whether students are getting or not proper results (Muijs and Reynolds, 2001; Anderson, 2004; Hattie and Timperley, 2007; Murillo et al., 2011). To do this, the teacher can ask questions or observe the work the student is performing (Pellicer and Anderson, 1995). However, the assessment should not stop here, but the teacher must motivate the students with feedback during their learning process (Muijs and Reynolds, 2001; Anderson, 2004; Hattie and Timperley, 2007; Murillo et al., 2011; Pegalajar-Palomino, 2011), which is also a way to motivate them.

With respect to the orientation that the teacher offers students in their learning process, research emphasizes it is important that the teacher solves any doubt that arises in the students to facilitate the adequate understanding of new concepts (Berliner, 1983; Anderson, 1989, 2004). Provide support ("Scaffolds") to students so that they can carry out the activities (Palincsar and Brown, 1984; Van de Grift, 2007) is important. Also, the research notes that, among other things, presenting new material in small steps, providing clear and detailed instructions (Rosenshine, 1979), asking questions and providing guided practice has been proved effective, especially with younger students and those who have lower academic abilities than the rest of their age group (Muijs and Reynolds, 2003; Houtveen et al., 2004; Houtveen and Van de Grift, 2006).

Additionally, it is important to highlight it is essential to take into consideration the diversity of needs and abilities within a classroom in order to ensure an education of good quality. This means answering the different needs, expectations, abilities, characteristics, motivations, cultures,

previous knowledge, learning rhythms, etc., present in the students (Murillo et al., 2011). Also, to consider activities that are challenging for all students is important (Houtveen et al., 1999; Ainscow et al., 2001; Ainscow, 2007).

To verify that these aspects are taking place is important to evaluate the educational work of the teacher in the classroom. For this reason it is essential to have instruments that are not only accurate, valid and reliable, but it is also important that the instruments are sensitive and allow capturing the differential behavior of teachers in their attention to students with different characteristics and needs. The assessment tools used to operationalize the quality of teachers to measure their teaching performance are various, such as teacher's materials, briefcases, questionnaires and surveys, interviews and observation (Jiménez, 1999; García and Congosto, 2000; Rodríguez and Ibarra, 2013).

The classroom observation is a proper procedure because it implies a good way of knowing the reality of the classroom, thus allowing us to dig deeper into the exchanges that occur within it every day (Mayorga and López, 2005), and provide evidence to identify areas that should be improved (Pianta and Hamre, 2009).

The objective of this study is to analyze the behavior of two teachers from Special Education in their interaction with students from Preschool and Primary Education who present different educational needs. The aim is to determine whether there are differences on their teaching practices depending on the students with who they interact and their needs. In the first place, we assume that the behavior of Special Education teachers during their educational work in the classroom is adjusted to the characteristics of students with special educational needs. Secondly, we assume that to the extent that the behavior of these teachers is in line with the characteristics of their students, this is indicative of good teaching practices.

MATERIALS AND METHODS

Methodology and Design

For the determination of behavioral patterns in the classroom, observational methodology has been used, whose design, taking into account the axes referred to units of study, temporality of registration and dimensionality (Anguera et al., 2001), is nomothetic, monitoring, and multidimensional.

Participants

The participants in this study have been selected intentionally. Given the institutional nature of this research, we contacted the Canary Islands Government Education Department to request authorization for collecting the information. They proposed two schools to us. Each school proposed a teacher of Special Education. These two teachers teach children of different educational stages. The first one works in TP, he is 52 years old and 22 years of experience. The second teacher works in Specialized Unit, he is 53 years old and 18 years of experience.

In order to collect the continuous flow of the behaviors that take place in the classroom, it is necessary to analyze the students' behaviors. For this reason, students' are studied in their interaction with the teacher. This research involves a

TABLE 1 | Participating students per teacher.

Teacher	Students	Educational stage	Age of students
Therapeutic Pedagogy	1	Preschool Education	5–6 years
	3	Primary School	8–9 years
Specialized Unit (*Aula Enclave*)	7	Preschool Education	5–6 years

total of 11 students of different educational stages that present specific educational needs and require a different educational response. The teacher of TP attends both students of Preschool Education and students of Primary School, while the teacher in the Specialized Unit works only students of Preschool Education. **Table 1** shows the number of students corresponding to each teacher, the educational stage in which they are and their ages.

Three observers participated in the coding of the videos. Two of them are students of Psychology and one has a Master in Educational Psychology.

Instruments

Observational Protocol for Teaching Functions in Primary School and Special Education (PROFUNDO-EPE, v.3., in Spanish) (Rodríguez-Dorta, 2015)

The observational instrument was based on the Protocol of Teaching Functions (PROFUNDO v2, in Spanish) (Rodríguez-Naveiras, 2011), created for the observation of teaching functions in an extracurricular program of psychoeducational intervention. PROFUNDO v2 is based on the model of teaching functions of Hernández-Jorge (2005).

The PROFUNDO v.2 has been adapted for the study of the educational functions in different educational stages: to university education, obtaining the Protocol of Observation of Teaching Functions in University (PROFUNDO-UNI, v.2, in Spanish) (Díaz, 2014) and Primary and Special Education, obtaining the instrument that is presented below, Protocol of Observation of Teaching Functions in Primary and Special Education (PROFUNDO-EPE, v.3, in Spanish) (Rodríguez-Dorta, 2015; **Table 2**).

The PROFUNDO-EPE, v.3 it is based on the Teaching Functions Model and analyzes those functions which can be directly observed: (a) organization function: planning of education and control over the context; (b) teacher's communicability function: teacher's ability to communicate the contents so that they are understood by students; (c) motivation function: teacher's ability to encourage students to learning; (d) behavior control function: group regulation, order and discipline; (e) orientation and advice function: guide students in their learning; (f) interaction function: teacher–student relationship to generate motivation, correct mistakes, and expand the information is working; (g) evaluation function: propose criteria to check whether the learning objectives have been achieved and whether the teaching process has been adequately performed.

The **Organization Function** is included in the code *Teacher's Organization* (TO). This code gathers all the tasks that the teacher develops in the classroom and has an organizational component

MACROCATEGORIES Teacher's Functions	CODES	
1. Organization	Teacher's Organization	TO
2. Teacher's Communicability	Teacher's Explanation	TE
3. Motivation	Reinforcement	RF
	Motivation	MO
4. Behavioral Control	Control	CL
5. Guidance and Advice	Guidance	GU
	Non-verbal Revision	NR
6. Interaction	General Interactions	GI
	Use of the Diary	SD
7. Evaluation	Verbal Revision	VR
8. Students' Interactions	Students Participation	SP
	Answer the Teacher	AT
	Classroom Disruption	CD
	General Interactions	GI
Instrumental Categories	Other Behaviors	X
	Unobservable	Y
	Teacher Leaves the Classroom	Z

(organization of teaching material, planning and structuring of the class and organization of students to work).

The **Teacher's Communicability Function** is collected through the code *Teacher's Explanation* (TE). This code represents the theoretical expositions of the teacher.

The **Motivation Function** is composed by two codes. The *Reinforcement* (RF) code formed by verbalizations or non-verbal teacher behaviors aimed at positively reinforcing students' behavior. The *Motivation* (MO) code represents teacher verbalizations that allow students to choose on some aspect of the task, verbalizations aimed at generating motivation toward the task or activity, verbalizations that highlight some aspect of the task to be performed or the anticipation of reward.

The **Behavior Control Function** is included in the *Control* (CL) code. This code is formed by all the teaching behaviors of criticism, threat, or punishment directed to the students.

The **Orientation and Advice Function** is included in the *Guidance* (GU) and *Non-verbal Revision* (NR) codes. The first one contains the indications, questions, clues and corrections directed to the students during the accomplishment of an activity. Also, collect the teacher's answers to questions raised by the students regarding what is being worked in class. The second includes the silent supervision of the activity or task of the students by the teacher.

The **Interaction Function** includes two codes. The *General Interactions* (GI) code formed by the verbalizations of the teacher directed to the students that are not related to the activity, task or theoretical content that is being worked. The code *Use of the Diary* (SD) formed by the behaviors of the teacher consisting of use the student's diary to communicate with their families.

Finally, the **Evaluation Function** is composed by the *Verbal Revision* (VR) code. This code consists of all those questions of the teacher directed to check if the students have acquired the contents or aspects worked.

The observation protocol also includes an eighth category to observe the student behavior when they interact with the teacher. Also, the protocol includes a last **instrument category** to complete the constant flow of the instructors' behavior and which collects other types of behavior which are not connected to the typified behaviors.

Thus, the category **Interventions of the Students** has four codes. Two of them include the positive interventions of the students. *Students Participation* (SP) formed by the interventions of the students on their own initiative directed to the teacher to present some idea or opinion, ask a question, etc. related to what is being worked on class. *Answer the Teacher* (AT) formed by the students' answers (verbally or through behavior) to a question, approach, comment, indication, etc. of the teacher referring to what is being worked on class. Negative interventions are listed in the *Classroom Disruption* (CD) code. This code consists of verbal or non-verbal behaviors of the students who are against the logical norms of the classroom (hitting or insulting a classmate, getting up and making noise in the middle of the class, breaking up material, etc.) that interrupt the rhythm of the class and call the attention of the teacher to apply a control. The GI code includes the neutral interventions of the students. This code represents students' verbalizations addressed to the teacher that are not related to the activity or content that is being worked on class.

The **Instrumental category** includes three codes. The code *Other Behaviors* (X) that is formed by those behaviors of the teacher that do not correspond with the teaching functions (to speak with a person who enters the classroom, to speak by the mobile, etc.). The code *Unobservable* (Y) formed by those moments in which it is impossible to observe or codify the behavior of the teacher. The code *Teacher Leaves the Classroom* (Z) that collects those moments in which the teacher leaves the classroom.

As it was mentioned before, this observational instrument is part of the instruments of evaluation of teacher behavior, having been cross-validated. The PROFUNDO v2 (Rodríguez-Naveiras, 2011) has been applied to evaluate instructors of extracurricular programs for students with high ability in Canary Islands and two Mexican States. The PROFUNDO-UNI v2 (Díaz, 2014) was applied to teachers in Canary Islands and Mexico.

Operationalization of Good Teaching Practices in Special Education

Given the particular characteristics of teaching Special Education, the PROFUNDO-EPE, v.3 is used to extract behavior patterns that correspond to good teaching practices in this area (**Table 3**), which are explained below.

After arranging the necessary materials for class or giving action guidelines to students to begin working (TO), it is important to motivate them by presenting, for example, the activity as something attractive or entertaining (MO). When the teacher presents the content in a theoretical way (TE) it is appropriate to continue to give guidance to students, organizing them to work and implement what has been explained (TO). The teacher should also encourage the students to work (MO), or try to check whether students have assimilated adequately

TABLE 3 | Behavioral patterns considered good teaching practices.

Criteria Conduct	Consequent Conducts
Teacher's Organization (TO)	- Motivation (MO)
Teacher's Explanation (TE)	- Teacher's Organization (TO) - Motivation (MO) - Verbal Revision (VR)
Guidance (GU)	- Motivation (MO) - Non-verbal Revision (NR) - Verbal Revision (VR)
Non-verbal Revision (NR)	- Reinforcement (RF) - Guidance (GU)
Students Participation (SP)	- Reinforcement (RF) - Guidance (GU) - Non-verbal Revision (NR) - Verbal Revision (VR)
Answer the Teacher (AT)	- Reinforcement (RF) - Guidance (GU) - Non-verbal Revision (NR) - Verbal Revision (VR)
Classroom Disruption (CD)	- Control (CL)
General Interactions (GI) of teacher	- General Interactions (GI) of students
General Interactions (GI) of students	- General Interactions (GI) of teacher

the content that was just explained (VR). Also, it is a good teaching practice when after assistances or instructions on how to perform a task (GU), the teacher encourages the students to work (MO), supervises in silence their work (NR) or checks that students have understood the instructions (VR). When the teacher is supervising in silence the work of the students (NR), it is appropriate in this context to give them positive feedback when they are doing their job correctly (RF), or to provide them with aid or instruction when they are not doing activities properly (GU).

When there are positive interventions of students, either on their own initiative (SP) or in response to the teacher (AT), good practices that teachers apply are positive feedback (RF), orient the students (GU), or supervise their interventions in silence (NR), or through questions (VR).

Also, it is appropriate that teachers apply a negative contingency (CL) after a disruptive behavior of the student (CD).

Finally, the GI of the teacher followed by the GI of the students or vice versa are good teaching practices. On the one hand, it contributes to the generation of a safe and reliable climate. On the other, it represents a contrast that facilitates the student's moments of disconnection, allowing them to return to the task with a higher level of attention.

Instruments of Registration and Coding

The teacher's behaviors are registered with two different video cameras: JVC and Sony.

For the coding process we used software of the Augenv.δ program for the evaluation of behaviors (Montero and Montero, 2012, unpublished).

Procedure

First, we obtained the signed informed consent from the student's parents and of from the teachers to record them. After, we proceeded to film 20 class hours for each teacher in the period between January and February 2011. The recording of the 20 h was carried out over several weeks being framed within the same quarter. This allowed us to collect a sufficient sample of teacher behavior in their educational work in the classroom.

In the TP classroom the students attended certain days, the rest is developed in the regular classroom with the corresponding teacher. The TP teacher attends Preschool Education students in a few days and Primary School students on other days. Therefore, we proceeded to record the days in which this teacher taught classes to the student of Preschool Education and the days in which he imparted classes to the students of Primary School. We recorded the entire period of the sessions.

In the Specialized Unit, the teacher attends students throughout all day. These students require special support in all or most of the areas or subjects of the curriculum. The space of the where the Specialized Unit is located has special conditions. It's not ordinary classroom, it has different spaces (kitchen, adapted bathroom, work room, mathematical room, and psychomotricity room). This place is adapted to the needs of the students. We couldn't record in all the space of this classroom; therefore we did into on consecutive days and only when the students were in the work room or in the math room.

For the selection of the sessions, two criteria were taken into account: we deleted the first 2 h of recording of each teacher to avoid the reactivity bias and selected those sessions where the teachers could be easily observed, so that their coding could be carried out without difficulty.

Observers signed a confidentiality agreement and were trained following a standardized procedure (Rodríguez-Naveiras, 2011; Cadenas et al., 2012; Díaz, 2014; Rodríguez-Dorta, 2015).

We determine the number of sessions to code through optimization study of the Theory of Generalizability (TG) (Blanco-Villaseñor, 1991; Blanco et al., 2000, 2010). In the case of the teacher of Therapeutic Pedagogy, since he attends students from both Preschool Education and of Primary School, we did the optimization for these two situations. However, we incorporated more sessions of the optimization study identifies as necessary to ensure stability of the behavioral patterns of teacher (Rodríguez-Dorta and Borges, 2015b).

The length of the sessions which were selected for the coding process was established using GT. A period of time is consider optimal when the generalizability coefficient achieves a value higher than 0.90. A total of four sessions of 25 min each were coded from the teacher of TP when he was interacting with students from Preschool. Three sessions of 20 min each were coded in the case of the students from Primary school. For the teacher of Specialized Unit, four sessions of 10 min each were coded.

Data Analysis

The analysis of inter-observer reliability was conducted through the Cohen's Kappa Coefficient (Cohen, 1960, 1968), using the

program of statistical analysis SPSS v.15 and through the Generalizability Theory (GT; Cronbach et al., 1972), using the programs EduG 6.0 and SAGT v.1.0.

To obtain behavioral patterns we used the sequential lag analysis through the program GSEQ v.5.1. (Bakeman and Quera, 1996), based in determining if one conduct follows another with a higher probability than the one expected randomly. We take a previous conduct or criteria from which we count the times that other (consequent) conducts follow it immediately after, with a first lag and after two conducts in a second, giving a positive or excitatory dependency when the value of Z is higher than 1,96. To determine the extent of the association between behaviors we have calculated the coefficient Yule's Q (Yule and Kendal, 1957 is cited Lloyd et al., 2013) through the program GSEQ v.5.1.

Optimization Study

The observational methodology is very costly in time and resources, which makes necessary to have valid, accurate and reliable instruments and procedures in place to implement it efficiently. In this sense, the Decision Study of the GT (Blanco-Villaseñor, 1991; Blanco and Anguera, 2000; Blanco et al., 2000, 2010) is especially useful because it allows to determine what the minimum of sessions and time to code are (Borges and Rodríguez-Dorta, 2015; Rodríguez-Dorta, 2015; Rodríguez-Dorta and Borges, 2015a,b).

Because this process involves the generalization of the same behaviors collected from a context and particular circumstances, this work was carried out for each of the contexts studied. First, the optimization session time is performed and secondly the number of sessions.

Optimizing the number of sessions to encode is usual in observational methodology. However, when the sessions are long-lasting, it is also important to optimize the time of encoding sessions. Since the GT works with discrete variables and time is a continuous variable, we have taken as units of time variable sections of 5 min, choosing three consecutive sections for each context. Next, a trained observer encoded these time sections (Borges and Rodríguez-Dorta, 2015; Rodríguez-Dorta, 2015; Rodríguez-Dorta and Borges, 2015a). Then, we performed a decision study of the time sections. The results allow to conclude that the duration of the session is relatively short in the three contexts, with a maximum of 25 min (**Table 4**).

TABLE 4 | Time sections optimization.

Unifaceta Crossover Design TxC. Random Estimation Plan. Measurement Plan C/T.			
Teacher	**Coef. G Rel.**	**Coef. G Abs.**	**Number of sections**
Therapeutic Pedagogy Preschool	0.912	0.906	5 (25 min)
Therapeutic Pedagogy Primary	0.916	0.914	4 (20 min)
Specialized Unit (*Aula Enclave*)	0.956	0.953	2 (10 min)

TABLE 5 | Optimization of the sessions.

Unifaceta Crossover Design TxC. Random Estimation Plan. Measurement Plan C/T.			
Teacher	**Coef. G Rel.**	**Coef. G Abs.**	**Number of sessions**
Therapeutic Pedagogy Preschool	0.970	0.970	2
Therapeutic Pedagogy Primary	0.901	0.901	2
Specialized Unit (*Aula Enclave*)	0.983	0.981	2

Afterward, we set the number of sessions to encode. We selected two sessions with the optimal duration of each context. A trained observer performed the coding of these sessions. Later, we made a decision study for the sessions. The results show that two sessions are needed in all contexts (**Table 5**).

To check whether the contribution of the decision study is sufficient, we began codifying the number of sessions set for each context and we continued to include sessions on the saturation criterion. In order to do this, we stopped to encode sessions when new relevant behavioral patterns do not appear with the addition of sessions. This procedure allows us to ensure stability of the behavioral patterns in the contexts studied (Rodríguez-Dorta and Borges, 2015b).

We found that coding time established by the decision study is not enough because significant new patterns take place with the inclusion of sessions. Therefore, for TP with students of Primary School we included one more session of the decision study identified as necessary, while for TP with students of Preschool Education and for Specialized Unit it was necessary to include two more sessions.

Inter-observer Reliability

Observation requires monitoring. So, following the criterion of Patterson (1982), the reliability was calculated at 20% of the encoded session (a session from each context: TP with students of Primary School, TP with students of Preschool Education and Specialized Unit with students of Preschool Education). We calculated the reliability of each observer with expert observer, through Kappa and TG, and among all with TG. In order to avoid bias, the observers were unaware of when reliability would be calculated. Values ranged between 0.83 and 1 for the Kappa index and 0.92 and 0.98, in TG, being suitable all indices obtained (Fleiss, 1981; Bakeman and Gottman, 1986; Hintze and Matthews, 2004).

RESULTS

Analysis of Behavioral Patterns

The differences on the behavior of teachers based on the context in which they perform their educational work are evident through the behavioral patterns that take place. Then, significant excitatory patterns in the first and second lag for

each criterion behavior in each educational context were shown. The residual value of the consequent conducts is in parentheses and the value of Q of Yule is in the next column. Those patterns that were considered good teaching practices are bolded.

When the criterion behavior is TO, this conduct is followed by TE or AT in Therapeutic Education with students of Primary School in the first lag. In Specialized Unit with students of Preschool Education, TO is followed by interventions of students on their own initiative (*Student Participation*, SP) or responses to the question posed by the teacher (AT) at first lag. In the second lag, it is followed by TO in both teachers (**Table 6**).

When the code TE was taken as a criterion behavior (see **Table 7**), meaningful patterns were obtained in TP with students of Preschool Education. In the first lag, this behavior is followed by interventions of students on their own initiative (SP) or a verbal verification by the teacher (VR). It is appropriate that teachers try to check that students understand and properly acquire the content they are working. In the second lag, TE is followed by aids or instructions of teacher (GU) or by students responses to the question posed by the teacher (AT).

In **Table 8**, the results when the behavior criterion is RF are presented, finding meaningful patterns in TP with students of Preschool Education and in Specialized Unit with student Preschool Education. In the first case, RF is followed by TE or VR, in first lag. In the second case, RF is followed by TO or GU.

When the criterion behavior is GU, a general pattern occurs in three cases. In the first lag, this behavior is followed by positive interventions by the student, SP and AT. In the second lag teachers continue to guide students in their learning, GU. Also, in the case of TP with students of Preschool Education, in the second lag, GU is followed by NR and in Specialized Unit by RF (**Table 9**).

On the other hand, when the criterion behavior is NR, significant patterns only occur in TP with students of Preschool Education in the first lag. The teacher applies a positive feedback (RF) or guides the task to verify if it is not being done well (GU). These behavioral patterns correspond to good teaching practices (**Table 9**).

The patterns obtained when the criterion behavior is VR are presented in **Table 10**. In this case a general pattern occurs in the three cases in the first lag. The criterion behavior is followed by AT.

In the second lag, in the context of TP with students of Preschool Education and Primary School, the teacher continues VR, monitoring the correct assimilation of the contents they are working.

It is also important to note that teacher continues to apply a positive feedback to the student's response of Preschool Education, RF, while with students of Primary School, the teacher continues to guide the students, GU.

When the criteria are the positive interventions of students (see **Table 11**) significant patterns are observed in all three cases. In the first lag, SP is followed either by TO or by GU. The latter behavioral pattern is a pattern of good teaching practices.

Interventions of the students on their own initiative are followed by RF only by the teacher of TP with students of Primary School. This behavioral pattern is also a good teaching practice. The code SP includes those interventions of students that are made on their own initiative, without distinguishing them by their content. This code includes interventions related to curriculum content or other related to organizational aspects. In the case of students of TP of Primary School it is more likely to perform interventions related to the contents they are working, because they cover more difficult content and they are older.

On the other hand, when the criterion behavior is AT, it is followed in all cases by RF, or GU in the first lag. GU and RF are necessary behaviors so that students can progress in their learning. Therefore, these behaviors are good teaching practices.

Also, in TP with both students of Preschool Education and students of Primary School AT is followed by *Review Verbal* (RV) and in the case of Therapeutic Pedagogy with students of Preschool Education in particular, it is followed by NR. These behavioral patterns are also patterns of good teaching practices. In the second lag, in TP with students of Preschool Education, it is followed by either TE or VR. The latter pattern is suitable as the teacher continues to check the proper assimilation of the contents.

Regarding the negative behaviors of students, when the criterion behavior is CD significant patterns are found in Therapeutic Education with students of Primary School. In this context, in the first lag, CD is followed by CL. This behavioral pattern is a good teaching practice (**Table 12**).

Finally, conversations about issues not related to the content given in class are also had in the classroom. These conversations are collected with the GI code. The goal of these interactions is to create a safe and reliable climate, stimulate or generate a contrast that allows students to relax and regain the level of attention when the teacher returns to class. This type of behavior is more expected in TP. Here, the curriculum addressed requires moments of relaxation to allow the students to maintain the appropriate level of attention. However, in Specialized Unit, by the characteristics of the students, they work basic and everyday contents which are approached from a ludic point of view. Therefore, the GI fails to meet its goal of stimulating contrast to become curricular content.

Significant behavioral patterns of GI by the teacher are presented in **Table 13**. In the first lag, these interactions have a response by the students in TP with both students from Preschool Education and Primary School. In the second lag, the teacher continued the conversation again with GI in response to the interventions of students.

Significant behavioral patterns of GI by the students are presented in **Table 14**. In the first lag, these interactions are answered by the teacher and, in the second lag, the conversation has continued again with GI of students.

These conversations are good teaching practices when their aim is to create a safe and trustworthy climate for the students or to create a stimulating contrast.

DISCUSSION

The challenge of education is to offer it in terms of quality and equity for all. One of the important aspects of this is to address and respond to the different educational needs presented by students. This requires assessing, among other things, the behavior of teachers in their professional performance in the classroom, as they are responsible to teach and make students learn (Hernández, 2006). Specifically, teachers dedicated to Special Education have special relevance in attention to this diversity in the classroom.

As we have seen in this work, the observational methodology offers us a lot of information but it is important to apply it properly. It is decisive to use valid, accurate, reliable observation instruments. In addition, due to the cost to the time and resources that can be assumed by this methodology, we need to have procedures which allow us to apply it efficiently. In this work we have begun using mathematical procedures as optimization GT (Blanco-Villaseñor, 1991; Blanco and Anguera, 2000; Blanco et al., 2000, 2010) to establish the minimum of codifying time necessary to obtain accurate information that allows its generalizability. From the results obtained in the optimization, we include new sessions. Thus was to ensure the stability of the conduct of the teachers we studied (Borges and Rodríguez-Dorta, 2015; Rodríguez-Dorta and Borges, 2015b).

The behavioral analysis of the teachers from this study allows us to see that they adapt their behavior according to the students with whom they work, giving the most adequate response to their needs.

As it was mentioned, the teacher of TP works with both students from Preschool and Primary Education. The results obtained indicate his behavioral patterns changes depending on the students with who he is working at every moment. Therefore he adapts his way of teaching according to the needs and characteristics of his students. The teacher from Specialized Unit also shows different behavioral patterns.

In the case of the teacher from the Specialized Unit and one of TP when he works with children from Primary School, there are significant patterns in TO follow by the response of the students, AT, and again TO. In addition, in the case of TP with students of Primary School to TO is followed, in the first lag, TE, which

is logical since the contents of this educational stage require theoretical explanations.

Non-verbal Revision produces significant patterns only in the case of the TP teacher when teaching classes to students of Preschool Education. Thus, after the NR applies RF to the student or continues to guide him on his learning (GU). On the other hand, the student's response to questions or approaches of the teacher receives non-verbal supervision (NR). Since there is only one student from this educational stage, the supervision is more intensive and continuous and the teacher is totally focused on him.

The code VR also produces a general pattern in all the cases (VR–AT). However, it is in the case of the teacher of TP where the teacher continues checking to if the students are understanding what they are working on after a response from the students (AT) or a theoretical explanation (TE). This verification process is of great importance in TP. However, in Specialized Unit the priority is another. This discipline supposes reinforcement for students with special educational needs whose objective is to get them to acquire the same knowledge as their regular classmates. However, in Specialized Unit, more than the verification of the acquisition of content, what is really important is to motivate and to guide the students to perform their tasks. Here, due to the educational needs presented by the students, the objective is they become more independent and they again autonomy, moving away from the contents raised with their regular classmates.

Regarding the positive interventions of the students, the answers of the students (AT) to questions and approaches of the teacher, receive both RF and GU by the teacher in all the cases. However, the SP on its own initiative only receives GU in TP with students from Preschool Education and Primary School. RF is obtained in TP with students from Primary School in particular. It is possible that the interventions on the student's own initiative refers to what they are doing, but on the other two cases, due to the age and characteristics of the students, they refer more to aspects of organizational and general character.

It is important to note how both teachers only use reinforcement to generate motivation. However, there are no behavioral patterns where the teacher encourages the students to the task by highlighting some positive aspect of it, show anticipation of reward or encourage the students to participate.

TABLE 6 | Organization function.

Teacher	Criteria Conduct	Consequent Conducts			
		Lag 1		Lag 2	
		Codes	Yule's Q	Codes	Yule's Q
Therapeutic Pedagogy Preschool	TO	–		TO (9.10)	0.82
Therapeutic Pedagogy Primary		**TE** (4.28)	0,68	**TO** (5.38)	0.56
		AT (2.85)	0,33		
Specialized Unit (*Aula Enclave*)		SP (2.35)	0,33	TO (4.77)	0.53
		AT (2.68)	0,34		

TO, Teacher's Organization; TE, Teacher's Explanation; SP, Student Participation; AT, Answer the Teacher.

TABLE 7 | Teacher's Communicability Function.

Teacher	Criteria Conduct	Consequent Conducts			
		Lag 1		Lag 2	
		Codes	Yule's Q	Codes	Yule's Q
Therapeutic Pedagogy	TE	**SP** (9.04)	0.81	**GU** (3.95)	0.49
Preschool		**VR** (5.57)	0.61	AT (3.61)	0.44

TE, Teacher's Explanation; SP, Student Participation; VR, Verbal Review; GU, Guidance; AT, Answer the Teacher.

TABLE 8 | Motivation Function.

Teacher	Criteria Conduct	Consequent Conducts			
		Lag 1		Lag 2	
		Codes	Yule's Q	Codes	Yule's Q
Therapeutic Pedagogy	RF	TE (5.88)	0.66	AT (8.13)	0.56
Preschool		VR (7.85)	0.56		
Specialized Unit (*Aula Enclave*)		TO (2.55)	0.35	–	
		GU (3.63)	0.44		

RF, Reinforcement; TE, Teacher's Explanation; VR, Verbal Review; TO, Teacher's Organization; GU, Guidance; AT, Answer the Teacher.

TABLE 9 | Guidance and Advice Function.

Teacher	Criteria Conduct	Consequent Conducts			
		Lag 1		Lag 2	
		Codes	Yule's Q	Codes	Yule's Q
Therapeutic Pedagogy	GU	TE (3.19)	0.42	**GU** (10.07)	0.60
Preschool		**SP** (8.70)	0.66	**NR** (3.89)	0.52
		AT (7.05)	0.43		
Therapeutic Pedagogy Primary		**SP** (2.80)	0.40	**GU** (4.24)	0.54
		AT (4.23)	0.49		
Specialized Unit (*Aula Enclave*)		**SP** (2.95)	0.38	**RF** (4.82)	0.54
		AT (7.85)	0.71	**GU** (4.35)	0.45
Therapeutic Pedagogy	NR	**RF** (4.11)	0.57	–	–
Preschool		**GU** (3.48)	0.48		

TE, Teacher's Explanation; SP, Student Participation; AT, Answer the Teacher; RF, Reinforcement; GU, Guidance; NR, Non-verbal Revision.

TABLE 10 | Evaluation Function.

Teacher	Criteria Conduct	Consequent Conducts			
		Lag 1		Lag 2	
		Codes	Yule's Q	Codes	Yule's Q
Therapeutic Pedagogy	VR	**AT** (27.37)	0.96	**RF** (14.97)	0.81
Preschool				**VR** (6.85)	0.43
Therapeutic Pedagogy		**AT** (16.00)	0.97	**GU** (3.54)	0.48
Primary				**VR** (7.46)	0.74
Specialized Unit (*Aula Enclave*)		AT (7.11)	0.89	–	–

AT, Answer the Teacher; RF, Reinforcement; VR, Verbal Review; GU, Guidance.

TABLE 11 | Positive Interventions of Students.

Teacher	Criteria Conduct	Consequent Conducts			
		Lag 1		Lag 2	
		Codes	Yule's Q	Codes	Yule's Q
Therapeutic Pedagogy	SP	**GU** (14.60)	0.85	TE (4.18)	0.58
Preschool				SP (10.96)	0.79
Therapeutic Pedagogy		TO (8.07)	0.75	SP (2.57)	0.38
Primary		**RF** (4.36)	0.68		
		GU (5.48)	0.63		
Specialized Unit (*Aula Enclave*)		TO (4.78)	0.56	–	–
Therapeutic Pedagogy	AT	**RF** (17.90)	0.91	TE (2.32)	0.30
Preschool		**GU** (6.55)	0.40	**VR** (2.47)	0.16
		NR (2.87)	0.40	AT (9.72)	0.51
		VR (4.32)	0.27		
Therapeutic Pedagogy		**RF** (3.39)	0.45	AT (7.85)	0.67
Primary		**GU** (6.11)	0.63		
		VR (5.29)	0.58		
Specialized Unit (*Aula Enclave*)		**RF** (8.89)	0.79	–	–
		GU (2.56)	0.31		

GU, Guidance; TO, Teacher's Organization; RF, Reinforcement; GU, Guía; NR, Non-verbal Revision; VR, Verbal Review; TE, Teacher's Explanation; SP, Student Participation; AT, Answer the Teacher.

TABLE 12 | Negative Behaviors of Students.

Teacher	Criteria Conduct	Consequent Conducts			
		Lag 1		Lag 2	
		Codes	Yule's Q	Codes	Yule's Q
Therapeutic Pedagogy Primary	CD	**CL** (19.31)	1.00	–	–

CD, Classroom Disruption; CL, Control.

TABLE 13 | Interaction Function.

Teacher	Criteria Conduct	Consequent Conducts			
		Lag 1		Lag 2	
		Codes	Yule's Q	Codes	Yule's Q
Therapeutic Pedagogy	TGI	SGI		TGI	
Preschool		(30.46)	0.99	(22.62)	0.97
Therapeutic Pedagogy		SGI		TGI	
Primary		(16.65)	0.98	(11.88)	0.94

SGI, General Interactions by the Students; TGI, General Interactions by the Teacher.

Aspects collected in the instrument of observation through the code MO.

With regard to the negative interventions of students, CD, only produce a significant pattern in TP with students of Primary School. This behavior is followed adequately by CL. This is logical because the teacher of TP teaches three students of Primary School, while Preschool Education only teaches a student. With a greater number of students the teacher requires a greater control and regulation of the group.

The code GI produces a significant reciprocal pattern in the two cases of TP (with students of Preschool Education and with students of Primary School) but not in Specialized Unit. These interactions, as discussed above, are important insofar as these interactions are intended to create a safe and trusting environment and to be a contrast encouraging allowing the students to rest and return to maintain the attention once the task is resumed. In Specialized Unit the contents are very practical. In this case these topics become part of the curriculum to be taught.

TABLE 14 | Neutral Interventions of Students.

Teacher	Criteria Conduct	Consequent Conducts			
		Lag 1		Lag 2	
		Codes	Yule's Q	Codes	Yule's Q
Therapeutic Pedagogy	SGI	TGI		SGI	
Preschool		(29.75)	0.99	(24.09)	0.98
Therapeutic Pedagogy		TGI		SGI	
Primary		(18.06)	0.99	(13.24)	0.95

TGI, General Interactions by the Teacher; SGI, General Interactions by the Students.

To the extent that the patterns developed by the teachers studied are adapted to the different needs presented by the students are indicative of good teaching practices. The patterns obtained indicate that the teachers observed offer support, help, and clear guidelines to perform the tasks through GU after interventions of the students (SP and AT) or after non-verbal supervision of their work (NR) (SP–GU, AT–GU, and NR–GU). This is in line with the results of the research that points out, as a relevant aspect for effective teaching, the importance of the teacher solving the possible doubts that arise in students (Berliner, 1983; Anderson, 1989, 2004) through support ("Scaffolds") to students so that they can carry out the activities (Palincsar and Brown, 1984; Van de Grift, 2007) and among other things, presenting new material in small steps, providing clear and detailed instructions, asking questions and providing guided practice (Rosenshine, 1979; Muijs and Reynolds, 2003; Houtveen et al., 2004; Houtveen and Van de Grift, 2006).

Another aspect to mention and which is indicator of good practices is they motivate their students to learn by giving them positive feedback (RF) after student interventions or non-verbal supervision of their work (NR–RF, SP–RF, and AT–RF). Several researches highlight the motivation of students as an important aspect for effective teaching. Thus, research refers, among other aspects, to strategies such as individualized attention, reinforcement, public and private praises, etc. (Murillo et al., 2011).

The teachers observed check their students assimilated the contents using questions or non-verbal supervision (VR and NR) after a theoretical explanation (TE) or a response of the students (AT) (TE–VR and AT–VR). The evaluation and the continuous monitoring of students have also been referred by researchers as an aspect which favors effective teaching. These aspects not only allow evaluating the achievements reached by the students, but also provide a motivation strategy for them by offering feedback on their learning process (Muijs and Reynolds, 2001; Anderson, 2004; Hattie and Timperley, 2007; Murillo et al., 2011; Pegalajar-Palomino, 2011). Specifically, literature refers to frequent supervision (Daloz, 1986; Stronge et al., 2004; Killen, 2005; Brookhart, 2009; Orlich et al., 2010; Murillo et al., 2011), asking questions or observing the work that students are doing (Pellicer and Anderson, 1995).

Another important issue on this topic is related with the discipline on the classroom. Thus, an indicative pattern of good practices in the teachers studied is that after CD the teacher applies CL. Applying an adequate control method to an inappropriate behavior allows the students to obtain the maximum benefit from teaching (Omoteso and Semudara, 2011).

Finally, these teachers try to generate a safe climate through GI and, at the same time, with this type of interactions they facilitate the students moments of disconnection that allows them to return to the task with a higher level of attention. Students should feel safe and comfortable to participate in the activities (Muijs and Reynolds, 2001). Literature refers to the climate that is generated in class as one important aspect which has to be taken into account for effective teaching. Thus, simple questions such as greeting or asking students for general aspects favor an adequate climate for learning (Hernández-Jorge, 2005).

These results show that the observational instrument used in this research, PROFUNDO-EPE, v.3, allows to capture the relevant aspects of Special Education professionals through a dynamic evaluation of teacher performance in the classroom. Thus, it is useful to evaluate the initial and permanent education of these teachers, allowing the detection of possible difficulties and establishing recommendations for improvement.

The limitation of this work is that it has been carried out only with two teachers. To check whether the conduct of special education teachers is different depending on the context and if it is appropriate, it is necessary to extend this study to a larger number of teachers in different contexts of Special Education.

Additionally, it is important to point out that the observational instrument used on this study shows, through the Teaching Functions, *what* the teachers does in their educational work in the classroom. However, it would be extremely interesting to be able to capture *how* they do it, that is, the specific strategies that the teacher use to explain, guide, reinforce or evaluate. These aspects have begun to be analyzed with university teacher through the Protocol of Observation of the Function of Explanation (PROFE in Spanish) (Borges et al., 2016b). This protocol has been designed to operationalize how the teacher transmits knowledge, collecting all those strategies, resources or observable styles that the teacher uses during the theoretical exposition of the contents.

This aspect has great relevance in order to capture behavioral patterns which represent good teaching practice. The research

about teachers and teaching effectiveness refers to specific strategies used by teachers within each of the teaching functions.

Undoubtedly, a particularly important aspect for achieving quality education is that it is capable of giving an appropriate answer to the different educational needs presented by students. Thus, education could maximize their abilities and ensure adequate personal and social development. Therefore, the efforts directed to assess that this is being achieved are a requirement on this area. This requires an evaluation on the development of the educational process that is formative evaluation (Tejada, 1999; López de la Llave and Pérez-Llantada, 2004). It is very important to provide quality education.

ETHICS STATEMENT

The results of this study are part of the research from a doctoral thesis ("*Evaluation of the process of teaching behavior in Education Primary and Special*") that was done and presented in the University of La Laguna. This doctoral thesis has the ethical approval of the Ethics Committee of Research and Animal Welfare (CEIBA, in Spanish) of the University of La Laguna, with registration number CEIBA2017-0224. In accordance with the Organic Law 15/1999 of December of Protection of Personal Data (1999, BOE n° 298 of December 14) the written informed consent was obtained from all individual participants included in the study (teachers and parents), where they agreed to participate in the investigation, as well as recording their behavior.

In addition, and following the guidelines of the aforementioned law, we request the written observer's confidentiality agreement.

AUTHOR CONTRIBUTIONS

MR-D has participated in the entire study. AB has also participated in the totality of the work (theoretical review, study planning, method and discussion), except in the analysis. Both authors have participated in the writing of the article.

ACKNOWLEDGMENT

We thank both the teachers who have agreed to be recorded in the development of their classes, as well as the parents who have authorized the recording. Also, we thank the observers who have codified the sessions used for this study.

REFERENCES

Ainscow, M. (2007). *Desarrollo de Escuelas Inclusivas. Ideas, Propuestas y Experiencias Para Mejorar las Instituciones Escolares*. Madrid: Narcea.

Ainscow, M., Hopkins, D., Southworth, G., and West, M. (2001). *Hacia Escuelas Eficaces Para Todos. Manual Para la Formación de Equipos Docentes*. Madrid: Morata.

Alegre, O. M. (2000). *Diversidad Humana y Educación*. Málaga: Aljibe.

Anderson, L. M. (1989). *The Effective Teacher: Study Guide and Readings*. New York, NY: Random House.

Anderson, L. M. (2004). *Increasing Teaching Effectiveness*. Paris: UNESCO-IIPE.

Anguera, M. T., Blanco, A., and Losada, J. L. (2001). Diseños observacionales, cuestión clave en el proceso de la metodología observacional. *Metodol. Cienc. Comport.* 3, 135–160.

Bain, K. (2006). *Lo Que Hacen Los Mejores Profesores Universitarios*. Valencia: Publicacions de la Universitat de Válencia.

Bakeman, R., and Gottman, J. M. (1986). *Observing Interaction: An Introduction to Sequencial Analysis*. Cambridge: Press Syndicate of the University of Cambridge.

Bakeman, R., and Quera, V. (1996). *Análisis de la Interacción. Análisis Secuencial con SDIS y GSEQ*. Madrid: Ra-Ma.

Berliner, D. C. (1983). Developing conceptions of classroom environments: some light on the T in classroom studies of ATI. *Educ. Psychol.* 18, 1–13. doi: 10.1080/00461528309529256

Blanco, A., and Anguera, M. T. (2000). "Evaluación de la calidad en el registro del comportamiento: aplicación a deportes de equipo," in *Métodos numéricos en Ciencias Sociales*, eds E. Oñate, F. García-Sicilia, and L. Ramallo (Barcelona: Centro Internacional de Métodos Numéricos en Ingeniería), 30–48.

Blanco, A., Castellano, J., and Hernández Mendo, A. (2000). Generalizabilidad de las observaciones de la acción del juego en el fútbol. *Psicothema* 12, 81–86.

Blanco, A., Sastre, S., and Escolano, E. (2010). Desarrollo ejecutivo temprano y Teoría de la Generalizabilidad: bebés típicos y prematuros. *Psicothema* 22, 221–226.

Blanco-Villaseñor, A. (1991). La Teoría de la Generalizabilidad aplicada a diseños observacionales. *Rev. Mex. Anál. Conducta* 14, 23–63. doi: 10.5514/rmac.v17.i3.23338

Borges, Á., Díaz, M., Rodríguez-Naveiras, E., González de la Fé, T., Domínguez, M., Isla, R., et al. (2016a). "Evaluación de buenas prácticas docentes mediante metodología observacional," in *Innovación Docente Para Convencidos. VI Jornadas de Innovación Educativa de la Universidad de La Laguna*, eds A. Vega and J. O. 'Dwyer (San Cristóbal de La Laguna: Universidad de La Laguna), 35–46.

Borges, Á., Falcón, C., and Díaz, M. (2016b). Creation of an observational instrument to operationalize the transmission of contents by university teachers. *Int. J. Soc. Sci. Stud.* 4, 82–89. doi: 10.11114/ijsss.v4i7.1596

Borges, A., and Rodríguez-Dorta, M. (2015). "Assessment of the teaching quality of teachers of primary and special education," in *Progress in Education*, Vol. 35, ed. R. V. Nata (New York, NY: Nova Science Publishers), 83–104.

Borges, Á., Rodríguez-Naveiras, E., Díaz, M., Domínguez, M., de la Fé, T. G., Dorta, R. L., et al. (2016c). The observational instrument of best practices: an observational instrument to measure best practice behaviors in university teachers. *Int. J. Soc. Sci. Stud.* 4, 71–78. doi: 10.11114/ijsss.v4i10.1868

Brookhart, S. M. (2009). "Assessment and examinations," in *International Handbook of Research on Teachers and Teaching*, eds L. J. Saha and A. G. Dworkin (New York, NY: Springer), 723–739. doi: 10.1007/978-0-387-73317-3_45

Brophy, J. E., and Good, T. (1986). *Learning from Teaching: A Developmental Perspective*. Boston, MA: Allyn & Bacon.

Brown, A. (2009). *Teaching Strategies: A Guide to Effective Instruction*. Boston MA: Houghton Mifflin Company.

Cadenas, M., Rodríguez, M., and Díaz, M. (2012). Los equipos de entrenamiento: una muestra para el estudio de los sesgos en fiabilidad entre parejas de observadores. *Rev. Investig. Divulg. Psicol. Logoped.* 2, 41–46.

Cid-Sabucedo, A., Pérez-Abellás, A., and Zabalza, M. (2009). Las prácticas de enseñanza declaradas de los "mejores profesores" de la universidad de Vigo. *RELIEVE* 15, 1–29.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psycholo. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement of partial credit. *Psychol. Bull.* 70, 213–220. doi: 10.1037/h0026256

Conklin, H. (2012). Tracing learning from divergent teacher education pathways into practice in middle grades classrooms. *J. Teacher Educ.* 63, 171–184. doi: 10.1177/0022487111426294

Cotton, K. (1995). *Effective Classroom*. London: Cassell.

Cronbach, L. J., Gleser, G. C., Nanda, H., and Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. Nueva York, NY: John Wiley and Sons.

Daloz, L. A. (1986). *Effective Teaching and Mentoring*. San Francisco, CA: Jossey-Bass.

Dalton, S. S. (2007). *Five Standards for Effective Teaching: How Succeed with All Student*. San Francisco, CA: Jossey-Bass.

Decree 157/1986 of October 24 on the Management of Therapeutic Pedagogy in an Integrated System (1986). *Decreto 157/1986, de 24 de octubre, de Gestión de la Pedagogía Terapéutica en un sistema integrado, BOC, n° 136, 12 de noviembre de 1986*.

Del Valle Ballón, J. (2012). La mirada, la disposición y la luz. El enfoque de buenas prácticas docentes y los estudios generales. *Blanco Negro* 3, 1–7.

Díaz, M. (2014). *Protocolo de Observación de Funciones Docentes en Universidad: Un Instrumento Para la Evaluación de la Conducta del Profesorado Universitario*. Doctoral thesis, Servicio de publicaciones Universidad de La Laguna, La Laguna.

Díaz, M., Borges, A., Valadez, M. D., and Zambrano, R. (2015). Valoración de buenas prácticas docentes a través de observación sistemática. *Univ. Psychol.* 14, 913–922. doi: 10.11144/Javeriana.upsy14-3.vbpd

Fleiss, J. L. (1981). *Statistical Method for Rates and Proportions*. Nueva York, NY: Wiley.

García, J. M., and Congosto, E. (2000). "Evaluación y calidad del profesorado," in *Evaluación y Gestión de la Calidad Educativa. Un Enfoque Metodológico*, ed. T. González (Málaga: Ediciones Aljibe, S.L.), 127–156.

Gómez Montes, J. M. (2005). Pautas y estrategias para entender y atender la diversidad en el aula. *Pulso* 28, 199–214.

González-Peiteado, M. (2013). Los estilos de enseñanza y aprendizaje como soporte de la actividad docente. *J. Learn. Styles* 11, 51–70.

Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487

Hernández, P. (1991). *Psicología de la Educación: Corrientes Actuales y Teorías Aplicadas*. Méjico: Trillas.

Hernández, P. (2006). *Psicología de la Educación. Educación del Pensamiento y Las Emociones*, 2ª Edición. Madrid: Narcea S.A. de Ediciones.

Hernández-Jorge, C. (2005). *Habilidades de Comunicación Para Profesionales*. Tenerife: ARTE Comunicación visual.

Hiebert, J., Wearne, D., and Taber, S. (1991). Fourth grades' gradual construction of decimal fractions during instruction using different physical representations. *Elem. Sch. J.* 91, 321–341. doi: 10.1086/461658

Hintze, J. M., and Matthews, W. J. (2004). The Generalizability of systematic direct observations across time and setting: a preliminary investigation of the psycometrics of behavioral onservation. *Sch. Psychol. Rev.* 33, 258–270.

Houtveen, A. A. M., Booij, N., de Jong, R., and Van de Grift, W. J. C. M. (1999). Adaptive instruction and pupil achievement. *Sch. Effect. Sch. Improv.* 10, 172–192. doi: 10.1076/sesi.10.2.172.3508

Houtveen, A. A. M., and Van de Grift, W. J. C. M. (2006). *Reading Instruction for Struggling Learners*. Utrecht: ISOR.

Houtveen, A. A. M., Van de Grift, W. J. C. M., and Creemers, B. P. M. (2004). Effective school improvement in mathematics. *Sch. Effect. Sch. Improv.* 15, 337–376. doi: 10.1080/09243450512331383242

Hunt, G. H., Wiseman, D. G., and Touzel, T. J. (2009). *Effective Teaching: Preparation and Implementation*. Springfield: Charles C Thomas.

Jiménez, B. (1999). "Evaluación de la docencia," in *Evaluación de Programas, Centros y Profesores*, ed. B. Jiménez (Madrid: Síntesis), 173–206.

Kallison, J. M. (1986). Effects of lesson organization on achievement. *Am. Educ. Res. J.* 23, 337–347. doi: 10.3102/00028312023002337

Killen, R. (2005). *Programming and Assessment for Quality Teaching and Learning*. Sidney: Thomsom. Killen.

Lloyd, B. P., Kennedy, C. H., and Yoder, P. J. (2013). Quantifying contingent relations from direct observation data: transitional probability comparisons versus Yule's Q. *J. Appl. Behav. Anal.* 46, 479–497. doi: 10.1002/jaba.45

López de la Llave, A., and Pérez-Llantada, M. C. (2004). *Evaluación de Programas en Psicología Aplicada (Salud, Intervenciones Sociales, Deporte, Calidad)*. Madrid: Dykinson.

Maddox, H., and Hoole, E. (1975). Performance decrement in the lectura. *Educ. Rev.* 28, 17–30. doi: 10.1080/0013191750280102

Martínez Geijo, P. (2007). *Los Estilos de Aprendizaje y Enseñanza Desde la Práctica en el Aula*. Bilbao: Mensajero.

Mayer, R. E., and Gallini, J. K. (1990). When is an illustration worth ten thousand words? *J. Educ. Psychol.* 82, 715–726. doi: 10.1037/0022-0663.82.4.715

Mayorga, M. J., and López, E. (2005). Observar para evaluar al profesorado: una experiencia práctica. *XXI Rev. Educ.* 7, 167–181.

Melton, R. F. (1978). Resolution of conflicting claims concerning the effect of behavioral objectives on student learning. *Rev. Educ. Res.* 48, 291–302.

Mortimore, P., Sammons, P. L., Stoll, L., Lewis, D., and Ecob, R. (1988). *School Matters: The Junior Years*. Wells: open Books.

Muijs, D., and Reynolds, D. (2001). *Effective Teaching. Evidence and Practice*. London: Sage.

Muijs, D., and Reynolds, D. (2003). Student background and teacher effects on achievement and attainment in mathematics: a longitudinal study. *Educ. Res. Eval.* 9, 289–314. doi: 10.1076/edre.9.3.289.15571

Murillo, F. J., Martínez, C. A., and Hernández, R. (2011). Decálogo para una enseñanza eficaz. *Rev. Iberoam. Sobre Calid. Eficacia Cambio Educ.* 9, 6–27.

Omoteso, B., and Semudara, A. (2011). The relationship between teachers' effectiveness and management of classroom misbehaviours in secondary schools. *Sci. Res. Pychol.* 2, 902–908. doi: 10.4236/psych.2011.29136

Organic Law 15/1999 of December of Protection of Personal Data (1999). *Ley Orgánica 15/1999 de 13 de diciembre, de Protección de Datos de Carácter Personal, BOE, n° 298, 14 de diciembre 1999*.

Orlich, D. C., Harder, R. J., Callahan, R. C., Trevisan, M. S., and Brown, A. H. (2010). *Teaching Strategies: A Guide to Effective Instruction*. Boston, MA: Wadsworth.

Palincsar, A. S., and Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cogn. Instr.* 1, 117–175. doi: 10.1207/s1532690xci0102_1

Patterson, G. R. (1982). *Coercive Family Process*. Eugene, OR: Catalia.

Pegalajar-Palomino, M. C. (2011). Estrategias didácticas en el proceso de enseñanza-aprendizaje de alumnos con discapacidad escolarizados en centros específicos de educación especial. *Rev. Educ. Inclusiva* 4, 103–116.

Pegalajar-Palomino, M. C. (2014). Importancia de la actividad formativa del docente en centros de educación especial. *Rev. Electrón. Int. Form. Profesorado* 17, 177–192. doi: 10.6018/reifop.17.1.181731

Pellicer, L. O., and Anderson, L. W. (1995). *An Handbook for Teacher Leaders*. Thousand Oaks. CA: Corwin Press.

Pianta, R. C., and Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: standardized observation can leverage capacity. *Educ. Res.* 38, 109–119. doi: 10.3102/0013189X09332374

Renkl, A., and Helmke, A. (1992). Discriminant effects of performance-oriented and structure-oriented mathematics tasks on achievement growth. *Contemp. Educ. Psycol.* 17, 47–55. doi: 10.1016/0361-476X(92)90045-Z

Rodríguez, G., and Ibarra, M. S. (2013). "La evaluación de los procesos y contextos educativos," in *Procesos y Contextos Educativos: Nuevas Perspectivas Para la Práctica Docente*, ed. F. Navaridas (Logroño: Genueve Ediciones), 257–294.

Rodríguez-Dorta, M. (2015). *Evaluación de Proceso del Comportamiento Docente en Educación Primaria y Especial*. Doctoral thesis, Universidad de La Laguna. Servicio de Publicaciones Universidad de La Laguna, La Laguna.

Rodríguez-Dorta, M., and Borges, A. (2015a). "Optimización del tiempo y número de sesiones para la observación de las funciones docentes en educación primaria y especial," in *XIII Congreso de Metodología de las Ciencias Sociales y de la Salud*, eds A. Borges and S. Hess (San Cristóbal de La Laguna: Universidad de La Laguna), 85–96.

Rodríguez-Dorta, M., and Borges, A. (2015b). Optimización y eficiencia en el análisis de datos en observacional. *Rev. Electrón. Metodol. Apl.* 21, 1–15.

Rodríguez-Naveiras, E. (2011). *PROFUNDO: Un Instrumento Para la Evaluación de Proceso de un Programa de Altas Capacidades*. Doctoral thesis, Servicio de Publicaciones Universidad de La Laguna, La Laguna.

Rosenshine, B. (1979). "Content, time and direct instruction", in *Research on Teaching*, eds P. L. Peterson and H. J. Walberg (Berkeley, CA: McCutchan), 28–56.

Ruiz Rodríguez, E. (2003). Adaptaciones curriculares individuales para alumnos con Síndrome de Down. *Revista Síndrome de Down* 76, 2–11.

Smith, L. R., and Cotton, M. L. (1980). Effect of lesson vagueness and discontinuity on student achievement and attitudes. *J. Educ. Psychol.* 72, 670–675. doi: 10.1037/0022-0663.72.5.670

Stronge, J. H., Tucker, P. D., and Hindman, J. L. (2004). *Handbook for Qualities of Effective Teachers*. Alexandria VA: Association for Supervision and Crriculum Development.

Sykes, G., Bird, T., and Kennedy, M. (2010). Teacher education: its problems and some prospects. *J. Teach. Educ.* 61, 464–476. doi: 10.1177/0022487110375804

Tejada, J. (1999). "La evaluación: su conceptualización," in *Evaluación de Programas, Centros y Profesores*, ed. B. Jiménez (Madrid: Síntesis, S.A.), 25–56.

Van de Grift, W. (2007). Quality of teaching in four European countries: a review of the literature and application of an assessment instrument. *Educ. Res.* 49, 127–152. doi: 10.1080/00131880701369651

Zavala, A., and de la Torre, G. (2015). "Fundamentos de la educación especial: modelos de intervención pedagógico-didácticos," in *Trastornos del Desarrollo y Problemas de Aprendizaje*, Vol. I, eds M. Hume and G. López (México: Ediciones Fontamara, S. A), 11–50.

Systematic Observation: Relevance of This Approach in Preschool Executive Function Assessment and Association with Later Academic Skills

Elena Escolano-Pérez[1]*, Maria Luisa Herrero-Nivela[1], Angel Blanco-Villaseñor[2] and M. Teresa Anguera[2]

[1] Faculty of Education, University of Zaragoza, Zaragoza, Spain, [2] Faculty of Psychology, University of Barcelona, Barcelona, Spain

Executive functions (EFs) are high-level cognitive processes that allow us to coordinate our actions, thoughts, and emotions, enabling us to perform complex tasks. An increasing number of studies have highlighted the role of EFs in building a solid foundation for subsequent development and learning and shown that EFs are associated with good adjustment and academic skills. The main objective of this study was to analyze whether EF levels in 44 Spanish children in the last year of preschool were associated with levels of literacy and math skills the following year, that is, in the first year of compulsory education. We used a multi-method design, which consisted of systematic observation to observe preschool children during play and selective methodology to assess their reading, writing, and math skills in the first year of compulsory primary education. General linear modeling was used to estimate the percentage of variability in academic skills in the first year of primary school that was explained by preschool EF abilities. The results showed that preschool EF level, together with participants and the instrument used to assess academic skills, explained 99% of the variance of subsequent academic performance. Another objective was to determine whether our findings were generalizable to the reference population. To make this determination, we estimated the optimal sample size for assessing preschool EFs. To do this, we performed a generalizability analysis. The resulting generalizability coefficient showed that our sample of 44 students was sufficient for assessing preschool EFs. Therefore, our results are generalizable to the reference population. Our results are consistent with previous reports that preschool EF abilities may be associated with subsequent literacy and math skills. Early assessment of EFs may therefore contribute to identifying children who are likely to experience later learning difficulties and guide the design of suitable interventions for the optimization of EFs.

Keywords: systematic observation, child development, executive functions, academic competences, preschoolers, generalizability

INTRODUCTION

Although not generally compulsory, preschool is essential for early childhood development. This stage of education can determine children's later development and learning and, consequently, performance and success at school and work, as well as in their personal and social lives (Duncan and Magnuson, 2013; Bartik, 2014). In these first years of life, the main neural connections that provide the basis for learning and behavior are established through the constant interaction of neurobiological factors and the stimulation of the child's surroundings (Bick and Nelson, 2017). During preschool, it is possible to take early action to avoid or compensate for situations arising from personal, family, and/or social inequalities that can subsequently have an impact on development and learning throughout childhood and into adulthood (Kaufman et al., 2015).

After finishing preschool, children begin compulsory primary education. Primary school presents children with a context that is very different from preschool: teacher–student interaction is less emotional; greater autonomy is expected of students; the curriculum is more oriented toward reading, writing, and mathematics; work periods are longer and require more sustained attention and concentration, etc. Because of these new characteristics and expectations, for many children the transition to primary school is a stressful period characterized by excessive demands and various difficulties (Veličković, 2013; Harper, 2016). In fact, some children who adapt well to preschool experience a decrease in skill level when they start primary school: they become less active, more easily distracted, less eager to learn and participate in class activities, more dependent, more insecure, and have more problems in their peer relationships (Veličković, 2013). This academic and socioemotional maladjustment can contribute to the likelihood that children will become inactive students in primary school and can even harm their overall well-being by causing additional health and emotional problems. By contrast, children who adjust well to this transition are generally successful in primary school and also later in life (Veličković, 2013; Harper, 2016).

Recent studies in this area (Blair and Raver, 2015; Moriguchi et al., 2016) have found that preschool executive functions (EFs) are essential to building a solid foundation for subsequent development and learning and are associated with school adjustment and academic success at the start of primary education. Consequently, research on preschool EFs has increased considerably over the past decade. However, many aspects of preschool EFs—including how best to evaluate them—remain poorly understood. Preschool EFs are an area of study in which conceptual aspects are better understood than aspects related to development and measurement (Willoughby and Blair, 2016). In order to help overcome these limitations, this study provides an example of how systematic observation, applied in children's natural context, can be an appropriate tool for assessing preschool EFs. On the basis of this assessment, we analyze the extent to which preschool EFs may be associated with academic skills 1 year later, in the first year of primary education. We also analyze whether the results obtained with the convenience sample recruited can be generalized to the reference population. Our use of generalizability theory (G theory) for this purpose represents a novel contribution in the measurement of preschool EFs in observational studies.

Preschool Executive Functions

EFs are a family of high-level cognitive processes that allow for conscious, goal-directed control of thoughts and actions, making it possible to solve problems effectively and efficiently, particularly in novel situations (Diamond, 2013; Carlson et al., 2016; Zelazo et al., 2016). In the preschool years, EFs consist of three main processes: working memory, inhibition, and cognitive flexibility (or shifting or switching) (Miyake et al., 2000; Diamond, 2013; Howard et al., 2015; Carlson et al., 2016; Moriguchi et al., 2016).

Working memory is the ability to hold information active in one's mind and mentally work with it for brief periods of time as a platform for guiding one's behavior. Two types of working memory are distinguished by the content: verbal or semantic working memory, on the one hand, and non-verbal or visuospatial working memory, on the other (Miyake et al., 2000; Diamond, 2013).

Inhibition refers to the ability to control one's behavior, thoughts, and/or attention in order to override a strong internal predisposition or external lure. It includes (a) behavioral inhibition (or inhibition of action) to control or cancel one's motor behavior, resist temptations, and not act impulsively; (b) cognitive inhibition to control and/or tune out thoughts and memories; and (c) resistance to distractor interference (or inhibition of attention) to select the information or stimulus one needs to complete a task while ignoring competing distractions (Friedman and Miyake, 2004).

Cognitive flexibility is the ability to quickly adapt one's course of thought or action to the changing demands of a situation. This involves being able to shift one's attention from one condition of a task—e.g., stimulus, dimension, or rule—to another (Miyake et al., 2000; Diamond, 2013).

These three EFs undergo considerable development during the preschool years (Anderson and Reidy, 2012; Howard et al., 2015; Nieto et al., 2016), coinciding with important changes in neuroanatomy and brain structures, especially in neural circuits of the prefrontal region that are particularly susceptible to experiential input during this period of rapid growth and plasticity (Bick and Nelson, 2017).

Preschool EFs and Later Academic Performance

As mentioned above, EFs are essential to the ability to perform academic tasks. Evidence for this claim has been obtained in samples of students of various ages, with and without learning difficulties, and with adequate and inadequate academic performance, independently of variables such as cultural context and socioeconomic level. EFs are so important in academic performance that they are even better predictors of academic performance than IQ (Viterbori et al., 2015; Purpura et al., 2017).

Several studies have shown that preschool EFs have an influence on students' later skills in literacy and mathematics, the curricular areas in which the effect of EFs has been most studied.

Preschool EFs and Literacy Skills

In the area of literacy skills, verbal working memory is related to phonological awareness, which is necessary for the output of words and phrases—both spoken and written—and therefore in reading and writing. In order to produce a word or sentence, children must be able to hold multiple sounds or words in their memory and combine them (Purpura et al., 2017). Studies have found that dyslexic children, who often have phonological problems, perform more poorly on working memory tasks than typically developing children (Varvara et al., 2014). Reading comprehension is another literacy skill in which working memory plays a major role. When we read, we must relate the ideas that appear in each sentence and paragraph with those we have just read in the previous sentences and paragraphs. These ideas must be stored and activated in our mind and combined in a new structure, forming a whole that gives meaning to the text. Essentially, as we read, working memory plays a key role in storing the intermediate and final products of our computations, allowing us to build and integrate the successive ideas we extract from the text (Cartwright, 2015; García-Madruga et al., 2016). Similarly, working memory is also required in the composition of written texts and phrases. But if, while reading a text, we encounter information that is irrelevant or not of interest to us—for example, when scanning a text for information on a particular subject—we must be able to inhibit, reject, and not be distracted by any information that does not meet our needs, keeping our attention on the information that is relevant to our goal (Cartwright, 2015; García-Madruga et al., 2016; Purpura et al., 2017). Thus, inhibition also plays a role in reading comprehension—and in verbal and written expression—in close interaction with working memory and, as explained below, cognitive flexibility. It is therefore clear that learning and academic tasks require the simultaneous participation of the various EF components (Colé et al., 2014; Cartwright, 2015; García-Madruga et al., 2016; Rapoport et al., 2016).

Various researchers have also shown the significant contribution of cognitive flexibility to literacy skills such as phonological and print awareness, word reading, and reading comprehension. Cognitive flexibility is needed to create cross-modal connections between spoken and written language and to access and integrate the different characteristics of print (phonology, morphology, syntax, semantics) during the process of word recognition. Cognitive flexibility also has a critical role in reading comprehension, as we need to process phonological codes in order to recognize the written words while also processing the meaning of the words (Colé et al., 2014; Cartwright, 2015). Cognitive flexibility is thus a key process for understanding specific reading comprehension difficulties (Engel de Abreu et al., 2014). Students with these difficulties are relatively good at decoding, so they sound like good readers, but they have problems with comprehension. They focus inflexibly on decoding processes (i.e., on word-level features of print) and pay only limited attention to meaning. They have difficulty

shifting their focus to the text's meaning or to simultaneously managing decoding and the construction of meaning (Colé et al., 2014; Cartwright, 2015).

Preschool EFs and Math Skills

Regarding the contribution of EFs to children's math skills, a substantial body of evidence shows that working memory is critical for mathematical proficiency. For example, calculation relies on working memory processes because it involves storing temporary information—the numbers involved in the operation, partial results, and the amount to be carried—and performing mental operations on this information until the final result is obtained. Working memory is especially important when the problem is presented verbally rather than visually (Clark et al., 2013; Rapoport et al., 2016). Nevertheless, several authors, in comparing primary school children with high and low working memory, found a significant difference in calculation ability even if the arithmetic operations were presented in written format (Viterbori et al., 2015). Number comparison is another math skill that requires holding multiple pieces of information (the numerals) in mind and combining or manipulating them in order to compare their magnitudes and identify the smallest or largest. Working memory also plays a role in the acquisition of new arithmetic facts—for example, addition and multiplication tables—because the operation and answer need to be held in mind together in order to strengthen the relationship between them (Cragg and Gilmore, 2014; Purpura et al., 2017). The influence of working memory in complex components of mathematics, such as problem-solving, can be illustrated in various ways. For example, when solving a problem, we must select the relevant information and hold it in mind. Some evidence suggests that poor problem solvers remember less relevant information than good problem solvers. The role of working memory in solving mathematical problems is closely related to a student's ability to access the right information (e.g., appropriate algorithms) from long-term memory (Viterbori et al., 2015; Purpura et al., 2017).

As for inhibition, it is important at younger ages to suppress less sophisticated strategies (e.g., counting on from the first addend) in order to use more sophisticated strategies (e.g., counting on from the larger addend). Inhibition is also necessary in order to suppress answers to related but incorrect number facts (for example, in response to 4×4, children must inhibit 8, the solution to $4 + 4$). Cross-operation errors such as these are generated by difficulty in inhibiting the incorrect responses in a set of possible and competing responses activated by the memory (Cragg and Gilmore, 2014; Viterbori et al., 2015). When a child is learning new concepts, inhibition—along with cognitive flexibility or shifting—is important in suppressing an automatic procedural approach and shifting attention toward the numerical relationships involved (Cragg and Gilmore, 2014). Inhibition also contributes to solving math problems, especially when the text of the problem contains irrelevant and distracting data that the child must suppress in order to develop an appropriate mental problem-solving model (Viterbori et al., 2015). Some studies show that students with mathematical difficulties have trouble inhibiting irrelevant information and focusing on

relevant information (Cragg and Gilmore, 2014; Viterbori et al., 2015). However, when solving complex mathematical problems, children must also have cognitive flexibility in order to switch between different procedures (e.g., adding and subtracting) or to look for an alternative problem-solving procedure after attempting to solve a problem using an unsuitable procedure. Cognitive flexibility also appears to be related to more abstract aspects of mathematics, such as cardinal number knowledge (Purpura et al., 2017). When children progress from applying the counting sequence to sets (one-to-one correspondence) to achieving quantity (the cardinal number that represents a sum or total number of existing elements), they shift from thinking about counting as a procedure to thinking about it as a conceptual process. Specifically, children must adapt their thinking and flexibly move from the procedural task of counting to understanding counting as providing quantitative information (Viterbori et al., 2015; Purpura et al., 2017).

It is therefore clear that EFs make significant contributions to young learners' overall mathematics and literacy performance.

Systematic Observation in Preschool

The recent literature on early childhood education and development increasingly argues that the assessment of development processes and learning during preschool should be done primarily through systematic observation in the natural learning context (Early Head Start National Resource Center, 2013; Jablon et al., 2013). The literature also stresses that play—an activity inseparable from a child's life—is an indispensable resource for the childhood teaching–learning process and for the systematic observation of children's progress and development (Nell and Drew, 2013; Fasulo et al., 2017). Systematic observation of a child's behavior during play makes it possible to obtain relevant data to describe, explain, and understand fundamental aspects of the child's development and learning (Federici et al., 2017; Otsuka and Jay, 2017), including the development of EFs. Accordingly, the literature on EFs indicates that given children's impulsive behaviors and linguistic, motor, and attentional limitations, the study of EF development in early childhood, like the tasks and tools used for their assessment, must be based on the children's everyday activities (Nieto et al., 2016), such as play. However, few studies have used systematic observation of children's play as a tool for obtaining objective and valid information about preschool EFs.

This lack of research may be due to certain difficulties associated with systematic observation, such as the high cost in terms of time (all observers must undergo rigorous prior training) and the painstaking process of collecting and recording the data (Portell et al., 2015). The time cost is even higher when the subjects observed are children, because of the additional complexity and difficulties inherent in working with young participants as a result of their developmental characteristics (behavioral instability, short attention span, and high fluctuation of motivation), the need to create a climate of trust to ensure the children's well-being and participation, and legal and ethical requirements that must be met in order to comply with international research guidelines (Shaw et al., 2011). Because of the need to obtain informed consent from parents or guardians

for children to participate in research, many studies involving children have small samples that are not very representative of the reference population. This could be a source of error and the results of such studies may not be generalizable to the reference population. However, new data analysis structures (such as G theory) are making it possible to overcome these limitations.

Generalizability Theory to Generalize Results from Systematic Observation of Preschool Behavior

In the field of education and development—and in the behavioral sciences generally—observed phenomena are often influenced by many factors, so the repetition of a particular experience or the use of a different instrument can modify the initial result considerably, leading one to wonder whether the observed values are interpretable or if they are the result of random fluctuations introduced by the act of measurement. This question is particularly important in behavioral observation designs. The use of G theory allows us to analyze the various sources of variance that can affect an observational measurement or measurement design and estimate the degree of generalization of a theoretical value with respect to specific conditions (Blanco-Villaseñor et al., 2014). However, G theory can be adapted to the specific conditions of each object of measurement, so its use in observational studies can contribute to the generalization of results and to improving their applicability on future occasions.

G theory assumes the existence of multiple sources of variance (variables or facets) in any measurement situation. This approach can estimate the accuracy of a measurement that is subject to multiple sources of error (Cardinet et al., 2010), allowing real variability to be separated from error variance. One of the important objectives of measurement is to try to identify and measure the components of variance that contribute to the error of an estimation and implement strategies that reduce the influence of these sources of error on the measurement.

As mentioned above, studies involving children often have a small sample size. On occasions, a "small" sample can be viewed as a possible limitation that could act as an additional source of measurement error. G theory allows us to analyze this source of variance and estimate the accuracy of the measurement in a studied sample. This makes it possible to estimate the degree to which the results obtained for a particular sample can be generalized to the reference population (Blanco-Villaseñor et al., 2014). Despite the advantages offered by this approach, few observational studies have used generalizability analysis, and even fewer have studied children. Fewer still have applied G theory to sample size estimation, given that G theory is normally used in observational studies to determine reliability and validity.

Aims of the Present Empirical Investigation

Given the background set out above, the objectives of this study were as follows:

(1) To determine whether different EF levels measured in children through systematic observation at the end of preschool are associated with different levels of literacy

and math skills the following year, that is, at the start of compulsory education.

(2) To determine whether the results obtained with the convenience sample recruited can be generalized to the reference population and, therefore, whether the studied sample is of sufficient size.

MATERIALS AND METHODS

Ethics Statement

The study was carried out in accordance with the recommendations of the ethics committee at Zaragoza University and the principles of the Declaration of Helsinki. Written informed consent was obtained from the parents of all the children who participated. Each child received a small reward (two chocolates) for participating.

Design

We used a multi-method design (Elliott, 2007; Sánchez-Algarra and Anguera, 2013; Kumschick et al., 2014; Mangelsdorf and Eid, 2015) consisting of systematic observation to observe preschool children during play and selective methodology to assess their reading, writing, and math skills the following year, that is, in the first year of compulsory primary education.

Systematic observation was non-participative and active and the behaviors observed were fully perceivable (Anguera, 2003; Shaughnessy et al., 2009; Bakeman and Quera, 2011).

The observational design was point, nomothetic, and multidimensional (Blanco-Villaseñor et al., 2003). It was point because a single session per participant was observed to assess each of the EFs analyzed; nomothetic because multiple observation units were analyzed; and multidimensional because several domains of EFs (working memory, inhibition, and mental flexibility) were analyzed within the theoretical model proposed by Miyake et al. (2000) and developed by other authors (e.g., Diamond, 2013).

Participants

Forty-four Spanish participants were recruited. They were all students, aged 5–6 years, in their last year of preschool (last year of non-compulsory education in Spain) at the same school when the study started. The school was located in a central moderate-to-high income neighborhood of a Spanish city with approximately 700,000 inhabitants. The vast majority of the students approached (95.65% of all the children in their last year of preschool education) participated in the study. The other

children (4.35%) did not participate as their parents did not provide their informed consent.

The students had to meet three inclusion criteria: (1) attendance at the targeted school since the second year of preschool education (age 3); (2) absence of the following disorders or risk factors: (a) birth weight <2000 g and/or gestational age <36 weeks or significant pre-, peri-, or postnatal events; (b) medical/neurological conditions affecting growth, development, or cognition (e.g., seizure) and sensory deficits (e.g., vision or hearing loss); (c) neurodevelopmental disorders (e.g., autism spectrum disorder, attention-deficit hyperactivity disorder, language disorder); (d) genetic conditions or syndromes; and (e) a first-degree relative with schizophrenia, bipolar disorder, or related disorders; and (3) an adequate IQ for their chronological age. The information to assess compliance with the first two criteria was provided by the children's parents, and IQ was tested using the Spanish Battery of Differential and General Abilities Tests (BADyG) (Yuste and Yuste, 2001).

The sample was a convenience sample formed by children who met the inclusion criteria and whose parents signed the informed consent form authorizing their participation. **Table 1** summarizes the main descriptive characteristics of the sample.

Games

In order to obtain videos of the preschool children during play, each participant was offered the chance to participate in five games. These games were based on other non-standardized games and tasks that had been used in various studies to assess preschool EFs (Anderson and Reidy, 2012). Through the observation of the children's spontaneous behavior in these games, it was possible to extract information about their EFs.

All of the games proposed to the children formed part of a fantasy story (the creation of a fantasy world is a characteristic of many children's games; Garris et al., 2002). This fantasy story—in which each participant acted as the protagonist—was set in space, a topic that the teachers had indicated was of interest to the participating children. Although instructions were given for each game as part of the fantasy story, at no time were the child's actions restricted or penalized in any way. Thus, the child was allowed to act freely throughout the course of the games.

Game 1: Preparing for the Journey

This game, based on the Backward Word Span task (Carlson, 2005; Diamond, 2013; Visu-Petra et al., 2014; Howard et al., 2015; Nieto et al., 2016), was used to observe behaviors indicative of the child's verbal working memory. To explain the game to the child, the adult told the following story: "We're going to take a trip to space in a big rocket ship. We need to prepare everything we'll need for our trip. I'm going to say the names of several of these

TABLE 1 | Descriptive characteristics of the sample.

N	M_{Age} (years) ± SD	Gender		M_{IQ} (IQ ± SD)	$M_{Birth\ weight}$ (kg) ± SD	$M_{Gestational\ age}$ (weeks) ± SD
		♂	♀			
44	5.73 ± 0.30	16	28	90.05 ± 8.5	3.15 ± 0.47	39.06 ± 1.91

things, and I want you to repeat them back to me in reverse order. I'll do two examples to help you understand better, and then you'll continue. Okay?" The words used (e.g., milk, water) were familiar concepts to children of this age and consistent with their level of vocabulary development.

Game 2: Our Travel Companions

This game, based on the Backward Animal Images Span task (Diamond, 2013), was used to observe behaviors indicative of the child's visuospatial working memory. To explain the game to the child, the adult told the following story: "Now we're going to meet our travel companions. I'm going to show you some photos of them. Take a good look because I'm going to set the photos on the table, and then I'll take them away. Then you'll have to arrange the photos in the opposite order from how I put them on the table. I'll do two examples to help you understand better, and then you'll continue. Okay?" The images all showed common animals that preschool children learn about in class (e.g., dog, pig).

Game 3: The Flight Begins

This game, based on a traditional imitation game called Simon Says (Strommen, 1973), was used to assess behavioral inhibition. To explain the game to the child, the adult told the following story: "Now we're flying in space! I'm going to indicate some actions and you have to do them. For example: If I say to touch your nose"—the adult performed this action while indicating it verbally—"you touch your nose." The child was then given time to perform the action. The adult then continued explaining the game: "Now I'm going to say some more actions, but only do them if I first say 'Simon Says'. If I don't say 'Simon Says' before indicating the action, don't do it; just hold still." The adult ordered an action while performing it simultaneously, but without first saying "Simon Says," leaving time for the child to remain still. In this game, therefore, in the absence of the words "Simon Says," the child was expected to be able to refrain from performing the action despite being told to and despite seeing the adult do it.

Game 4: The Day-Night Planet

This game, based on the Day-Night Task (Gerstadt et al., 1994; Carlson, 2005), was used to observe behaviors indicative of the child's capacity for resistance to distractor interference. To explain the game to the child, the adult told the following story: "We've landed on a new planet! On this planet, when you see the sun"—a picture of a sun appeared on a computer screen— "it's nighttime. When you see the moon"—a picture of a moon appeared on a computer screen—"it's daytime. The sun and the moon are going to appear on the screen quickly, one at a time. Pay attention, because when you see the sun"—the picture of the sun once again appeared on the screen—"you have to say 'night' as fast as you can, and when you see the moon"—the picture of the moon once again appeared on the screen—"you have to say 'day' as fast as you can." Two images of the sun and two images of the moon were shown alternately, as an example, to ensure that the participant had understood the instructions.

Game 5: Martians

This game is based on the Shape School game, which was created by Espy (1997) to assess cognitive flexibility and resistance to distractor interference in preschool children. To explain the game to the child, the adult told the following story: "Let's meet the inhabitants of this new planet!"—the adult showed the child a piece of cardboard with red, blue, and yellow squares and circles representing neutral facial expressions—"Look. These are the Martians who live on this planet. Their name is their color. Tell me the names of all the inhabitants of this planet as quickly as you can." The adult then displayed another piece of cardboard showing Martians with happy and sad faces. The adult said to the child: "Now some of the Martians are sad because they have to go home. Tell me, as quickly as possible, the name of the Martians with a happy expression but not the name of those with a frustrated face." This allowed the observation of behaviors related to resistance to distractor interference (i.e., resisting the sad faces and therefore not saying their color). Afterward, the adult displayed a third piece of cardboard showing some of the previous Martians, as well as some new Martians wearing hats. All of the Martians had a neutral face. The adult said to the child: "New Martians have arrived! These new Martians are wearing a hat, and their name is the shape of their figure. Take a good look and tell me the names of all the Martians as quickly as possible. Remember that the name of the Martians who aren't wearing a hat is their color and the name of the Martians wearing a hat is their shape." This allowed the observation of behaviors related to cognitive flexibility. Later, the adult displayed a fourth piece of cardboard showing both types of Martians (with and without a hat) with happy or frustrated faces. The adult said to the child: "Now there are Martians with a happy expression and others with a frustrated face. As quickly as possible, say the name of the happy Martians. Remember that the name of the Martians without a hat is their color and the name of the Martians with a hat is their shape." This allowed the observation of behaviors indicative of the child's capacity for resistance to distractor interference (as the child had to refrain from naming Martians with frustrated faces) and cognitive flexibility (as the child had to switch between shape and color to name the happy Martians depending on whether or not the Martian was wearing a hat).

Instruments for Collecting Data through Systematic Observation

In systematic observation (Anguera, 2003; Sánchez-Algarra and Anguera, 2013; Arias-Pujol and Anguera, 2017), a distinction is made between recording instruments (i.e., those used to record or code data) and observation instruments (purpose-designed instruments to analyze a given subject).

Recording Instruments

A Sony HDR-CX115 video camera was used to record the activity of each preschool child during the games.

The open-source software application Lince (Gabin et al., 2012) was used to code actions indicative of the preschool children's EFs. This program can be downloaded for free from http://lom.observesport.com/. Lince can be used to code all types of behavior as it is the observer who imports the video recordings

and corresponding observation instrument into the program. The program allows observers to simultaneously view the video recordings, the observation instrument, and the dataset being created.

Observation Instrument

As required by the nature of our systematic observation design, we built an *ad hoc* instrument fully adapted to the context of interest to capture the children's level of EF, using games (tasks) performed by the children. As the design was multidimensional, we built an instrument combining a field format and category systems (Sánchez-Algarra and Anguera, 2013; Castañer et al., 2016). The instrument had seven dimensions, each of which

formed the basis for a category system of exhaustive and mutually exclusive categories. The seven dimensions corresponded to three types of criteria: three fixed criteria, which remained unchanged throughout the observation session; one mixed criterion, which remained unchanged for part of the session; and three variable criteria, which changed frequently throughout the sessions and corresponded to the behaviors that were observed and coded. The observation instrument is shown in **Table 2**.

Standard Instruments

The two standard instruments used in this study justify the incorporation of selective methodology in the systematic observation and, consequently, a multimethod approach.

TABLE 2 | Observation instrument.

Criterion	Dimension	Category systems	Category code	Category description
Fixed	Participant	Participant	01	Participant 1 is performing the task and is being assessed
			02	Participant 2 is performing the task and is being assessed
			03	Participant 3 is performing the task and is being assessed
			04	Participant 4 is performing the task and is being assessed
			…	Participant … is performing the task and is being assessed
	Gender	Female	F	The participant performing the task is female
		Male	M	The participant performing the task is male
	Executive Function Task	Backward Word Span	BWS	Task for assessing verbal working memory
		Backward Animal Images Span:	BAIS	Task for assessing visuospatial working memory
		Simon Says	S	Task for assessing behavioral inhibition
		Day-Night Task	DNT	Task for assessing resistance to distractor interference
		Shape School	SS	Task for assessing cognitive flexibility primarily but also inhibition
Mixed	Phase	Phase 1	P1	Phase 1 of the task the participant is performing
		Phase 2	P2	Phase 2 of the task the participant is performing (only possible in the Shape School tasks)
		Phase 3	P3	Phase 3 of the task the participant is performing (only possible in the Shape School task)
		Phase 4	P4	Phase 4 of the task the participant is performing (only possible in the Shape School task)
Variable	Item	Item 1	1	Response stimulus 1 in task
		Item 2	2	Response stimulus 2 in task
		Item 3	3	Response stimulus 3 in task
		Item 4	4	Response stimulus 4 in task
		Item 5	5	Response stimulus 5 in task
		Item 6	6	Response stimulus 6 in task
		Item 7	7	Response stimulus 7 in task
		Item 8	8	Response stimulus 8 in task
		…	…	
		Item 40	40	Response stimulus 40 in task
	Performance	Correct	Co	The participant performs adequately for the specific task item
		Incorrect	Inc.	The participant does not perform adequately for the specific task item
		Self-Correct	Aut	The participant starts an action, realizes and corrects a mistake, and does not complete the action
		Omission	Om	The participant does not respond when required to do so by the specific task
	Adult	Explains	AdEx	The adult explains to the participant what the task consists of by giving instructions and explaining the rules of the game
		Corrects	AdCg	The adult corrects the participant when he/she does something wrong, leading the participant to either correct the mistake and or to modify his/her approach accordingly

BADyG: Assessment of Intellectual Ability

The BADyG (Yuste and Yuste, 2001) was used to assess intellectual ability and confirm that the children had an adequate IQ for their chronological age (third inclusion criterion). The BADyG is a Spanish battery of nine tests that have proven to provide a reliable measure (high Cronbach's alpha values) of the intellectual abilities of school children in numerous studies (Castejón et al., 2016; Veas et al., 2016). In our study, we used the level-1 battery designed for use in preschool children (BADyG-I).

The BADyG-I assesses three global performance items: (1) Verbal Intelligence, assessed through Numerical-Quantitative Concepts (1a), Information (1b), and Graphic Vocabulary (1c); (2) Non-verbal Intelligence, assessed through Non-verbal Mental Ability (2a), Reasoning with Figures (2b), and Logic Puzzles (2c); and (3) General Intelligence and IQ, assessed using the scores from the previous tests. Each test is composed of 18 items, each consisting of five pictures. The students must mark with an X the picture that matches the statement read out by the test administrator.

The children were also administered the complementary Perception and Coordination Graphomotor skills test to assess their ability to coordinate vision and manual movements during the reproduction of 12 simple geometric figures.

PAIB 1: Assessment of Academic Skills

The PAIB 1 (*Prueba de aspectos instrumentales básicos: Lectura, escritura y conceptos numéricos*; Galve-Manzano et al., 2009) was used to assess academic skills in reading, writing, and numeracy. These skills are considered to be the most important pillars for academic success (Cutler and Graham, 2008).

The PAIB 1 consists of eight subtests with activities that the children must complete with a pencil and paper. The activities are structured to resemble typical classroom activities. The PAIB 1 has demonstrated reliability (Galve-Manzano et al., 2009).

The eight subtests are:

1. *Basic Aspects of Mathematics* (three tests):
1.1. *Numeracy.* A 10-item test involving activities related to numbers with which the children are familiar.
1.2. *Calculation.* Six simple arithmetic operations.
1.3. *Problem resolution.* Eight math problems.
2. *Basic Aspects of Reading and Writing* (five tests):
2.1. *Reading comprehension.* Seven-item reading comprehension test consisting of sentences with different levels of grammatical complexity. Each item consists of four true or false sentences about a drawing. The children must choose the true sentences.
2.2. *Writing.* Subtest designed to assess writing skills through four tests or tasks involving different cognitive processes.
2.2.1. *Spelling.* This subtest consists of two tests: (a) *Word dictation*, where the children must write down words with different syllable structures read out by the test administrator, and (b) *Sentence dictation*, where the children must write down four sentences read out by the administrator.
2.2.2. *Vocabulary.* This subtest also consists of two tests: (a) *Sentence composition (I)*, in which the children must write

a sentence using all of three words provided, and (b) *Sentence composition (II)*, where the children are shown four drawings and asked to think about what is happening and to write a sentence for each of them.

A score is calculated for each of the eight tests, together with a total score for math, a total score for reading and writing, and a total score for math, reading, and writing combined.

Data Analysis Software

Ensuring the quality of the data collected is an essential part of systematic observation. We assessed this by calculating intra- and interobserver reliability for 30 sessions using the intraclass correlation coefficient in SAS 9.1.3 (Schlotzhauer and Littell, 1997; SAS Institute Inc., 2004).

The data used to address the first study objective were analyzed in the general linear model (GLM) in SAS 9.1.3 (Schlotzhauer and Littell, 1997; SAS Institute Inc., 2004).

The generalizability analysis to assess sample size (second study objective) was performed in EduG 6.0-e (Cardinet et al., 2010).

Procedure

The study was approved by the school management team and the parents of the children in the last year of preschool education were informed about the goals and nature of the study. They were asked to consent to their children participating in the study and to give their permission to have them video recorded while playing. They were also asked questions to assess compliance with the first two inclusion criteria: (1) attendance at the school since the second year of preschool education (age 3) and (2) absence of certain disorders or risk factors. Anonymity and compliance with ethical principles were guaranteed.

Students for whom parents gave their informed consent to participate in the study and who met the first two inclusion criteria were tested for IQ to ensure that they also met the third criterion, which was an adequate IQ for their chronological age. This was tested using the BADyG-I, which was administered to the group as a whole in two sessions held on non-consecutive days. Each session lasted approximately 30 min. The tests were administered according to the instructions in the BADyG-I manual for children in preschool education. They were administered by the same person, with the help of three others. They were scored automatically using the computer software feature provided with the BADyG-I. Each child was scored on verbal intelligence, non-verbal intelligence, and general intelligence. All the students had an adequate IQ for their chronological age and were therefore admitted into the study.

To fulfill the requirements of systematic observation, several exploratory play sessions were held prior to the definitive systematic observation. A child and the researcher were present in each session. These sessions were held at the school, but not in the children's usual classroom to avoid distractions. They were held during school hours and the children were allowed to take their usual breaks. The aim of this exploratory phase

was to guarantee the consistency of subsequent decisions and collect information to guide the construction of the observation instrument (Anguera, 2003). Specifically, the exploratory sessions were intended to verify that the children understood and were interested in the games, thus ensuring that they would participate readily and naturally. The children's involvement in the games is what would make it possible to systematically observe actions indicative of their EFs. The exploratory sessions also allowed the researchers to determine the approximate length of time that the children would spend on the games. On this basis, the researchers were able to determine how many sessions would be needed in order for each participant to play all the games. These steps were taken in order to ensure that the games could be included in the children's regular play routines without altering their activities or the context. Each day, the students' regular preschool schedule included periods of playtime as well as other regular activities that are common in school settings (psychomotor activities, reading and writing, lunch, rest periods, etc.). Thus, the exploratory sessions consisted of three children playing, on an individual basis, each of the five games described above after receiving the aforementioned explanations. The first child played all five games in a row, in a single session and in the following order: Games 1, 2, 3, 4, and 5. The session lasted 32 min, longer than the usual time allocated for play in the children's school routines. As a result, to avoid altering the children's daily school activities, we decided to offer the second child the chance to play the games in two sessions on different days. Thus, Games 1, 2, and 3 (involving the trip to space and the preparation thereof) were offered during the first session and Games 4 and 5 (set on the destination planet) were offered during the second session. The first session lasted 17.45 min and the second session lasted 8.20 min, and therefore was in line with the usual amount of playtime in the children's school routine. The same approach was used for the third child. The first session lasted 15.30 min and the second session lasted 7.10 min, thus respecting the usual amount of playtime in the daily school routine.

On the basis of this exploratory analysis involving three children who played the five games individually and freely, the following decisions were made:

(a) The games were deemed to be useful and appropriate for the systematic observation of preschool EFs, as they consisted of games that the children understood and found interesting. The children expressed their satisfaction and enjoyment of the games and exhibited spontaneous play activity.

(b) Each participant would be observed during two play sessions on different days. Each session would have a maximum duration of 20 min. With this arrangement, the students' usual play routines would be respected and their school activities would remain unchanged.

(c) Each participant would play Games 1, 2, and 3 in the first session and Games 4 and 5 in the second session.

The sessions were video recorded for later viewing. The recordings were used to integrate information about the children's EFs during completion of the different games and information from the theoretical framework on EFs in children with the ultimate aim of building the observation instrument. Different versions of the instrument were built and improved on until the definitive version shown in **Table 2** was achieved.

In the definitive systematic observation stage, each participant completed all the EF games. This was done at the school, again outside the children's classrooms and without interference from their teacher or other students. The games were played on two separate days. On the first day, the children played Game 1 (Preparing for the Journey), Game 2 (Our Travel Companions), and Game 3 (The Flight Begins) in a single session. The mean time spent on these games was 16.33 min. A week later, they played Game 4 (The Day-Night Planet) and Game 5 (Martians), again in a single session. The mean time spent on these games was 9.31 min. None of the sessions exceeded the length of the children's usual playtime; thus, their daily school routines were maintained. All the sessions were video recorded.

The video recordings were imported into Lince and coded using the *ad hoc* observation instrument for assessing EF (**Table 2**). The data recorded were converted into a matrix of codes that was tested for reliability (intraclass correlation coefficient ≥ 0.95).

The following year, when the children were in their first year of compulsory education, they were administered the PAIB 1 to test their reading, writing, and math skills. They completed the test as a group, in two sessions on non-consecutive days, and it was administered by the same adults who had administered the BADyG-I the previous year following the instructions in the manual. The first session lasted approximately 45 min and the second session was slightly shorter, at 40 min. The tests were corrected automatically by computer and a score was given for each of the eight tests, together with a total score for math, a total score for reading and writing, and a total score for these combined.

Data Analysis

GLMs were used to analyze the data to address the first study objective, which was to investigate whether different levels of EF in preschool children were associated with different levels of reading, writing, and mathematical skills the following year, at the start of compulsory education. GLMs indicate the percentage of variance in the dependent (response) variable (in our case, level of academic skills) that is explained by a series of independent (explanatory) variables (in our case, EF level and other variables that we will specify further on).

In order to estimate these models, it was first necessary to transform the data corresponding to the categories in the Performance dimension in the observation instrument into an appropriate format. To do this, we first transformed the data corresponding to the execution of each game into raw scores, assigning 2 points to the Correct category, 1 point to the Self-Correct category, and 0 points to the Incorrect and Omission categories. This resulted in a raw score per participant per EF game completed (EF level).

The converted data were now suitable for fitting various GLMs in SAS. Academic skills level was used as the response

variable in all the models. The explanatory variables were EF level in all cases and, depending on the model, participants, gender, EF game, and academic skills assessment instrument, together with their different interactions. The coefficient of determination (R^2) was calculated for all models. This coefficient (expressed as a percentage) indicates the extent to which the model (with its explanatory variables) explains the variance in the response variable (reading, writing, and mathematical skills).

To address the second objective of the study, i.e., to determine whether our systematic observations were generalizable to the reference population from which the sample was drawn, we calculated the generalizability coefficient using the G theory software program EduG. We used a measurement design with EF level and academic skills instrument as the differentiation facets and participants as the instrumentation facet.

RESULTS

Tables **3** and **4** show the most relevant results for the primary study objective, which consisted of estimating a GLM that would provide the best explanation for the variance in literacy and math skills.

Table 3 shows the two models that provided the best fit. The first had three explanatory variables (EF level, participants, and academic skills instrument, together with their interactions), while the second had five explanatory variables (EF level, EF games, participants, academic skills instrument, and gender, also with their respective interactions). In both cases, there were significant differences, indicating that level of academic skills in the first year of compulsory education was explained by three variables in the first model and five in the second one. The three-variable model accounted for 99% of the variance

TABLE 3 | Two type 1 overall general linear models, one with three variables and another with five.

Source	df	SS	MS	*F*-value	Pr > *F*
Three facets with interactions (EF level, participants, academic skills instrument)					
Model	1264	956167.91	756.46	234.71	<0.0001
Error	1617	5211.43	3.22		
Total corrected	2881	961379.34			
R^2					0.9946
Five facets with interactions (EF level, EF games, participants, academic skills instrument, and gender)					
Model	281	896834.70	3191.58	128.56	<0.0001
Error	2600	64544.64	24.82		
Total corrected	2881	961379.34			
R^2					0.9329

TABLE 4 | Two type 1 specific general linear models, one with three variables and another with five.

Source	df	SS	MS	*F*-value	Pr > *F*
Three facets: EF level, participants, and academic skills instrument with their relevant interactions					
EF level	2	1269.17	634.58	196.90	<0.0001
Participants	43	46290.20	1076.52	334.02	<0.0001
EF level × participants	69	1535.20	22.25	6.90	<0.0001
Academic skills instrument	10	837107.72	83710.77	25973.7	<0.0001
EF level × academic skills instrument	20	1917.42	95.87	29.75	<0.0001
Participants × academic skills instrument	430	66537.60	154.74	48.01	<0.0001
EF level × participants × academic skills instrument	690	1510.59	2.19	0.68	1.0000
Five facets: EF level, EF games, participants, academic skills instrument, and gender with their relevant interactions					
EF games	4	87.18	17.44	0.70	0.6217
EF level	2	1277.52	638.76	25.73	<0.0001
EF games × EF level	10	1518.12	151.81	6.12	<0.0001
Participants	43	44832.11	1042.61	42.00	<0.0001
EF games × participants	201	3763.32	18.72	0.75	0.9952
Academic skills instrument	10	837973.09	83797.31	3375.54	<0.0001
Gender × academic skills instrument	10	7383.36	738.34	29.74	<0.0001

($R^2 = 0.99$), while the five-variable model accounted for 93% ($R^2 = 0.93$). Both models, therefore, provided a very good fit, although the three-variable model slightly outperformed the five-variable one. The results suggest that the additional variables in the second model (EF games and gender) did not contribute anything to the overall model. On the contrary, they appeared to somehow distort it as it explained less of the variance.

The variables in the three-variable model (EF level, participants, and academic skills instrument) explained almost all of the variance in reading, writing, and math, and their power was not improved by the addition of more variables.

Table 4 shows the results for the three- and five-variable models, including the individual components of variance and their relevant interactions. Interactions that did not make a significant contribution have been omitted. In the three-variable model in **Table 4**, all the components and their interactions showed significant differences, except for the largest order interaction component EF level × participant × academic skills instrument (residual error of the model). In brief, EF level, participants, and academic skills instrument contributed significantly to explaining 99% of variations in literacy and math skills in the first year of compulsory education. The 1% of unexplained variance suggests the existence of variance

components that were not included in our study. This is supported by the fact that when we included other variables contemplated in our analysis (e.g., in the five-variable model), these not only reduced the fit of the model, but also, in some cases, offered no significant differences (**Table 4**), indicating that they did not explain variations in academic skills as they contributed nothing to the overall model. This was the case, for example, for EF game (0.6217) and EF game × participants (0.9952). A similar situation was seen for gender and a number of its interactions (e.g., gender × EF level), which were eliminated from **Table 4** as they did not make any relevant contribution to explaining the variance in academic skills. Significant differences were, however, obtained for gender × academic skills instrument, and for EF level, EF games × EF level, participants, and academic skills instrument, meaning that they also contributed to explaining variability.

The results for the generalizability analysis are summarized in **Tables 5** and **6**. **Table 5** shows the estimated variance components. The academic skills instrument has a large influence on the facet, accounting for 88.2% of all variance in the three-facet design. As can be seen, the rest of the facet and its interactions contributed very little to design variability.

Table 6 summarizes the results of the G study. The generalizability coefficient $[\xi \rho^2(\delta) = 1]$ indicates that the sample

TABLE 5 | Analysis and G study (generalizability theory) to estimate the generalizability of the results obtained for the sample of participants using the three-facet design, EF level (L) × participants (P) × academic skills instrument (S).

				Components				
Source	SS	df	MS	Random	Mixed	Corrected	%	SE
L	1269.00	2	634.50	1.08	1.08	1.08	0.2	0.93
P	46290.00	43	1076.51	27.44	27.44	27.44	3.8	6.89
S	837107.00	10	83710.70	632.29	632.29	632.29	88.2	258.90
LP	1535.00	86	17.85	1.46	1.46	1.46	0.2	0.24
LS	1917.00	20	95.85	2.14	2.14	2.14	0.3	0.66
PS	66537.00	430	154.74	50.99	50.99	50.99	7.1	3.51
LPS	1510.00	860	1.76	1.76	1.76	1.76	0.2	0.08
Total	956165.00	1451					100	

TABLE 6 | G study (generalizability theory) to estimate the generalizability of the results obtained for the sample of participants using the three-facet measurement design, EF level (L) × participants (P) × academic skills instrument (S).

Source of variance	Differentiation variance	Source of variance	Relative error variance	% Relative	Absolute error variance	% Absolute
L	1.08		.....		.....	
	.....	P	.....		0.62	33.6
S	632.29		.....		.....	
	.....	LP	0.03	2.7	0.03	1.8
LS	2.14		.....		.....	
	.....	PS	1.16	94.1	1.16	62.4
	.....	LPS	0.04	3.2	0.04	2.2
Sum of variances	635.51		1.23	100%	1.86	100%
Standard deviation	25.21		Relative SE: 1.11		Absolute SE: 1.36	
Coef_G relative $\xi \rho^2(\delta)$	1.00					
Coef_G absolute $\xi \rho^2(\Delta)$	1.00					

size (44 participants) was sufficient for accurately generalizing the results to the larger universe from which the sample was obtained.

DISCUSSION

The results of this study show that preschool EF level together with participants and academic skills instrument explained 99% of variations detected in literacy and math skills of children in their first year of compulsory education. In addition, our findings appear to be highly generalizable to the reference population from which the sample was drawn.

Overall, our results are consistent with reports in the literature that EFs have a key role in reading, writing, and math skills and that early assessment of these functions can help to identify children who are likely to present later learning difficulties (Engel de Abreu et al., 2014; McClelland et al., 2014; Viterbori et al., 2015; Moriguchi et al., 2016; Nieto et al., 2016; Purpura et al., 2017).

The interaction gender × EF level was not significant in our results, indicating the absence of significant differences in EF between boys and girls. While our systematic observation that EFs develop at a similar rate in boys and girls finds some support in the literature (Anderson, 2002; Li et al., 2009), several studies have reported that girls have slightly higher EF abilities than boys (Clark et al., 2013; Mansouri et al., 2016), at least in the case of certain components, such as inhibition (Mansouri et al., 2016). Nonetheless, there have also been reports of boys outperforming girls in components such as working memory (Dias et al., 2013). Differences in EF abilities between the sexes have been attributed to biological differences in frontal and temporal lobe function in children. They would thus be attributable to different brain growth patterns, which appear to follow the prefrontal cortex connections involved in the different EF components (Ngun et al., 2011). However, this is an area that requires further research.

The development of EFs is the result of interactions between biological growth factors and individual experiences, suggesting that they are malleable and as such candidates for targeted interventions (Diamond, 2013; Traverso et al., 2015; García-Madruga et al., 2016). The results of our study have important implications for educational practice. Assessment of EFs in preschool children may identify children whose EF level is lower than expected for their age and could therefore present later learning difficulties, enabling thus early interventions aimed at optimizing EFs with the ultimate goal of improving essential academic skills, such as reading, writing, and math. A growing number of interventional strategies are proving to be effective in this respect and many revolve around everyday activities, meaning they do not require a costly infrastructure (Anderson and Reidy, 2012; Diamond, 2014; Zelazo et al., 2016).

Children who start school with delays or gaps in the skills or EFs required for learning have been seen to continue to have difficulties throughout school, and the gaps tend to increase as the children move up through the school system (Clark et al., 2013). Therefore, early assessment of EFs, followed by early intervention when necessary, should be implemented as an educational action in all school systems. In addition, the benefits of early interventions persist into later life. In short,

early interventions targeting EFs can benefit children's cognitive, social, and emotional development, but they can also benefit development in later years, contributing to personal and career success, health, and quality of life in general (Diamond, 2013; Howard et al., 2015; Moriguchi et al., 2016). There is abundant literature showing that effective investment in early childhood education has a greater impact than later interventions and that the effects persist beyond the duration of the intervention, benefiting thus not only individuals but society as a whole. A country's socioeconomic progress and the well-being of its citizens are closely linked to academic achievements, which, in turn, are associated with adequate EF development in the early years of life. The benefits of early intervention in EFs thus far outweigh their potential costs, as the return on investment brings benefits to both children and the nation (Duncan and Magnuson, 2013; Bartik, 2014; Kaufman et al., 2015).

As children get older, they are presented with increasingly demanding tasks and academic challenges, but their EFs also improve. The improvement, however, is irregular (i.e., it is characterized by cycles of jumps and drops), so the relationship between these variables could show variations with age (Cragg and Gilmore, 2014; Viterbori et al., 2015; Moriguchi et al., 2016). It would also be interesting to investigate this aspect of EF further.

This study has contributed to knowledge in the area of preschool EFs and to systematic observation, as it demonstrates once again that the systematic observation of behavior in natural settings is a particularly apt scientific method for studies in the areas of development and education. This method offers endless opportunities for expanding knowledge in these areas, particularly in young children.

Systematic observation aims to describe and explain phenomena that occur in natural settings (Anguera, 2003) and aside from home, there is no more natural setting for children than school. School has an obvious impact on a child's development and life in general. Together with family, it is the factor that influences early development most (Bronfenbrenner, 1989). One of the greatest merits of systematic observation thus is that it captures development and learning almost as it occurs in everyday life, not in the controlled, artificial environment of a laboratory, enabling thus the rigorous analysis of everyday behavior in a person's natural settings (Anguera, 2003; Early Head Start National Resource Center, 2013). The most recent literature on the development of EFs highlights both the need for and the benefits of assessing EFs while children are performing routine, everyday activities in familiar contexts, as this is where EFs are developed, not in the controlled structure of a laboratory (Willoughby et al., 2012; Nieto et al., 2016). Before these recommendations, EFs were typically assessed using tasks completed by children in clinical or laboratory settings or surveys or questionnaires on their behavior filled in by third parties, such as parents or teachers. Both systems have their limitations. Tasks completed in a laboratory-like setting will reflect how a child behaves in this artificial, controlled setting but not in the real world, and any findings thus will have low ecological validity (Miranda et al., 2016). In the second case, while third parties can provide information on how children behave in a greater variety of situations, the reliability of this information is questionable

for numerous reasons. The answers might be biased by recall or a desire to answer what is "socially acceptable," for example, or the person may be unfamiliar with or fail to perceive certain behaviors (Wertz, 2014). Systematic observation, by contrast, has high ecological validity as it captures spontaneous behavior in natural settings. It also has an additional advantage that the behaviors are observed and coded by one or more people who are experts not only in the "what" but also in the "how" (Anguera, 2003). Our study thus contributes to advancing research in early EFs, as it was conducted in line with the latest guidelines for research in this area. We hope that more studies will take on board this recommendation to assess EFs in natural settings.

One issue of increasing concern to methodologists and researchers in field of education and development and in the social sciences in general is the quality of data gathered during the research phase, as this has an obvious impact on findings and subsequent decisions. The importance of reliable data is a given in all methodological approaches, but being able to offer the necessary guarantees of quality is particularly challenging when studying spontaneous behavior in natural settings. When perceivable human behavior is observed without the constraints imposed by external controls, the data collected are more likely to contain more errors and more serious errors, potentially jeopardizing the validity of the research.

One means of addressing the different risks that can affect the accuracy of a dataset is to design a quality control procedure that analyzes how different facets or potential sources of variance affect different measurements designs and also provides a measure of the magnitude of error. The relatively recent use of G theory to calculate the reliability and validity of observational data is an important step in this direction, as is its lesser known application for estimating effective sample size. We used this novel feature of G theory to avoid underpowering, which is a frequent limitation of studies conducted in children.

AUTHOR CONTRIBUTIONS

EE-P contributed to conceptual structure, collecting data, and systematic observation. MH-N involved in collecting data. AB-V performed data analysis and results. MTA contributed to conceptual structure and systematic observation. All authors contributed to documenting, drafting and writing the manuscript, and gave their approval to the final version to be published.

FUNDING

REFERENCES

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychol.* 8, 71–82. doi: 10.1076/chin.8.2.71.8724

Anderson, P. J., and Reidy, N. (2012). Assessing executive function in preschoolers. *Neuropsychol. Rev.* 22, 345–360. doi: 10.1007/s11065-012-9220-3

Anguera, M. T. (2003). "Observational methods (General)," in *Encyclopedia of Psychological Assessment*, Vol. 2, ed. R. Fernández Ballesteros (London: Sage), 632–637.

Arias-Pujol, E., and Anguera, M. T. (2017). Observation of interactions in adolescent group therapy: a mixed methods study. *Front. Psychol.* 8:1188. doi: 10.3389/fpsyg.2017.01188

Bakeman, R., and Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge: Cambridge University Press.

Bartik, T. K. (2014). *From Preschool to Prosperity: The Economic Payoff to Early Childhood Education*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Bick, J., and Nelson, C. A. (2017). Early experience and brain development. *WIREs Cogn. Sci.* 8:e1387. doi: 10.1002/wcs.1387

Blair, C., and Raver, C. C. (2015). School readiness and self-regulation: a developmental psychobiological approach. *Annu. Rev. Psychol.* 66, 711–731. doi: 10.1146/annurev-psych-010814-015221

Blanco-Villaseñor, A., Castellano, J., Hernández Mendo, A., Sánchez-López, C. R., and Usabiaga, O. (2014). Aplicación de la TG en el deporte para el estudio de la fiabilidad, validez y estimación de la muestra. *Rev. Psicol. Deporte* 23, 131–137.

Blanco-Villaseñor, A., Losada, J. L., and Anguera, M. T. (2003). Data analysis techniques in observational designs applied to the environment-behaviour relation. *Medio Ambient. Comport. Hum.* 4, 111–126.

Bronfenbrenner, U. (1989). Ecological systems theory. *Ann. Child Dev.* 6, 187–249.

Cardinet, J., Johnson, S., and Pini, G. (2010). *Applying Generalizability Theory using EduG*. Londres: Routledge.

Carlson, S. A. (2005). Developmentally sensitive measures of executive function in preschool children. *Dev. Neuropsychol.* 28, 595–616. doi: 10.1207/s15326942dn2802_3

Carlson, S. M., Faja, S., and Beck, D. M. (2016). "Incorporating early development into the measurement of executive function: the need for a continuum of measures across development," in *Executive Function in Preschool-Age Children: Integrating Measurement, Neurodevelopment, and Translational Research*, eds J. A. Griffin, P. McCardle, and L. S. Freund (Washington, DC: American Psychological Association), 45–64.

Cartwright, K. B. (2015). *Executive Skills and Reading Comprehension: A Guide for Educators*. New York, NY: Guilford Press.

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Canton, A., and Hileno, R. (2016). Goal Scoring in Soccer: a polar coordinate analysis of motor skills used by Lionel Messi. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Castejón, J. L., Gilar, R., Veas, A., and Miñano, P. (2016). Differences in learning strategies, goal orientations, and self-concept between overachieving, normal-achieving, and underachieving secondary students. *Front. Psychol.* 7:1438. doi: 10.3389/fpsyg.2016.01438

Clark, C. A. C., Sheffield, T. D., Wiebe, S. A., and Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Dev.* 84, 662–677. doi: 10.1111/j.1467-8624.2012.01854.x

Colé, P., Duncan, L. G., and Blaye, A. (2014). Cognitive flexibility predicts early reading skills. *Front. Psychol.* 5:565. doi: 10.3389/fpsyg.2014.00565

Cragg, L., and Gilmore, C. (2014). Skills underlying mathematics: the role of executive function skills in the development of mathematics proficiency. *Trends Neurosci. Educ.* 3, 63–68. doi: 10.1016/j.tine.2013.12.001

Cutler, L., and Graham, S. (2008). Primary grade writing instruction: a national survey. *J. Educ. Psychol.* 100, 907–919. doi: 10.1037/a0012656

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Diamond, A. (2014). Want to optimize executive functions and academic outcomes? Simple, just nourish the human spirit. *Minn. Symp. Child Psychol.* 37, 205–232.

Dias, N. M., Menezes, A., and Seabra, A. G. (2013). Age differences in executive functions within a sample of Brazilian children and adolescents. *Span. J. Psychol.* 16:E9. doi: 10.1017/sjp.2013.12

Duncan, G. J., and Magnuson, K. (2013). Investing in preschool programs. *J. Econ. Perspect.* 27, 109–132. doi: 10.1257/jep.27.2.109

Early Head Start National Resource Center (2013). *Observation: The Heart of Individualizing Responsive Care.* Washington, DC: Early Head Start National Resource Center at ZERO TO THREE.

Elliott, J. (2007). Multimethod approaches in educational research. *Int. J. Disabil. Dev. Educ.* 51, 135–149. doi: 10.1080/10349120410001687364

Engel de Abreu, P. M., Abreu, N., Nikaedo, C. C., Puglisi, M. L., Tourinho, C. J., Miranda, M. C., et al. (2014). Executive functioning and reading achievement in school: a study of Brazilian children assessed by their teachers as "poor readers". *Front. Psychol.* 5:550. doi: 10.3389/fpsyg.2014.00550

Espy, K. A. (1997). The shape school: assessing executive function in preschool children. *Dev. Neuropsychol.* 13, 495–499. doi: 10.1080/87565649709540690

Fasulo, A., Shukla, J., and Bennett, S. (2017). Find the hidden object. Understanding play psychological assessments. *Front. Psychol.* 8:323. doi: 10.3389/fpsyg.2017.00323

Federici, S., Meloni, F., Catarinella, A., and Mazzeschi, C. (2017). Models of disability in children's pretend play: measurement of cognitive representations and affective expression using the affect in play scale. *Front. Psychol.* 8:794. doi: 10.3389/fpsyg.2017.00794

Friedman, N. P., and Miyake, A. (2004). The relations among inhibition and interference control function: a latent-variable analysis. *J. Exp. Psychol. Gen.* 133, 101–135. doi: 10.1037/0096-3445.133.1.101

Gabin, B., Camerino, O., Anguera, M. T., and Castañer, M. (2012). Lince: multiplatform sport analysis software. *Proc. Soc. Behav. Sci.* 46, 4692–4694. doi: 10.1016/j.sbspro.2012.06.320

Galve-Manzano, J. L., Ramos-Sánchez, J. L., Martínez-Arias, R., and Trallero-Sanz, M. (2009). *Prueba de Aspectos Instrumentales Básicos. Lectura, Escritura y Conceptos Numéricos. PAIB 1. [Test of Basic Instrumental Aspects. Reading, writing and numerical concepts].* Madrid: CEPE.

García-Madruga, J. A., Gómez-Veiga, I., and Vila, J. O. (2016). Executive functions and the improvement of thinking abilities: the intervention in reading comprehension. *Front. Psychol.* 7:58. doi: 10.3389/fpsyg.2016.00058

Garris, R., Ahlers, R., and Driskell, J. E. (2002). Games, motivation, and learning: a research and practice model. *Simulat. Gam.* 33, 441–467. doi: 10.1177/1046878102238607

Gerstadt, C. L., Hong, Y. J., and Diamond, A. (1994). The relationship between cognition and action: performance of children 3½-7 years old on a stroop-like day-night test. *Cognition* 53, 129–153. doi: 10.1016/0010-0277(94)90068-X

Harper, L. J. (2016). Supporting young children's transitions to school: recommendations for families. *Early Childh. Educ. J.* 44, 653–659. doi: 10.1007/s10643-015-0752-z

Howard, S. J., Okely, A. D., and Ellis, Y. G. (2015). Evaluation of a differentiation model of preschoolers' executive functions. *Front. Psychol.* 6:285. doi: 10.3389/fpsyg.2015.00285

Jablon, R., Dombro, A., and Dichtelmiller, M. (2013). *The Power of Observation: Birth to Age 8.* Washington, DC: Teaching Strategies.

Kaufman, M. J., Kaufman, S. R., and Nelson, E. C. (2015). Beginning together: reforming schools by investing in early childhood education. *Schools* 12, 133–149. doi: 10.1086/680698

Kumschick, I. R., Beck, L., Eid, M., Witte, G., Klann-Delius, G., Heuser, I., et al. (2014). READING and FEELING: the effects of a literature-based intervention designed to increase emotional competence in second and third graders. *Front. Psychol.* 5:1448. doi: 10.3389/fpsyg.2014.01448

Li, C. S., Zhang, S., Duann, J. R., Yan, P., Sinha, R., and Mazure, C. M. (2009). Gender differences in cognitive control: an extended investigation of the stop signal task. *Brain Imaging Behav.* 3, 262–276. doi: 10.1007/s11682-009-9068-1

Mangelsdorf, J., and Eid, M. (2015). What makes a thriver? Unifying the concepts of posttraumatic and postecstatic growth. *Front. Psychol.* 6:813. doi: 10.3389/fpsyg.2015.00813

Mansouri, F. A., Fehring, D. J., Gaillard, A., Jaberzadeh, S., and Parkington, H. (2016). Sex dependency of inhibitory control functions. *Biol. Sex Differ.* 7:11. doi: 10.1186/s13293-016-0065-y

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., et al. (2014). Predictors of early growth in academic achievement: the head-toes-knees-shoulders task. *Front. Psychol.* 5:599. doi: 10.3389/fpsyg.2014.00599

Miranda, A., Colomer, C., Mercader, J., Fernández, I., and Presentación, M. J. (2016). Performance-based tests versus behavioral ratings in the assessment of executive functioning in preschoolers: associations with ADHD symptoms and reading achievement. *Front. Psychol.* 6:545. doi: 10.3389/fpsyg.2015.00545

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Moriguchi, Y., Chevalier, N., and Zelazo, P. H. (2016). Editorial: development of executive function during childhood. *Front. Psychol.* 7:6. doi: 10.3389/fpsyg.2016.00006

Nell, M., and Drew, W. (2013). *From Play to Practice: Connecting Children's Play to Children's Learning.* Washington, DC: NAEYC.

Ngun, T. C., Ghahramani, N., Sánchez, F. J., Bocklandt, S., and Vilain, E. (2011). The genetics of sex differences in brain and behavior. *Front. Neuroendocrinol.* 32:227–246. doi: 10.1016/j.yfrne.2010.10.001

Nieto, M., Ros, L., Medina, G., Ricarte, J. J., and Latorre, J. M. (2016). Assessing executive functions in preschoolers using shape school task. *Front. Psychol.* 7:1489. doi: 10.3389/fpsyg.2016.01489

Otsuka, K., and Jay, T. (2017). Understanding and supporting block play: video observation research on preschoolers' block play to identify features associated with the development of abstract thinking. *Early Child Dev. Care* 187, 990–1003. doi: 10.1080/03004430.2016.1234466

Portell, M., Anguera, M. T., Chacón-Moscoso, S., and Sanduvete-Chaves, S. (2015). Guidelines for reporting evaluations based on observational methodology. *Psicothema* 27, 283–289. doi: 10.7334/psicothema2014.276

Purpura, D. J., Schmitt, S. A., and Ganley, C. M. (2017). Foundations of mathematics and literacy: the role of executive functioning components. *J. Exp. Child Psychol.* 153, 15–34. doi: 10.1016/j.jecp.2016.08.010

Rapoport, S., Rubinsten, O., and Katzir, T. (2016). Teachers' beliefs and practices regarding the role of executive functions in reading and arithmetic. *Front. Psychol.* 7:1567. doi: 10.3389/fpsyg.2016.01567

Sánchez-Algarra, P., and Anguera, M. T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: impact of recording and coding among predominating perspectives. *Qual. Quant.* 4, 1237–1257. doi: 10.1007/s11135-012-9764-6

SAS Institute Inc. (2004). *SAS 9.1.3 Help and Documentation.* Cary, NC: SAS Institute Inc.

Schlotzhauer, S. D., and Littell, R. C. (1997). *SAS System for Elementary Statistical Analysis.* Cary, NC: SAS Institute Inc.

Shaughnessy, J., Zechmeister, E., and Zechmeister, J. (2009). *Research Methods in Psychology.* Boston, MA: McGraw-Hill.

Shaw, C., Brady, L. M., and Davey, C. (2011). *Guidelines for Research with Children and Young People.* London: National Children's Bureau Research Centre.

Strommen, E. A. (1973). Verbal self-regulation in a children's game: impulsive errors on "simon says". *Child Dev.* 44, 849–853. doi: 10.2307/1127737

Traverso, L., Viterbori, P., and Usai, M. C. (2015). Improving executive function in childhood: evaluation of a training intervention for 5-year-old children. *Front. Psychol.* 6:525. doi: 10.3389/fpsyg.2015.00525

Varvara, P., Varuzza, C., Sorrentino, A. C. P., Vicari, S., and Menghini, D. (2014). Executive functions in developmental dyslexia. *Front. Hum. Neurosci.* 8:120. doi: 10.3389/fnhum.2014.00120

Veas, A., Gilar, R., Miñano, P., and Castejón, J. L. (2016). Estimation of the proportion of underachievers in compulsory secondary education in Spain: an application of the Rasch model. *Front. Psychol.* 7:303. doi: 10.3389/fpsyg.2016.00303

Veličković, S. (2013). Problems discontinuity on the first level of the school system. *Int. J. Cogn. Res. Sci. Eng. Educ.* 1, 135–143.

Visu-Petra, L., Stanciu, O., Benga, O., Miclea, M., and Cheie, L. (2014). Longitudinal and concurrent links between memory span, anxiety symptoms, and subsequent executive functioning in young children. *Front. Psychol.* 5:443. doi: 10.3389/fpsyg.2014.00443

Viterbori, P., Usai, M. C., Traverso, L., and De Franchis, V. (2015). How preschool executive functioning predicts several aspects of math achievement in Grades 1 and 3: a longitudinal study. *J. Exp. Child Psychol.* 140, 38–55. doi: 10.1016/j.jecp.2015.06.014

Wertz, F. J. (2014). Qualitative inquiry in the history of psychology. *Qual. Psychol.* 1, 4–16. doi: 10.1007/s12124-014-9293-z

Willoughby, M. T., and Blair, C. B. (2016). Measuring executive function in early childhood: a case for formative measurement. *Psychol. Assess.* 28, 319–330. doi: 10.1037/pas0000152

Willoughby, M. T., Blair, C. B., Wirth, R. J., and Greenberg, M. (2012). The measurement of executive function at age 5: psychometric properties and relationship to academic achievement. *Psychol. Assess.* 24, 226–239. doi: 10.1037/a0025361

Yuste, C., and Yuste, D. (2001). *Batería de Aptitudes Diferenciales y Generales BADyG. [Battery of General and Differential Aptitudes BADyG].* Madrid: CEPE.

Zelazo, P. D., Blair, C. B., and Willoughby, M. T. (2016). *Executive Function: Implications for Education.* Washington, DC: U.S. Department of Education.

Check for updates

Is Reading Instruction Evidence-Based? Analyzing Teaching Practices Using T-Patterns

Natalia Suárez [1]*, Carmen R. Sánchez [2], Juan E. Jiménez [1] and M. Teresa Anguera [3]

[1] Departamento de Psicología Evolutiva y de la Educación, University of La Laguna, San Cristóbal de La Laguna, Spain, [2] Departamento de Psicología Clínica, Psicobiología y Metodología, University of La Laguna, San Cristóbal de La Laguna, Spain, [3] Faculty of Psychology, Institute of Neurosciences, University of Barcelona, Barcelona, Spain

The main goal of this study was to analyze whether primary teachers use evidence-based reading instruction for primary-grade readers. The study sample consisted of six teachers whose teaching was recorded. The observation instrument used was developed *ad hoc* for this study. The recording instrument used was Match Vision Studio. The data analysis was performed using SAS, GT version 2.0 E, and THEME. The results indicated that the teaching practices used most frequently and for the longest duration were: feedback (i.e., correcting the student when reading); fluency (i.e., individual and group reading, both out loud and silently, with and without intonation); literal or inference comprehension exercises (i.e., summarizing, asking questions); and use of educational resources (i.e., stories, songs, poems). Later, we conducted analyses of T-Patterns that showed the sequence of instruction in detail. We can conclude that <50% of the teaching practices used by the majority of teachers were based on the recommendations of the National Reading Panel (NRP). Only one teacher followed best practices. The same was the case for instructional time spent on the five essential components of reading, with the exception of teacher E., who dedicated 70.31% of class time implementing best practices. Teaching practices (i.e., learners' activities) designed and implemented to exercise and master alphabetic knowledge and phonological awareness skills were used less frequently in the classroom.

Keywords: teaching practices, reading instruction, T-Patterns, observational methodology, resources, National Reading Panel, components of reading, teaching experience

INTRODUCTION

There has always considerable interest in exploring how to teach reading and thus bring pupils to appropriate levels of reading proficiency (EACEA/Eurydice, 2011). The National Institute of Child Health and Human Development (2000) identified basic skills that constitute reading competency and the best practices in literacy instruction. Generally speaking, different programs have been developed that propose different ways of targeting the teaching of reading. The direct instruction method (Carnine and Kameenui, 1992; Chard and Jungjohann, 2006; Coyne et al., 2007), based on behavioral theory, is a form of instruction where the teacher is the main axis and works through modeling; this method explicitly uses practices for teaching reading that break the process down into small units, and follows a clear sequence involving repetition and reinforcement. Scaffolding (Temple et al., 2011), based on constructivist principles, consists of having children build their own

learning with the help and guidance of their teacher. Psychomotricity, the development of spatial orientation, handedness, and the growing awareness of one's body (Pinker, 2001; Scarborough, 2002; Slavin, 2003), together with respect for one's own pace of learning, are practices based on maturational theory (Fons, 2008). Innatist theory focuses on teaching reading at an early age (Al Otaiba and Fuchs, 2002; De Arcangelo, 2003; Foorman et al., 2003; Fons, 2012; Pascual et al., 2013). Proponents of sociocultural theory promote practices that encourage family, social, cultural, and educational involvement, as all these will play a role in a child's reading development (Purcell-Gates et al., 2002; Fetsco and McClure, 2005; Porta and Ison, 2011; Greenhoot et al., 2014); hence the importance of providing a book-rich environment (Dickinson and Tabors, 2001) that is high in both quality and quantity (Porta, 2008). Finally, the development of phonological awareness through the teaching of sounds is one of the premises of psycholinguistic theory (Pearson, 2001; Rayner et al., 2002; Fletcher-Flinn, 2014). Despite the existence of countless different approaches for teaching reading, the results of international and national tests of basic reading skills [International Association for the Evaluation of Educational Achievement (IEA); Reading Achievement, Progress in International Reading Literacy Study (Mullis et al., 2012; National Assessment of Educational Progress, 2015); Programme for International Student Assessment (Organization for Economic Co-operation and Development, 2012, 2015)]; indicate that it is necessary to improve and promote effective practices.

EVIDENCE-BASED PRACTICES FOR TEACHING READING

Scientific research has demonstrated that the teaching of reading should begin at an early age through teaching practices designed and implemented to exercise and master basic skills that constitute reading competency as defined by the National Reading Panel (2000) (i.e., phonological awareness, alphabetic knowledge, vocabulary, fluency, and comprehension). Phonological awareness means the ability to detect and manipulate sound segments of spoken words (Pufpaff, 2009). Findings on phonological awareness have shown that this is a key skill in the early years of a child's schooling. Many studies have confirmed that this skill is a good predictor of future reading performance (Porta et al., 2010; Suárez et al., 2013; Kjeldsen et al., 2014; Del Campo et al., 2015). Alphabetic knowledge refers to the knowledge of the rules for grapheme-phoneme (G-P) and phoneme-grapheme (P-G) conversion; fluency, which is described as the ability to read texts rapidly and accurately, using appropriate intonation within the reading context; vocabulary, i.e., learning the meaning and use of words in a given context; and comprehension, which refers to a child's ability to reason about, reflect on, and understand what they are reading (Jiménez et al., 2012). Teaching these essential components of reading not only helps children learn to read (National Reading Panel, 2000) it is also helpful for children at risk of exhibiting learning difficulties (Adams, 2001; Foorman and Torgesen, 2001; Tunmer

and Arrow, 2013). Numerous recommendations of instructional practices to promote these basic skills have emerged from research findings. The National Reading Panel (2000) developed specific recommendations for activities to teach phonemic awareness. These include isolating, identifying, categorizing, substituting, adding, and deleting phonemes. In the same vein, it has been found that when two or more tasks of segmenting (e.g., dividing a word up into sounds) and deletion (e.g., removing a sound from a given word) are combined, the effect size is much greater.

With respect to alphabetic knowledge, findings have shown that it is better to combine the teaching of sounds with that of the printed letter (Ehri et al., 2001; Stevenson, 2004; Caravolas et al., 2005; Hatcher et al., 2006). It has been shown that the most effective of all the programs using different phonics methods to teach this skill for teaching reading are those that are: synthetic (converting letters to sounds, mixing sounds to form words), analytic (identifying words and their sounds), spelling-based (transforming sounds into letters), contextual (using sound-letter correspondence and finding unknown words in a text), and analogical (using parts of written words to find new ones) (National Reading Panel, 2000). In addition, it's important to note that using a systematic instructional sequence (i.e., easier to more complex and most common letters and letter patterns first) providing ample opportunities for practice and employing evidence-based methods of phonics instruction results in better student outcomes (Armbruster et al., 2001).

Fluency is another skill that predicts reading success. Teachers should teach their pupils to read texts accurately, quickly, and effortlessly, using the correct pronunciation (Nichols et al., 2008), and rapidly, precisely, and with the appropriate intonation (Allington, 1983). It has been shown that guided oral instruction, the use of tutoring, and the involvement of the child's immediate environment have a positive influence on rapid reading (National Reading Panel, 2000). Consolidating this skill also contributes to improving comprehension, as the pupil can free up more cognitive resources for understanding a text (National Institute of Child Health and Human Development, 2000; Hirsch, 2007). Teachers need to use activities focused on: repeated reading of the same text (Rasinski, 2003), independent reading of carefully selected text (Allington, 2000), or practicing expression (Schwanenflugel and Benjamin, 2012), and repeated oral reading with feedback (Armbruster et al., 2001).

Teaching vocabulary also has a direct influence on reading comprehension and vice versa (Perfetti et al., 2005; Hirsch, 2007; Strasser et al., 2013). This skill should be taught early on, and it should focus on the use of strategies such as the use of new technologies, the indirect method, and repeated exposure to words and their meanings (Joshi, 2005; Perfetti et al., 2005; Hirst, 2007; Strasser et al., 2013). Instruction should include multiple exposures to a word, careful selection of words, deepening the meaning of the words, connecting familiar and new words and teaching compound or familiar words (Lane, 2014).

As for comprehension, defined as the skill in which intentional thinking is developed, whereby the meaning of words is constructed through interaction between the text and the reader (Durkin, 1993), a number of practices have proven effective, such

as: monitoring comprehension, cooperative learning, the use of graphic and semantic organizers, the use of question-and-answer formats, generating questions, recognizing story structure, and summarizing. In addition, teachers need to help children: activate their prior knowledge, provide ample opportunities to use comprehension strategies (i.e., lower, summarize), read and work with different types of texts (i.e., narratives, expository), use questions to facilitate discussion (Shanahan et al., 2010).

In sum, teachers need to incorporate activities aimed at helping children to discover the sounds of phonemes, associating sounds with the corresponding graphic symbols, creating a link between readings of texts or stories, working with previous knowledge and lexicon, this will help the development of skills such as: phonological awareness, alphabetic knowledge, fluency, comprehension and vocabulary (National Reading Panel, 2000). It has been said that it is important not only that teachers be aware of and understand these components, but also that they know how to work with them to contribute to reading success (Cunningham et al., 2009; Joshi et al., 2009; Kaiser et al., 2009; Podhajski et al., 2009). We must first find out how teachers evaluated actually teach reading, and establish whether their teaching practices are based on the recommendations of scientific research; this is the main aim of the present study.

MATERIALS AND METHODS

We employed systematic observation, which is widely used in a range of contexts (Castañer et al., 2013, 2016), as it fulfills the basic requirements proposed by Anguera (1979, 2003): habitual behavior, natural context, and perceptivity. These conditions are all guaranteed in the events tracked in our study. The choice of methodology is also justified, as we used an *ad hoc* observation instrument to record, analyze, and interpret how teachers of the sample teach reading.

The observational design can be classified as Nomothetic/Follow-up/Multidimensional (N/F/M) (Blanco-Villaseñor et al., 2003; Sánchez-Algarra and Anguera, 2013; Portell et al., 2015), where *nomothetic* refers to the observation of various different teachers; *follow-up* refers to recording the behaviors or situations that arise over a period of time; and *multidimensional* refers to the fact that more than one dimension of the participant's response is taken into account. We carried out non-participatory observation of teachers on the island of Tenerife (Canary Islands, Spain), in the classroom context, while they were teaching their pupils how to read. Observation was active, governed by scientific criteria, characterized by total perceptibility, and performed by direct observation of the film shot.

Participants

Our study involved six teachers aged 25 to 50, with 3 to 25 years of teaching experience. Each of them interests us as a case study, individually, and without any pretension of generalizing the results.

The teachers were employed at different preschools and elementary schools on the island of Tenerife. Two of these schools were in a suburban area, one was in a rural area and one was in an urban area. The selection criteria essentially involved ensuring the participants were teachers of language arts and that they spent an average of 1 hour each day on teaching reading.

Materials

The classroom sessions were recorded using four digital video cameras, four stands and two recorders. Both hardware (two complete computer workstations and two pairs of headphones) and software (Windows Movie Maker by Microsoft, for video editing) were used to observe the teachers' behavior. For recording the data, we used Match Vision 3.0 (Perea et al., 2006). The data quality analysis was done using Generalizability Study (GT) version 2.0.E (Ysewijn, 1996) and the (SAS Institute Inc, 1999) 9.1 statistical package. THEME (Magnusson, 1988) was used to analyze the teachers' behavioral patterns.

The observation of a natural context requires the use of an observation instrument. The observation tool used here was *ad hoc* and combines a field format and systems of categories. The field format is formed by the dimensions of the instrument, and a system of categories has been constructed from each one of them.

This instrument was created using the information obtained from the reality observed, and the dimensions are based on innatist, maturational, behaviorist, sociocultural, corrective, repetitive, and psycholinguistic theories (see Suárez et al., 2013; Jiménez et al., 2014). Systems of categories are characterized by their high degree of structure and their adaptation to the previously defined research question (Anguera, 2003). They also respect the assumptions of mutual exclusivity (e.g., a single behavior cannot be associated with two categories) and exhaustiveness (e.g., a category system covers all possible behaviors ascribed to it). This instrument covers the practices carried out by teachers when teaching reading, and is made up of 14 dimensions. Each criterion has allowed the construction of an exhaustive and mutually exclusive category system.

Below is a presentation of the instrument used in this study (see **Table 1**). The acronyms shown in the following table (which reflect the wording of the categories in Spanish) were used to record the behavior in the Match Vision Studio program (Perea et al., 2006).

The dimensions refer to whether the teacher carried out teaching practices based on: phonological awareness, alphabetic knowledge, fluency, vocabulary, and comprehension activities. In addition, the observation instrument includes other reading teaching practices based on the use of resources, reinforcement, feedback, modeling, guided oral instruction, homework, reading and writing, and psychomotricity activities.

Procedure

Before the recordings were made, authorization was obtained from both the teachers and the pupils' parents. All participants provided a written informed consent prior to their participation. The dates and times of the recording sessions were scheduled in advance (taking into account the school timetable). On the first days, the cameras were tested to ensure they were being used properly. Afterward, the cameras were set up in the classrooms 10 minutes prior to the start of the agreed session.

TABLE 1 | Observation instrument of practices used for teaching reading.

Dimensions	Categories	Acronyms (in Spanish)
Alphabetic knowledge activities	Teaching sounds	ES
	Teaching letter names	ENL
	Teaching the rule (with graphic symbol)	ERAG
	Teaching the rule (without graphic symbol)	ERSAG
	Rhyming: linking words that rhyme (words can be written in any format: on the board, in notebooks, etc.)	RP
Phonological awareness activities	Stimulating children to become aware of sound (without visual aid)	ECSSV
	Saying words that start with or contain a given sound	DPESD
	Dividing words into syllables	SPS
	Playing "I spy"	JVV
	Separating words in a sentence (clapping or counting)	SPF
	Building words through sounds	CPS
	Rhyming: saying words with the same endings (without graphical symbol as visual aid)	RM
	Saying a sentence with a given words or sound	DFSP
	Alphabet cards	TA
Use of resources	Using made-up stories	HCI
	Songs	CC
	Poetry	PES
	Written material from different sources (menu, children's letters)	EDFMC
	Using written texts or stories	UTCE
	Getting children involved (correcting errors, helping each other, play-acting or role-playing)	HPCAD
	Riddles	ADZ
	Metaphors	UM
	Stories with a moral	MRA
Comprehension previous knowledge activities	Using words the children say	UPN
	Asking children to talk about their experiences	LPNEV
	Reviewing content covered in the past	REPA
Reinforcement	Positive (verbal or tangible)	P(VM)
	Negative (verbal)	NE-VE
Feedback	Correcting reading (mispronunciation, getting lost, skipping a line)	CLPDS
	Pointing out where child went wrong	SDSE
	Offering examples	PEJP
	Alerting child to an error ("no, that's not right")	AA-NO
	Review and correction of children's work	RCTN
	Asking questions	PG
	The teacher writes the word or sentence wrong	PEMLP
	Repeating the reading	RL
Modeling	The teacher shows the steps to be taken, when the child makes a mistake or to remind the children of content	PEPEQ
Fluency activities	Individual reading (out loud)	LI-VA
	Individual reading (out loud and rapidly)	LIVAR
	Individual reading (silently)	LISIL
	Group reading (out loud)	LG-VA
	Group reading (out loud and rapidly)	LGVAR
	Group reading (silently)	LGSIL
	Fluent reading (rapid, accurate and precise)	FLUI

(Continued)

TABLE 1 | Continued

Dimensions	Categories	Acronyms (in Spanish)
	Rapid reading (the teachers says "faster")	LR
	Individual reading with intonation	LEFI
	Group reading with intonation	LEFG
	Individual reading without intonation	LSEFI
	Group reading without intonation	LSEFG
Guided oral instruction	Teaching the steps to be followed (saying them)	EDPS
	Instruction on how to intone sentence or text	IEF
Homework	Doing different activities at home	RVAC
	Reading at home	LECS
Reading and writing activities	Read the word and then write it	LPLE
	Read the sentence and then write it	LFLE
	Writing and reading a sentence	EFL
	Writing and reading a word	EPL
	Dictation of letters, words or sentences	DICPF
	Copying letters, words or sentences	CLPF
	Completing activities or exercises	CA
Psychomotricity activities	Spatial orientation (up, down, left, right, in, out)	OE-AB
	Temporal orientation (yesterday, tomorrow, next year, in a month's time)	OT-AM
	Rhythm (rhythmic sequences)	R-SR
	Body awareness (knowing parts of the body)	EC-CP
Literal or inference comprehension activities	Teaching the parts of a book	EPLB
	Relating illustrations to text	RITEX
	Using news and current events	UNRE
	Using exercises from the book to work on comprehension	REL
	Using questions to check for comprehension of a text	UPC
	Summarizing a text	RTEX
Vocabulary activities	Teaching the meaning of words	ESP
	Using the dictionary	UD

Two cameras and their stands were used to record each teacher. One camera was set up at the back of the classroom, with a full view of the space, to record all instances of teacher-student and student-student interaction. Another camera was placed near the teacher's desk, to record teacher-student interaction and offer a more detailed observation of the teaching. A total of 10 hours of recordings were made for each teacher (1 hour a day, twice a week) in December 2011 and January 2012. Overall, 42 sessions were used in this study.

Over the course of this process, two observers received four training sessions in the use of Match Vision Studio Premium (Perea et al., 2006). Once the training was completed, each observer viewed the same sessions on different occasions (with 15 days in between viewings), so that both intra- and inter-rater reliability could be calculated.

Data Analysis

Data quality was analyzed with Generalizability Theory (Cronbach et al., 1972) to calculate inter- and intra-rater reliability, and the validity of the instrument used. A measurement plan was also developed to calculate the optimal number of sessions required to run the study.

For the measurement plan, the results showed that the absolute and relative generalizability measures were acceptable (at 0.970 and 0.989) at 30 sessions, and that 40 sessions were needed to reach 0.977 and 0.992, respectively. In this sense, a total of 42 sessions were used to have the same number of sessions per teacher.

Regarding inter- and intra-observer reliability, a four-faceted SRC/O (Session, Criterion, Category/Observer) design was used, and analysis showed the greatest percentage of variability to be related to the Criterion facet (33%), while the Observer facet showed no variability at all. The absolute generalizability

coefficient was 0.999, and the relative coefficient was also 0.999, showing a high inter-rater reliability.

With respect to the intra-rater reliability, using a four-faceted SRC/M (Session, Criterion, Category/Moment) design, analysis showed that 32% of variability corresponded to the Session facet and 33% corresponded to Criterion, while Moment showed no variability. The absolute and relative generalizability coefficients obtained for Observer 1 were both 0.999. The absolute and relative coefficients for Observer 2 were both 0.997, showing high intra-rater reliability too.

Analyses of validity showed low measures of both absolute (0.000) and relative (0.000) generalizability, which is a clear sign that the test meets specificity criteria.

Next, we analyzed the frequency and duration of the behaviors exhibited by the teachers participating in the study. To determine whether their practices were in line with what the research recommends, we analyzed the frequency and duration of each teacher's use of dimensions mentioned above.

Finally, the T-patterns were analyzed to study the instructional sequence for each teacher. T-pattern detection is used to identify hidden patterns within sequential datasets (Magnusson, 1996, 2000, 2005; Magnusson et al., 2015), and in several fields (Brill et al., 2015; Burgoon et al., 2015; Castañer et al., 2015). A temporal pattern (T-pattern) is essentially a combination of events that occur in the same order with temporal distances between each other that remain relatively invariant in relation to the null hypothesis that each component is independent and is randomly distributed over time. The basic premise here is that the interactive flow or chain of behavior is governed by structures of variable stability that can be visualized by detecting these underlying T-patterns. We considered patterns that had a minimum occurrence of 7 and $p < 0.05$.

RESULTS

The results showed that the practice used most was feedback, followed by the use of resources, fluency activities, and comprehension previous knowledge activities. Used to a lesser extent were reading-writing activities, (tangible or verbal) reinforcement of correct performance of exercises and reading, alphabetic knowledge activities such as: teaching sounds, letter names, and rules using a visual aid (see **Table 2**).

We also saw that none of the teachers used practices based on the recommendations of the National Reading Panel (2000) more than 50% of the time. Measured in terms of instruction time, all the teachers spent <50% of their time teaching the five essential components of reading, with the exception of Teacher E, who spent 70.46% of class time teaching these components. The data also showed that the most common practice was fluency (see **Figure 1**), followed by literal or inference comprehension activities (see **Figure 2**), and comprehension previous knowledge activities (see **Figure 3**). We also found that teaching alphabetic knowledge activities (see **Figure 4**), phonological awareness activities (see **Figure 5**) and vocabulary activities (see **Figure 6**) were the components addressed the least in class.

To observe whether these practices formed part of a work routine, we analyzed the T-patterns of the six teachers. The results showed that Teacher A was constantly working with comprehension and vocabulary, but that his activities focused exclusively on asking questions and teaching the meaning of words. Teacher B's work routine was based on using activities for developing the five essential components of reading. Thus, in this classroom we observed instruction based on teaching fluency through activities such as rapid, accurate, and precise reading and individual/group reading, as well as joint comprehension and vocabulary work in the form of activities such as relating illustrations to text, doing exercises from the book, asking questions, and studying the meanings of words. Also typical for this teacher's work was running many different activities for teaching phonological awareness and alphabetic knowledge. Teacher C worked first on alphabetic knowledge and then on phonological awareness. No other best practices were observed in that teacher's classroom. Teacher D taught comprehension and vocabulary, but did not demonstrate any appropriate practices related to developing fluency, phonological awareness or alphabetic knowledge. Teacher E's work routine was focused on activities involving rapid, fluent, and accurate reading as well as individual/group and silent reading. There were no other best practices identified in this teacher's sequence of instruction (see **Figure 7**). The selection of this pattern is due to the fact that it is the teacher who uses most of his time to teach evidence-based components. If we analyze the results, we observe a stable T-pattern over time. The T-Patterns are plotted as dendrograms, the interpretation is performed from top to bottom, and beginning with the most elementary levels of the dendrogram. The T-pattern that is analyzed, consists of two dendrograms. The first one indicates that Teacher E works comprehension by asking the children to relate their experiences (LPNEV) and reviewing the contents worked in the classroom in relation to reading (REPA). Later, he uses positive reinforcement (P_VM) and negative reinforcement (NE_VE). Also, as for the feedback, he corrects the student when he is wrong (CLPDS), he indicates where the error is when reading (SDSE), provides examples (PEJP) and rejects when he is wrong (AA_NO). The second dendrogram indicates that the teacher works fluency through the individual reading aloud and fast (LGVAR), group and silent (LGSIL), and fast reading (LR). In relation to reading and writing, this teacher firstly asks the children to read and then write the word (LPLE) or phrases (LFLE) and vice versa (EPL-EFL). He also uses activities such as dictation (DICPF), copying (CLPF) complete words or phrases (CA). In addition, he instructs with activities that develop psychomotricity, such as orientation in space (OE_AB) or time (OT_AM), rhythm with rhythmic sequences (R_SR) and body schema (EC_CP). Teacher F used practices based on teaching phonological awareness and alphabetic knowledge activities such as: saying words that start with a given sound, dividing words up into syllables, rhyming, and teaching rules using aids. This teacher also worked on fluency activities, asking the children to read out loud using different combinations as well as quickly, accurately, and precisely. No activities aimed at developing vocabulary and comprehensions were observed in this classroom.

TABLE 2 | Data on the frequency and duration of teacher's reading practices.

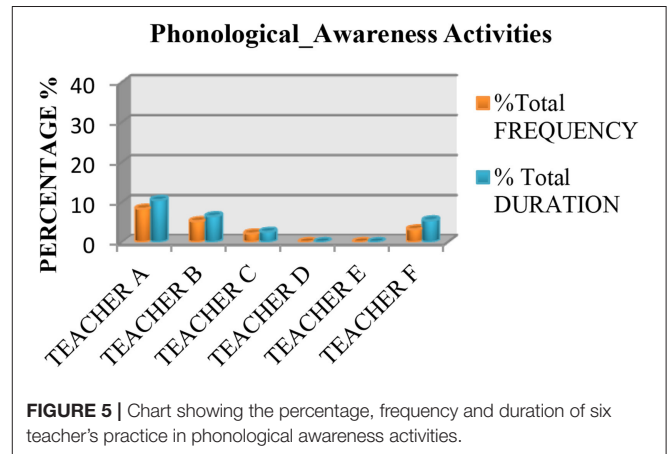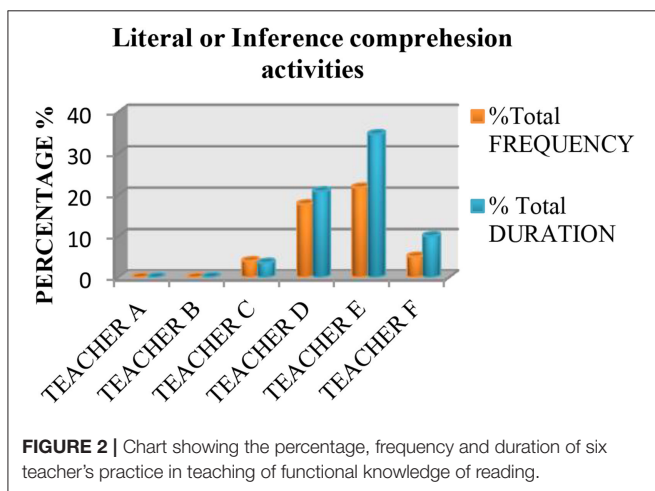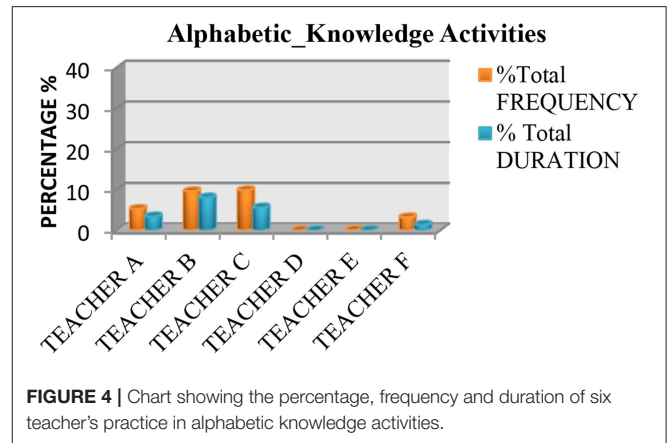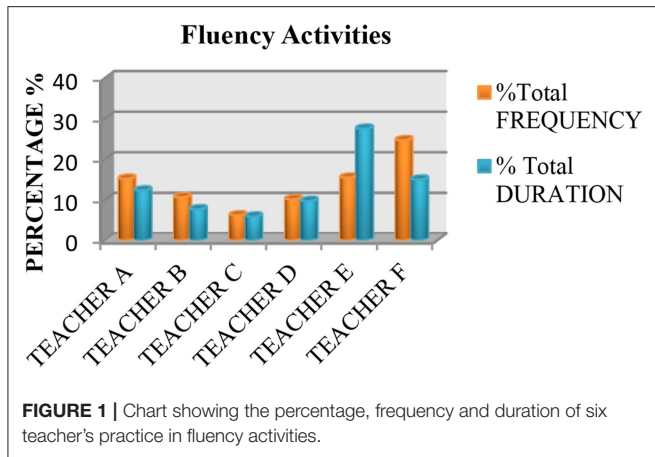| PRACTICES		TEACHERS					
		A (%)	B (%)	C (%)	D (%)	E (%)	F (%)
Alphabetic knowledge activities	FR	5.27	9.62	9.87	0	0.08	3.21
	DT	3.43	8.08	5.61	0	0.05	1.39
Phonological awareness activities	FR	8.37	5.24	2.16	0	0	3.21
	DT	10.53	6.48	2.57	0	0	5.44
Use of resources	FR	10.66	6.82	8.83	20.05	8.79	4.59
	DT	15.60	5.09	18.34	18.46	8.13	3.11
Comprehension previous knowledge activities	FR	6.19	2.15	1.12	4.36	7.43	1.15
	DT	10.58	1.30	10.96	0.81	4.15	2.76
Reinforcement	FR	7.45	13.74	10.56	11.34	3.41	10.55
	DT	3.13	6.14	3.77	2.20	0.84	2.42
Feedback	FR	29.24	32.66	50.82	22.96	37.57	35.09
	DT	23.50	47.60	42.63	29.53	13.41	29.36
Modeling	FR	0	0.71	0.08	2.32	0	0.69
	DT	0	1.97	0.43	4.13	0	0.97
Fluency Activities	FR	15.25	10.69	6.23	10.17	15.54	24.77
	DT	12.36	7.71	5.93	9.76	27.58	15.04
Guided oral instruction	FR	0.80	0.21	0.08	1.45	0.68	0
	DT	3.37	0.41	0.40	1.96	5.46	0
Homework	FR	0	2.30	0.08	0.58	0.08	2.06
	DT	0	2.44	0.11	0.08	0.08	1.23
Reading and writing activities	FR	8.37	13.57	5.37	4.94	0.25	6.19
	DT	6.61	7.27	5.33	6.93	0.42	25.46
Psychomotricity activities	FR	2.17	0.57	0.60	0.29	0.34	1.37
	DT	3.44	2.30	0.16	0.95	0.55	1.46
Literal or inference comprehension activities	FR	3.78	0.007	3.98	17.73	21.77	5.04
	DT	2.91	0.08	3.56	20.83	34.63	10
Vocabulary activities	FR	2.40	1.72	0	3.77	4.01	2.06
	DT	4.49	3.04	0	4.34	4.05	1.34
Total	FR	99.95	99.91	99.78	99.96	99.95	99.98
	DT	99.95	99.95	99.80	99.98	99.35	99.98

FR, frequency; DT, duration.

We can conclude that no teacher followed a sequence of instruction that was based on teaching all of the components recommended in the scientific literature. Three teachers did not consistently work on vocabulary or comprehension. Three did not include activities for working on phonological awareness or alphabetic knowledge in their practice. We also found that some of the activities run for some of the components were insufficient. For instance, there was no use in certain cases of practices involving isolation, identification, or deletion, and in some cases the teachers even confused phonological awareness with alphabetic knowledge: our observations included situations where the teachers were teaching this skill with alphabet cards hanging on the wall for all the pupils to see, even though they were only meant to be teaching the sounds. For alphabetic knowledge, the activities focused on teaching the name of the letter, the rule, and rhymes.

DISCUSSION

The case studies presented here through observational methodology have allowed us to analyze if the reading teaching practices used by the teachers in the classroom context are evidence based. That is, we have tried to investigate if these practices promote the skills prescribed by the NRP (i.e., alphabetic knowledge, vocabulary, fluency, comprehension, phonological awareness).

Our findings showed that none of our teachers used practices based on the recommendations of the National Reading Panel (2000) more than 50% of the time. What is more, the T-pattern analysis showed that no teacher studied, had an instruction sequence that was based on some of the key components. The practice that was used the most was feedback, followed by the use of resources, fluency activities, and previous knowledge comprehension activities. To a lesser extent, we saw the use
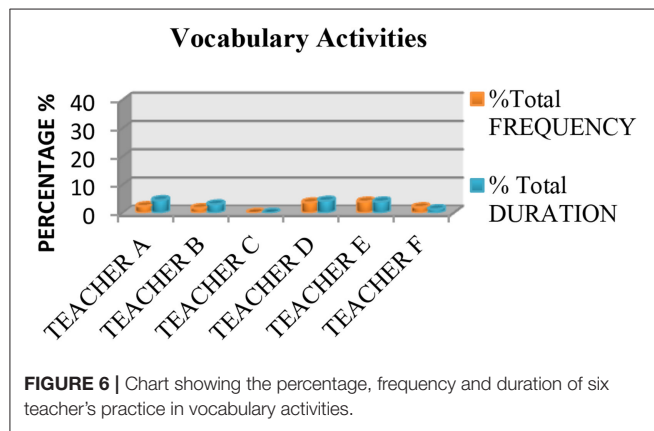
FIGURE 1 | Chart showing the percentage, frequency and duration of six teacher's practice in fluency activities.

FIGURE 4 | Chart showing the percentage, frequency and duration of six teacher's practice in alphabetic knowledge activities.

FIGURE 2 | Chart showing the percentage, frequency and duration of six teacher's practice in teaching of functional knowledge of reading.

FIGURE 5 | Chart showing the percentage, frequency and duration of six teacher's practice in phonological awareness activities.

FIGURE 3 | Chart showing the percentage, frequency and duration of six teacher's practice in comprehension activities.

which Barragán and Medina (2008) observed practices in six preschool classrooms, the results showed that practices differ as a function of how the classroom is organized and what material is available. Also, Ríos et al. (2010), working with two third and fourth grade teachers, identified two profiles of practice types: situational (e.g., working on the basis of situations that arise in the classroom, using newspapers, letters, etc.) and instructional (e.g., teaching letter names, linking letters with sounds). Also worth mentioning is the work by Fons-Esteve and Buisán-Serradell (2012), who used natural observation and systematic recording to analyze the practices of 71 preschool and elementary school teachers. Their results showed that 39% of these teachers used instructional practices, 18% used multidimensional practices, and 14% used situational practices. Looking at all this research, we see that the main strategies analyzed focused on the instructional characteristics, and classified practices as instructional, situational or multidimensional and in terms of the available resources.
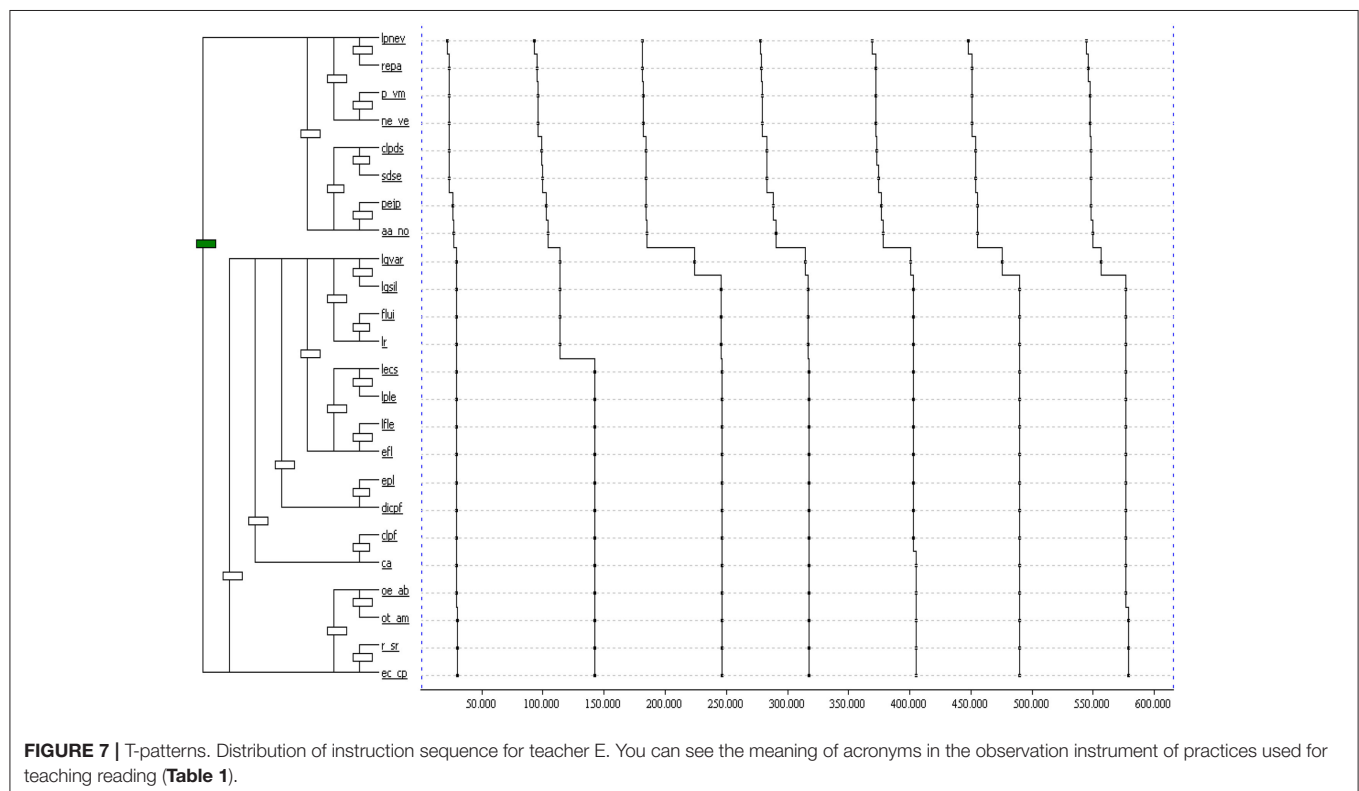
However, a common denominator that we observed in the abovementioned studies and which we present here was a far cry from the activities recommended by the National Reading Panel (2000), which insists, for instance, on the need to teach alphabetic knowledge through methods that are synthetic

of reading-writing activities, reinforcement aimed at providing (tangible or verbal) praise, reading and writing, and alphabetic knowledge activities. In one of the few studies conducted in this field, Tolchinsky and y Ríos (2009) found that teachers used explicit, early, systematic teaching. In another study, in

(converting letters to sounds, mixing sounds to form words), analytic (identifying words and their sounds), spelling-based (transforming sounds into letters), contextual (using sound-letter correspondence and finding unknown words in a text), and analogical (using parts of written words to find new ones). With respect to teaching vocabulary, we only observed practices related to teaching the meaning of words and using the dictionary. It has been recommended that when teaching this component, new technologies should also be used (Ito, 2009; Smeets et al., 2014; Bus et al., 2015), as well as the indirect method and repeated exposure to words and their meaning (Daniels, 1994, 1996; Dole et al., 1995); also, this component should be taught early on to promote reading success and comprehension (Joshi, 2005).

FIGURE 6 | Chart showing the percentage, frequency and duration of six teacher's practice in vocabulary activities.

With respect to comprehension, only three teachers carried out activities of this type, such as linking an illustration with a text, asking questions or summarizing. These exercises should be complemented with monitoring comprehension, cooperative learning, the use of graphic and semantic organizers and recognizing story structure, all of which are activities that have been shown to predict reading success (National Reading Panel, 2000). These results are in line with those obtained in other studies (Moats and Foorman, 2003; Foorman and Moats, 2004; Moats, 2009), where it was found that teachers were not using evidence-based practices.

One alternative would be to promote professional development among teachers to help them keep their knowledge up-to-date. We are aware that participation rates in this type of training are low, as evidenced by the data obtained through the Progress in International Reading Literacy Study (PIRLS) (Mullis et al., 2007); which showed that teachers in Spain receive less training in teaching reading than their counterparts in Bulgaria or Lithuania. The fact is that teacher quality predicts pupils' academic success (European Commission, 2008). Teachers should therefore be given the tools they need to teach properly, using research-based practices. Training programs should therefore address both the fundamentals of theory and educational research on the development and structure of language and reading; offer effective strategies and materials for teaching reading and writing; teach techniques for evaluating a pupil's reading performance as measured by the different components; expose teachers to new technologies; and help teachers strike the right balance between theory and

FIGURE 7 | T-patterns. Distribution of instruction sequence for teacher E. You can see the meaning of acronyms in the observation instrument of practices used for teaching reading (Table 1).

practice (IRA, 2007). A clear example of this can be found in the DIPELEC (*Diploma de Especialización en Enseñanza de la Lectura*), the first postgraduate diploma in reading instruction to be offered in Spain (http://fg.ull.es/grados-posgrados/estudios/diploma-de-especializacion-en-ensenanza-de-la-lectura/). The difficulty now lies in convincing teachers of the need to obtain up-to-date training and change their consolidated teaching practices. Including best practices in legislation and offering compensation to teachers might serve as a good start. A limitation of this study was not analyze how these practices could influence reading performance amongst schoolchildren. Future lines of research should explore this aspect.

AUTHOR CONTRIBUTIONS

NS: This author's grant was used to run the project *Integrando creencias y prácticas de enseñanza de la lectura* (Integrating beliefs and practices about teaching reading), ref: PSI2009-11662. She participated actively in the observation of the teachers, carried out the analyses of the teaching practices, and was responsible for the literature review and the drafting of this manuscript. CS: Supervised the design and preparation of the study, was responsible for handling and analyzing the data, offered guidance on methodology, and helped review the manuscript. JJ: As principal investigator, supervised the project and the preparation of the study, offered guidance for the theoretical component, and was responsible for reviewing this manuscript. MTA: Carried out the analyses of the teaching practices using T-patterns, offered guidance on methodology, and helped review this manuscript. All authors approved the final version of this article.

REFERENCES

Adams, M. J. (2001). "Alphabetic anxiety and explicit systematic phonics instruction: a cognitive science perspective," in *Handbook of Early Literacy Research*, Vol. 1, eds S. B. Neuman and D. K. Dickinson (New York, NY: Guilford), 66–80.

Al Otaiba, S., and Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: a review of the literature. *Remed. Spec. Educ.* 23:300. doi: 10.1177/07419325020230050501

Allington, R. (1983). Content coverage and contextual reading in reading groups. *J. Read. Beh.* 16, 85–96.

Allington, R. L. (2000). *What Really Matters for Struggling Readers: Designing Research-Based Programs*. Boston, MA: Longman.

Anguera, M. T. (1979). Observational typology. *Qual. Quant.* 13, 449–484.

Anguera, M. T. (2003). "Observational Methods (General)", in *Encyclopedia of Psychological Assessment,* Vol. 2, ed R. Fernández-Ballesteros (London: Sage), 632–637.

Armbruster, B. B., Lehr, F., and Osborn, J. (2001). *Put Reading First: The Research Building Blocks for Teaching Children to Read*. Washington, DC: National Institute for Literacy.

Barragán, C., and Medina, M. M. (2008). Las prácticas de la lectura y escritura en Educación Infantil [The practices of reading and writing in Early Childhood Education]. *Rev. Educ.* 10, 149–165.

Blanco-Villaseñor, A., Losada, J. L., and Anguera, M. T. (2003). Analytic techniques in observational designs in environment-behavior relation. *Med. Amb. Comport. Hum.* 4, 111–126.

Brill, M., Jonsson, G. K., Magnusson, M. S., and Schwab, F. (2015). "Immersive dynamics: presence experiences and patterns of attention," in *Discovering Hidden Temporal Patterns in Behavior and Interactions: T-Pattern Detection and Analysis with THEME*, eds M. S. Magnusson, J. K. Burgoon, M. Casarrubea, and D. McNeill (New York, NY: Springer), 183–194.

Burgoon, J. K., Wilson, D., Hass, M., and Schuetzlr, R. (2015). "Interactive deception in group decision-making: new insights from communication pattern analysis," in *Discovering Hidden Temporal Patterns in Behavior and Interactions: T-Pattern Detection and Analysis with THEME*, eds M. S. Magnusson, J. K. Burgoon, M. Casarrubea, and D. McNeill (New York, NY: Springer), 37–61.

Bus, A. G., Takacs, Z. K., and Kegel, C. A. (2015). Affordances and limitations of electronic storybooks for young children's emergent literacy. *Dev. Rev.* 35, 79–97. doi: 10.1016/j.dr.2014.12.004

Caravolas, M., Kessler, B., Hulme, C., and Snowling, M. (2005). Effects of orthographic consistency, frequency, and letter knowledge on children's vowel spelling development. *J. Exp. Child Psychol.* 92, 307–321. doi: 10.1016/j.jecp.2005.08.001

Carnine, D., and Kameenui, E. (1992). *Higher Order Thinking: Designing Curriculum for Mainstreamed Students*. Austin, TX: Pro-Ed.

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Canton, A., and Hileno, R. (2016). Goal scoring in soccer: a polar coordinate analysis of motor skills used by Lionel Messi. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Castañer, M., Camerino, O., Anguera, M. T., and Jonsson, G. (2015). "Paraverbal communicative teaching T-patterns using SOCIN and SOPROX observational systems," in *Discovering Hidden Temporal Patterns in Behavior and Interactions: T-Pattern Detection and Analysis with THEME*, eds M. S. Magnusson, J. K. Burgoon, M. Casarrubea, and D. McNeill (New York, NY: Springer), 83–100.

Castañer, M., Camerino, O., Anguera, M. T., and Jonsson, G. K. (2013). Kinesics and proxemics communication of expert and novice PE teachers. *Qual. Quant.* 47, 1813–1829. doi: 10.1007/s11135-011-9628-5

Chard, D. J., and Jungjohann, K. (2006). *Scaffolding Instruction for Success in Mathematics Learning, Intersection: Mathematics Education Sharing Common Grounds*. Houston TX: Exxon-Mobil Foundation.

Coyne, M. D., Kame'enui, E. J., and Carnine, D. W. (2007). *Effective Teaching Strategies That Accommodate Diverse Learners, 3rd Edn*. Upper Saddle River, NJ: Pearson: Merrill Prentice Hall.

Cronbach, L. J., Gleser, G. C., Nanda, H., and Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York, NY: Wiley.

Cunningham, A. E., Zibulsky, J., Stanovich, K. E., and Stanovich, P. J. (2009). How teachers would spend their time teaching language arts: the mismatch between self-reported and best practices. *J. Learn. Dis*. 42, 418–430. doi: 10.1177/0022219409339063

Daniels, M. (1994). The effect of sign language on hearing children's language development. *Comm. Educ*. 43, 291–298.

Daniels, M. (1996). Bilingual, bimodal education for hearing kindergarten students. *Sign. Lang. Stud*. 90, 25–37. doi: 10.1353/sls.1996.0003

De Arcangelo, M. (2003). On the mind of a child: a conversation with Sally Shaywitz. *Educ. Lead*. 60, 6–10.

Del Campo, R., Buchanan, W. R., Abbott, R. D., and Berninger, V. W. (2015). Levels of phonology related to reading and writing in middle childhood. *Read. Writ*. 28, 183–198. doi: 10.1007/s11145-014-9520-5

Dickinson, D., and Tabors, P. (2001). *Young Children Learning at Home and School. Beginning Literacy with Language*. Baltimore: Paul H. Brookes.

Dole, J. A., Sloan, C., and Trathen, W. (1995). Teaching vocabulary within the context of literature. *J. Read*. 38, 452–460.

Durkin, D. (1993). *Teaching them to Read*. Des Moines: Allyn and Bacon.

EACEA/Eurydice (2011). *Teaching Reading in Europe: Contexts, Policies and Practices*. Brussels: Eurydice.

Ehri, L., Nunes, S., Willows, D., Valeska, B., Yaghoub, Z., and Yaghoub, T. (2001). Phonemic awareness instruction helps children learn to read: evidence from the national reading panel's meta-analysis. *Read. Res. Quart*. 36, 250–287. doi: 10.1598/RRQ.36.3.2

European Commission (2008). *Conclusions of the Presidency, Brussels European Council*. Available online at: http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/ec/99410.pdf

Fetsco, T., and McClure, J. (2005). *Educational Psychology; An Integrated Approach to Classroom Decisions*. Boston, MA: Allyn and Bacon.

Fletcher-Flinn, C. (2014). Learning to read as the formation of a dynamic system: evidence for dynamic stability in phonological recoding. *Front. Psych*. 5:660. doi: 10.3389/fpsyg.2014.00660

Fons, M. (2008). Despacito y buena letra [Slow and good letter]. *Inf. Aprend*. 45, 38–41.

Fons, M. (2012). Nacidos para leer… ¿Por qué? [Born to read…Why?] 0 a 6. *Aula de Inf*. 67, 8–12.

Fons-Esteve, M., and Buisán-Serradell, C. (2012). Entrevistas y observaciones de aula: otra mirada a los perfiles de prácticas docentes. [Interviews and classroom observations: another look at the profiles of teaching practices]. *Cult. Educ*. 24, 401–413. doi: 10.1174/113564012803998776

Foorman, B. R., and Moats, L. C. (2004). Conditions for sustaining research-based practices in early reading instruction. *Remed. Spec. Educ*. 25, 51–60. doi: 10.1177/07419325040250010601

Foorman, B. R., and Torgesen, J. (2001) Critical elements of classroom and small-group instruction promote reading success in all children. *Lear. Disabil. Res. Pract*. 16, 203–212. doi: 10.1111/0938-8982.00020

Foorman, B., Chen, D., Carlson, C., Moats, L., Francis, D., and Fletcher, J. M. (2003). The necessity of the alphabetic principle to phonemic awareness instruction. *Read. Writ. Int. J*. 16, 289–324.

Greenhoot, A. F., Beyer, A. M., and Curtis, J. (2014). More than pretty pictures? How illustrations affect parent-child story reading and children's story recall. *Front. Psych*. 5:738. doi: 10.3389/fpsyg.2014.00738

Hatcher, P. J., Hulme, C., Miles, J. N., Carroll, J. M., Hatcher, J., Gibbs, S., et al. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: a randomised controlled trial. *J. Child Psych. Psychiatry* 47, 820–827. doi: 10.1111/j.1469-7610.2005.01559.x

Hirsch, J. (2007). La comprensión lectora requiere conocimiento de vocabulario y del mundo: hallazgos científicos sobre el bajón de cuarto grado y el estancamiento en los puntajes nacionales de comprensión [Reading comprehension requires knowledge of vocabulary and the world: scientific findings on the fourth-degree downturn and stagnation in national comprehension scores]. *Estud. Públic*. 108, 229–252.

Hirst, D. (2007). *Method for Responding to Reading Comprehension Questions*. U.S Patent Application No. 11/462,265. Available online at: https://docs.google.com/viewer?url=patentimages.storage.googleapis.com/pdfs/US20070031803.pdf

IRA (2007). *Teaching Reading Well: A Synthesis of the International Reading Association's Research on Teacher Preparation for Reading Instruction*. Newark, DE: IRA.

Ito, M. (2009). *Engineering Play. A Cultural History of Children's Software*. Cambridge, MA: The MIT Press.

Jiménez, J. E., Crespo, P., and Suárez, N. (2012). "Perspectivas futuras en la identificación temprana e intervención de la dislexia" [Future prospects in the early identification of and intervention in dyslexia], in *Dislexia en Español, Prevalencia e Indicadores Cognitivos, Culturales, Familiares y Biológicos*, ed J. E. Jiménez (Madrid: Pirámide), 251–272.

Jiménez, J. E., Rodríguez, C., Suárez, N., and O'Shanahan, I. (2014). ¿Coinciden nuestras ideas con lo que dicen las teorías científicas sobre el aprendizaje de la lectura? [Do our ideas coincide with what the scientific theories say about learning to read?] *Rev. Esp. Ped*. 259, 395–412.

Joshi, R. (2005). Vocabulary: a critical component of comprehension. *Read. Writ. Q*. 21, 209–219. doi: 10.1080/10573560590949278

Joshi, R. M., Binks, E., Hougen, M., Dahlgren, M., Oker-Dean, F., and Smith, D. L. (2009). Why elementary teachers might be inadequately prepared to teach reading? *J. Learn. Dis*. 42, 444–457. doi: 10.1177/0022219409338736

Kaiser, L., Rosenfield, S., and Gravois, T. (2009). Teachers' perception of satisfaction, skill development, and skill application after instructional consultation services. *J. Learn. Dis*. 42, 444–457. doi: 10.1177/0022219409339062

Kjeldsen, A., Kärnä, A., Niemi, P., Olofsson, Å., and Witting, K. (2014). Gains from training in phonological awareness in kindergarten predict reading comprehension in grade 9. *Sci. Stud. Read*. 18, 452–467. doi: 10.1080/10888438.2014.940080

Lane, H. (2014). *Evidence-Based Reading Instruction for Grades K-5 (Document No. IC-12)*. Retrieved from University of Florida, Collaboration for Effective Educator, Development, Accountability, and Reform Center website: Avalaible online at: http://ceedar.education.ufl.edu/tools/innovation-configurations/

Magnusson, M. S. (1988). "Le temps et les patterns syntaxiques du comportement humain: modèle, méthode et le programme THEME," in *Revue des Conditions de Travail*. Actes du *Premier Colloque National D'ergonomie Scolaire* (Marseille, Octares: Université de Lille), 284–314.

Magnusson, M. S. (1996). Hidden real-time patterns in intra- and inter-individual behavior. *Eur. J. Psych. Assess*. 12, 112–123. doi: 10.1027/1015-5759.12.2.112

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behav. Res. Meth. Instr. Comp*. 32, 93–110. doi: 10.3758/BF03200792

Magnusson, M. S. (2005). "Understanding social interaction: discovering hidden structure with model and algorithms," in *The Hidden Structure of Interactions: From Neurons to Culture Patterns*, eds L. Anolli, S. Duncan, M. S. Magnusson, and G. Riva (Amsterdam: IOS Press), 4–20.

Magnusson, M. S., Burgoon, J. K., and Casarrubea, M. (Eds.). (2015). *Discovering Hidden Temporal Patterns in Behavior and Interaction. T-Pattern detection and analysis with THEME*. New York, NY: Springer.

Moats, L. (2009). Knowledge foundations for teaching reading and spelling. *Read. Writ*. 22, 379–399.

Moats, L. C., and Foorman, B. R. (2003). Measuring teachers' content knowledge of language and reading. *Ann. Dys*. 53, 23–45.

Mullis, I. V., Martin, M. O., Kennedy, A. M., and Foy, P. (2007). *PIRLS 2006, International Report in Reading*. Boston, MA: IEA. Available online at: https://timssandpirls.bc.edu/pirls2006/intl_rpt.html

Mullis, I. V., Martin, M. O., Kennedy, A. M., Foy, P. and Drucker, K. T. (2012). *PIRLS 2011, International Results in Reading*. Boston, MA: IEA. Available online at: https://timssandpirls.bc.edu/pirls2011/international-results-pirls.html

National Assessment of Educational Progress (2015). *The Nation's Report Card: 2015 Mathematics and Reading Grade 4-12 Assessments*. Department of education. Available online at: http://www.nationsreportcard.gov/reading_math_2015/#reading?grade=4

National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel, Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Instruction*. Washington, DC: Government Printing Office. Available online at: https://www1.nichd.nih.gov/publications/pubs/nrp/Pages/report.aspx

National Reading Panel (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction: Reports of the Subgroups*. Bethesda, MD: National Institute of Child Health and Human Development. Available online at: https://www.nichd.nih.gov/about/org/der/branches/cdbb/nationalreadingpanelpubs

Nichols, W., Rupley, W., and Rasinski, T. (2008). Fluency in learning to read for meaning: going beyond repeated readings. *Lit. Res. Instruc.* 48, 1–13. doi: 10.1080/19388070802161906

Organization for Economic Co-operation and Development (2012). *PISA 2012. Programa para la Evaluación de Alumnos 2012. Informe español. Volumen I: Informe de Resultados y Contexto. [Program for the Evaluation of Students 2012. Spanish Report. Volume I: Report of results and context]*. Madrid: Ministerio de Educación, Cultura y Deporte.

Organization for Economic Co-operation and Development (2015). PISA 2015. *Programa para la Evaluación de Alumnos 2015. Resultados Clave. [Program for the Evaluation of Students 2015. Key Results]*. Madrid: Ministerio de Educación, Cultura y Deporte. Available online at: https://www.oecd.org/pisa/pisa-2015-results-in-focus-ESP.pdf

Pascual, R., Madrid, D., and Mayorga, M. J. (2013). Aprendizaje precoz de la lectura: reflexiones teóricas y desde la experiencia en el aula. [Early learning of reading: theoretical reflections and from the experience in the classroom]. *Ocnos* 10, 91–106. doi: 10.18239/ocnos

Pearson, P. D. (2001). "Learning to teach reading: The status of the knowledge base", in *Learning to Teach Reading: Setting the Research Agenda*, ed C. M. Roller (Newark, DE: International Reading Association), 4–19.

Perea, A., Castellano, J., and Alday, N. (2006). *MatchVision Studio Premium V.3.* Vitoria: Universidad del País Vasco.

Perfetti, C. A., Landi, N., and Oakhill, J. (2005). "The acquisition of reading comprehension skill", in *The Science of Reading: A Handbook*, eds N. J. Snowing and C. Hulme (Oxford, USA: Blackwell Publishing), 227–247.

Pinker, S. (2001). "Language acquisition," in *An Invitation to Cognitive Science: Language*, eds L. Gleitman, D. Osherson, and M. Liberman (Cambridge: MIT Press), 1–41.

Podhajski, B., Mather, N., Nathan, J., and Sammons, J. (2009). Professional development in scientifically based reading instruction teacher knowledge and reading outcomes. *J. Learn. Disabil.* 42, 403–417. doi: 10.1177/0022219409338737

Porta, E. (2008). *Hacia un Enfoque Comprehensivo del Aprendizaje Lingüístico Inicial. Implementación del Modelo y Evaluación de sus efectos sobre el Rendimiento lector.[Towards a Comprehensive Approach to Initial Language Learning. Implementation of the Model and Evaluation of its Effects on Reader Performance]*. Tesis doctoral. Facultad de Filosofía y Letras, Universidad Nacional de Cuyo. Available online at: http://ffyl.uncu.edu.ar/?id_rubrique=417

Porta, M. E., Harper, L., and Kraft, R. (2010). Hemispheric asymmetry profiles during beginning reading. Effects of reading level and word type. *Dev. Neuropsych.* 35, 96–114. doi: 10.1080/87565640903325717

Porta, M., and Ison, M. (2011). Hacia un enfoque comprensivo del aprendizaje lingüístico inicial como proceso cognitivo. [Towards a comprehensive approach to early language learning as a cognitive process]. *Rev. Iberoam. Educ.* 55, 243–260.

Portell, M., Anguera, M. T., Chacón, S., and Sanduvete, S. (2015). Guidelines for reporting evaluations based on observational methodology (GREOM). *Psicothema* 27, 283–289. doi: 10.7334/psicothema2014.276

Pufpaff, L. A. (2009). A developmental continuum of phonological sensitivity skills. *Psychol. Schools* 46, 679–691. doi: 10.1002/pits.20407

Purcell-Gates, V., Degener, S. C., Jacobson, E., and Soler, M. (2002). Impact of authentic adult literacy instruction on adult literacy practices. *Read. Res. Quart.* 37, 70–92. doi: 10.1598/RRQ.37.1.3

Rasinski, T. V. (2003). *The Fluent Reader: Oral Reading Strategies for Building Word Recognition, Fluency, and Comprehension.* Scholastic.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., and Seidenberg, M. S. (2002). How should reading be taught? *Sci. Amer. Ed.* 286, 70–77.

Ríos, I., Fernández, P., and Gallardo, I. (2010). *La contribución de las prácticas de aula. A los logros de aprendizaje. [The contribution of classroom practices. To learning achievements]*. II Congrés Internacional de didactiques, 322–329. Available online at: http://hdl.handle.net/10503/12311

Sánchez-Algarra, P., and Anguera, M. T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: impact of recording and coding predominating perspectives. *Qual. Quant.* 47, 1237–1257. doi: 10.1007/s11135-012-9764-6

SAS Institute Inc (1999). *SAS/STAT User's Guide, Version 8.* Cary, NC: SAS Institute Inc.

Scarborough, H. (2002). "Connecting early language and literacy to later reading (dis)abilities: Evidence, theory and practice", in *Handbook of Early Literacy Research*, eds S. Neuman, and D. Dickinson (London: The Guilford Press), 97–110.

Schwanenflugel, P. J., and Benjamin, R. G. (2012). "Reading expressiveness: The neglected aspect of reading fluency," in *Fluency Instruction: Research-Based Best Practices*, 2nd Edn, eds T. Rasinski, C. Blachowicz, and K. Lems (New York, NY: Guilford), 35–54.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., et al. (2010). *Improving reading comprehension in kindergarten through 3rd grade: IES practice guide (NCEE 2010-4038)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Available online at: whatworks.ed.gov/publications/practiceguides

Slavin, R. E. (2003). *Educational Psychology. Theory and Practice.* Boston: Allyn and Bacon.

Smeets, D. J., van Dijken, M. J., and Bus, A. G. (2014). Using electronic storybooks to support word learning in children with severe language impairments. *J. Learn. Dis.* 47, 435–449. doi: 10.1177/0022219412467069

Stevenson, R. B. (2004). Constructing knowledge of educational practices from case studies. *Environ. Educ. Res.* 10, 39–51. doi: 10.1080/13504620320001 73698

Strasser, K., del Río, F., and Larraín, A. (2013). Profundidad y amplitud del vocabulario: ¿Cuál es su rol en la comprensión de historias en la edad pre-escolar? Vocabulary depth and breadth: Their role in preschoolers' story comprehension. *Est. Psicol.* 34, 221–225. doi: 10.1174/021093913806751401

Suárez, P. M., García, M., and Cuetos, F. (2013). Variables predictoras de la lectura y la escritura en castellano. [Predictors of reading and writing in Spanish]. *Infan. Aprend.* 36, 77–89. doi: 10.1174/021037013804826537

Temple, C., Ogle, D., Crawford, A., and Freppon, P. (2011). *All Children Read: Teaching for Literacy in Today's Diverse Classrooms, 3rd Edn.* Boston: Pearson.

Tolchinsky, L., and y Ríos, I. (2009). ¿Qué dicen los maestros que hacen para enseñar a leer y a escribir? [What do teachers say to teach reading and writing?] *Aula de Innov. Educat*, 174, 1–7.

Tunmer, W. E., and Arrow, A. W. (2013). "Phonics instruction," in *International Guide to Student Achievement* eds J. Hattie and M. Anderman (New York, NY: Routledge), 316–319.

Ysewijn, P. (1996). *GT Software for Generalizability Studies.* Mimeograph.

Observation of Communication by Physical Education Teachers: Detecting Patterns in Verbal Behavior

Abraham García-Fariña[1]*, F. Jiménez-Jiménez[1] and M. Teresa Anguera[2]

[1] *Department of Specific Didactics, Faculty of Education, University of La Laguna, Santa Cruz de Tenerife, Spain,* [2] *Faculty of Psychology, Institute of Neurosciences, University of Barcelona, Barcelona, Spain*

The aim of this study was to analyze the verbal behavior of primary school physical education teachers in a natural classroom setting in order to investigate patterns in social constructivist communication strategies before and after participation in a training program designed to familiarize teachers with these strategies. The participants were three experienced physical education teachers interacting separately with 65 students over a series of classes. Written informed consent was obtained from all the students' parents or legal guardians. An indirect observation tool (ADDEF) was designed specifically for the study within the theoretical framework, and consisted of a combined field format, with three dimensions, and category systems. Each dimension formed the basis for building a subsequent system of exhaustive and mutually exclusive categories. Twenty-nine sessions, grouped into two separate modules, were coded using the Atlas.ti 7 program, and a total of 1991 units (messages containing constructivist discursive strategies) were recorded. Analysis of intraobserver reliability showed almost perfect agreement. Lag sequential analysis, which is a powerful statistical technique based on the calculation of conditional and unconditional probabilities in prospective and retrospective lags, was performed in GSEQ5 software to search for verbal behavior patterns before and after the training program. At both time points, we detected a pattern formed by requests for information combined with the incorporation of students' contributions into the teachers' discourse and re-elaborations of answers. In the post-training phase, we detected new and stronger patterns in certain sessions, indicating that programs combining theoretical and practical knowledge can effectively increase teachers' repertoire of discursive strategies and ultimately promote active engagement in learning. This has important implications for the evaluation and development of teacher effectiveness in practice and formal education programs.

Keywords: communicative strategies, social constructivism, systematic observation, physical education, instructional communication

INTRODUCTION

Analysis of patterns in instructional communication allows teachers to reflect on their use of discursive strategies, check that these are aligned with their teaching goals, and resolve to incorporate them as a strategic part of their teaching.

Instructional communication patterns have been detected in the teaching of science (Cazden, 1988; Lemke, 1990) and mathematics (Lobato et al., 2005) and include the

initiation-response-evaluation (IRE) pattern and the elicitation-response-evaluation (ERE) pattern (Bowers and Nickerson, 2001). Both patterns, or sequences, begin with a question designed to actively engage the students in the construction of knowledge. Nonetheless, it has been claimed that IRE sequences can deny students and teachers the opportunity for debate and negotiation (Wright and Forrest, 2007).

Social constructivism theory (Vygotsky, 1978) attaches great importance to dialogue between the agents engaged in the teaching and learning process. The general principle underlying this theory is that students can be helped to build knowledge by stimulating their higher mental processes through language-mediated interaction with their social and cultural environments. Edwards and Mercer (1989), claim that the value of educational discourse lies above all in its potential as a tool for negotiating students' previous representations and using these as scaffolding to build new knowledge throughout teacher-student interactions. This idea that language, as a modulator of an interactive system, influences cognitive and perceptual processes has also been highlighted by Lupyan (2012).

According to Coll and Onrubia (2001), instructional communication, which they refer to as "discursive strategies," can serve three important pedagogical functions. It can (a) lead to the establishment of an initial platform for shared representations, where students' previous knowledge can be linked to the learning objective through discursive strategies involving questions or references to specific or social frameworks; (b) help students to adopt a positive attitude to learning through the use of meta-statements, incorporation of student contributions into their discourse, and characterization of knowledge as something shared; and (c) increase students' knowledge by guiding them toward increasingly complex representations. To achieve this, teachers can adopt a range of discursive strategies, such as re-elaboration of student contributions, categorization and labeling of certain aspects of content or context, abbreviation of expressions, modification of references used to talk about content, and use of recapitulation, summaries, and synthesis. By incorporating these and similar discursive strategies, which are defined by Coll and Onrubia (2001, p. 24) as a particular form of verbal communication used to guide the construction of knowledge, teachers can increase the impact and effectiveness of their instructional communication. Constructivist strategies are a valuable methodological resource, and they acquire meaning in context and at a given moment during a class. In a study on how to develop tools for an effective classroom, Powell and Kalina (2009) claimed that teachers need to use constructivist strategies and resources, such as examples linked to the topic being taught, questions to assess learning, and discussion and dialogue to recapitulate.

Several authors have analyzed social constructivism in the field of physical education through a theoretical lens. Constructivist physical educators value students' contributions, actively involve them in the construction of knowledge, and draw parallels between what is being taught and the students' personal experiences (Azzarito and Ennis, 2003). The main principles underlying the social constructivism theory (higher mental processes, language, mediation, cultural influence, and zone of proximal development) can all be applied to physical education, which involves teaching and learning about the development of motor skills and higher mental processes while enabling the exploration of concepts through action and language (Ussher and Gibbes, 2002). Authors such as Rovegno and Dolly (2006) and Ussher and Gibbes (2002) have also analyzed the constructivist perspective underlying diverse physical and sport education models, including the Teaching Games for Understanding (TGfU) and the sport education, personal and social responsibility, and adventure-based learning models. In all these models, dialogue between teachers and students regarding actions is critical.

The emergence of new sport education models centered around the intentional use of communicative strategies has had an important role in the creation of constructivist understanding (Morgan and Kingston, 2008). The TGfU model, considered by Light (2008) to be a good example of a social constructivist approach to teaching physical education, is perhaps the best-known example (Bunker and Thorpe, 1982; Kirk and MacPhail, 2002; Oslin and Mitchell, 2006). This model stresses the importance of using questions as a key communicative strategy for promoting reflection and tactical awareness among players, and accordingly, stimulates teachers' interest in the verbal behavior of students in relation to the meaning they attribute to the actions they perform (Wallian and Chang, 2007). As the TGfU model is built on problem-solving activities, high-quality questions are critical. These need to be planned and carefully constructed to ensure that they prompt critical thinking and favor the development of problem-solving skills (Dyson et al., 2004; Mitchell et al., 2006; Hubball et al., 2007). Questions addressed to the group help the students as a whole to scaffold knowledge, creating a learning environment that engages the students in the construction of knowledge (Harvey and Light, 2015) and helps them to learn to learn (Light, 2014). In teaching models that use a similar approach to the TGfU model, eliciting information from students in the form of questions is considered a key discursive strategy for building knowledge. Rink (1998) considers that "instructional strategies" used in the teaching of physical education (e.g., questions, references to existing knowledge, linking to other topics, and recapitulations) are themselves a methodological resource.

Webster (2010) proposed six skills that physical educators should master in order to improve the effectiveness of their instructional communication processes and increase student motivation. The first three are rhetorical communication skills (being clear, content relevance, and using humor) (Chesebro and Wanzer, 2006), while the second three are relational communication skills (immediacy, communication style and listening). For each of these skills, Webster proposed a series of specific instructional strategies.

Other studies in the field of physical education have analyzed the communication of content relevance. Webster et al. (2012), for example, analyzed the different ways in which teachers communicated content relevance and also the frequency with which they reported doing so according to whether they were expert or novices. Webster et al. (2011, 2013), in turn, analyzed how students perceived this communication of content relevance.

The results showed that expert teachers communicate content relevance more frequently and that this strategy appears to instill in students a desire to keep learning. Other studies have analyzed instructional communication among physical educators from the perspective of need-supportive interactions (Haerens et al., 2013). Finally, a study of middle-school students' perceptions of instructional choices by physical education teachers found that these choices appeared to satisfy autonomy needs and promote student engagement (Agbuga et al., 2016). Overall, the different studies undertaken in this area show that the communication strategies (Anguera and Izquierdo, 2006) employed by physical educators have a significant effect on different aspects of learning.

The main aim of this study was to investigate whether it was possible to detect patterns in instructional communication strategies used by primary school physical education teachers. A secondary aim was to determine whether participation in a training intervention designed to teach social constructivist communication skills led to changes in practice.

MATERIALS AND METHODS

Design

To investigate the presence of constructivist discursive strategies (Coll and Onrubia, 2001), we designed a systematic observation study (Anguera, 2003; Castañer et al., 2016, 2017; Anguera et al., 2017) based on indirect observation (Lacy and Darst, 1985; Allison, 1990; Eckrich et al., 1994; Coleman and Mitchell, 2001; Anguera et al., 2018) to analyze the verbal behavior of physical education teachers in a natural classroom setting.

The nature and requirements of the study justified the use of a Nomothetic/Follow-up/Multidimensional design, which corresponds to quadrant IV of the observational methodology designs (Blanco-Villaseñor et al., 2003; Sánchez-Algarra and Anguera, 2013). The design was: (a) nomothetic because we analyzed the instructional communication, or verbal behavior, of three physical education teachers acting individually; (b) "follow-up" because we collected data over a series of successive sessions (intersessional follow-up) and also recorded each session in full, without interruption (intrasessional follow-up); and c) "multidimensional," because although we were investigating just one overall response level or dimension (i.e., the teachers' instructional communication), the ad hoc observation instrument, which was derived from Coll and Onrubia (2001) social constructivist framework, unveiled three levels of response or dimensions (see description of observation instrument).

To investigate changes in the patterns detected following participation in a training activity focused on discursive strategies from a constructivist approach, we organized a collaborative action research program designed to familiarize physical education teachers with the use and value of these strategies as a methodological resource. Collaborative action research programs are accredited models (Carr and Kemmis, 1986; Elliott, 1991) that encourage interpretation and critical thinking to help teachers to reflect on and evaluate their practices and introduce changes that will make these more effective. The collaborative action research program designed for the present study was held over a 4-month period and was led by the first author. The program consisted of eight sessions, held every 2 weeks. It was held in the period between the teaching of the first and second modules. The participants learnt about and discussed social constructivist strategies and alternatives, and reflected on how these could improve their teaching. Because the collaborative action research program was interpreted as a training event, we use the terms "pre-training" and "post-training" in our presentation of data and results.

Participants

We analyzed three physical education teachers (1 man and 2 women) with more than 2 years' experience who taught a total of 65 students with a mean age of 10.7 years. The students were from 3 years at different schools and included 26 first-second class students, 19 fifth-class students, and 20 sixth-class students).

This study was carried out in accordance with the recommendations of Ethical Committee of the University of La Laguna (Spain) with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Observation Instrument

We used an *ad hoc* observation instrument (Anguera et al., 2007), called Analysis of Educational Discourse in Physical Education, or ADDEF as per its Spanish acronym (**Table 1**). The instrument was designed to discriminate between and record discursive strategies used by physical education teachers. It was suited to the multidimensional design of the study, and consisted of a combined field format and category system, which is the most recommendable system for studies of this type (Lacy and Darst, 1985; Castañer et al., 2013; Portell et al., 2015). We built a category system for each of the three dimensions or criteria derived from Coll and Onrubia's (2001) social constructivist theory regarding discursive strategies for the classroom: (1) exploration and activation of previous knowledge, (2) attribution of positive meaning by students to the concepts being taught, and (3) progressive establishment of increasingly expert and complex representations of the subject matter. **Table 2** shows the three category systems, which are formed, respectively by three, seven, and four exhaustive, mutually exclusive categories.

Recording Instrument

The transcripts of the teachers' lessons were coded using the qualitative analysis program Atlas.ti v. 7.1.8. (**Figure 1**) Lag sequential analysis was performed in GSEQ 5.1 (Bakeman and Quera, 2011).

Procedure

For the data collection stage, 29 sessions corresponding to two teaching modules were recorded. The first module (consisting of six sessions for teacher #1, four sessions for teacher #2, and five sessions for teacher #3) was taught before the collaborative action research program and the second module (consisting of five sessions for teacher #1, five sessions for teacher #2, and four sessions for teacher #3) was taught after the program. A total of 1,991 messages containing the discursive strategies analyzed were recorded: 719 before the program and 1,272 afterwards. All

TABLE 1 | ADDEF Observation Instrument.

Criterion 1. Exploration and activation of previous knowledge

---

Use of social framework (A1)

References to social situations/events (or their meanings) related to the subject matter or task at hand with the aim of establishing sharing meanings in relation to these situations/events.

Example: You have to jump like a frog

Use of specific framework (A2)

References to specific previously shared learning experiences, clearly highlighting their relationship with the subject matter or task at hand, seeking to establish shared meanings.

Example: At the beginning of the course we practiced moving from one point to another; today we are going to do sprints.

Request for information (A3)

Use of strategies to obtain relevant information from the students on the subject matter or task at hand, but without mention of a social or specific framework.

Example: How many different ways did they throw the ball?

Criterion 2. Attribution of positive meaning by students to the concepts being taught

---

Use of meta-statements before the task (B1)

References to what is going to be done or to what might occur, without linking these to a previous activity, and only including messages that refer to the subsequent learning activity.

Example: We are going to play the 10-pass game so that the player who is about to receive the ball in movement learns to get free.

Use of meta-statements during the task (B2)

References that remind students about the goal of the task, i.e., about what it is they are trying to improve.

Example: We are practicing our aim and learning to move the cones.

Incorporation of students' contributions into the teacher's discourse (B3)

Literal or near-literal incorporation into the teacher's discourse of elicited or spontaneous verbal contributions from the students about what they are learning.

Example: As Laura says, I have to move faster.

Incorporation of students' actions into the teacher's discourse (B4)

Incorporation into the teacher's discourse of a specific aspect of a student's motor behavior, with specific reference to the student involved, with the aim of guiding learning.

Example: Did you see how Luis moves his feet when skipping?

Characterization of knowledge as something shared (B5)

References to the subject matter or the task at hand, or their results, systematically using the first person plural (we), and drawing attention to what has been learned or is about to be learned, with the inclusion of a positive evaluation.

Example: We have successfully kept the ball in the air.

Acknowledgment of acquired personal knowledge (B6)

References to current tasks or their results using the second or third person singular or plural (you, he/she, they) and highlighting something that has been learned.

Example: Sandra, your shot was very good; you positioned your hands and feet just like we said you should earlier.

Praise for verbal contribution or action (B7)

References to current activities or their results using the second or third person singular or plural (you, he/she, they) in response to a motor behavior or verbal comment by a student or group of students, but without mention of a specific type of learning.

Examples Very good! Nice! Perfect! Great! Excellent!

Criterion 3. Progressive establishment of increasingly expert and complex representations of subject matter

---

Re-elaboration of student contributions (C1)

Re-elaboration of a spontaneous or elicited motor or verbal contribution from a student, where the teacher expands, develops, reorganizes, trims, or corrects the relevant information.

Example: Michael says that if we throw the ball in the air, we push our bodies upwards, and if we throw it in front of us, we push our bodies forwards.

Characterization and labeling of aspects of content or context C2

Redefinition and characterization of a concept, contextual aspects, an activity or its results; the teacher may do this spontaneously or use labels typically employed by the students.

Example: The leg in front is called the drive leg.

Introduction of different referential expressions (C3)

Introduction of new referents (spatial, temporal, tactical-strategic, biomechanic-technical and/or physical-physiological) in relation to the task the students are about to start, or to an object or concept. The task/object/concept is clearly identified and highlighted.

Example: When running in a hurdle race, it's not a good idea to jump over the hurdle when you are very close to it, as we can hurt ourselves. We are going to try to do it at a fast pace, with our front leg in a semi-bent position.

Cognitive transfer of learning to a future situation (C4)

Description and/or justification of how the object of the lesion or task can be applied in a future situation.

Example: We are going to work on our spatial-temporal perception, and this will help us to know whether we can cross the road safely or not when we see a car coming.

TABLE 2 | Number and percentage of discursive strategies used before and after participation in the collaborative action research program.

Teachers	Teacher #1				Teacher #2				Teacher #3			
Phases	Pre-training		Post-training		Pre-training		Post-training		Pre-training		Post-training	
Categories (see Table 1)	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
A1 (social framework)	20	6.49	34	6.59	4	1.41	6	1.34	2	1.57	0	0.00
A2 (specific framework)	11	3.57	24	4.65	10	3.52	10	2.23	3	2.36	1	0.33
A3 (request for information)	53	17.21	175	33.91	64	22.54	143	31.85	62	48.82	149	48.53
B1 (meta-statements before task)	4	1.30	4	0.78	1	0.35	11	2.45	0	0.00	1	0.33
B2 (meta-statements during task)	3	0.97	2	0.39	1	0.35	10	2.23	0	0.00	0	0.00
B3 (incorporation of students' contributions)	28	9.09	111	21.51	31	10.92	73	16.26	25	19.69	79	25.73
B4 (incorporation of students' actions)	2	0.65	11	2.13	0	0.00	10	2.23	0	0.00	0	0.00
B5 (characterization of knowledge as shared)	0	0.00	1	0.19	0	0.00	19	4.23	0	0.00	0	0.00
B6 (acknowledgement of acquired knowledge)	1	0.32	26	5.04	3	1.06	35	7.80	0	0.00	0	0.00
B7 (praise for verbal contribution/action)	169	54.87	82	15.89	161	56.69	69	15.37	16	12.60	36	11.73
C1 (re-elaboration of student contribution)	8	2.60	40	7.75	8	2.82	33	7.35	18	14.17	33	10.75
C2 (characterization/labeling of content/context)	4	1.30	0	0.00	0	0.00	11	2.45	1	0.79	7	2.28
C3 (introduction of referential expressions)	2	0.65	6	1.16	0	0.00	7	1.56	0	0.00	1	0.33
C4 (cognitive transfer to future situation)	3	0.97	0	0.00	1	0.35	12	2.67	0	0.00	0	0.00
Total	308	100	516	100	284	100	449	100	127	100	307	100

the sessions were recorded using a Panasonic HDC-HS100 video camera fitted with a wireless audio recording system (AKGPR81 + PT81).

The intraobserver reliability of the data was checked using Krippendorf's canonical agreement coefficient (Krippendorf, 2004), which is an adaptation of Cohen's kappa statistic (Cohen, 1960), used to analyze at least three datasets collected at three different points in time. The analysis was performed in HOISAN (v. 1.6.3.3) (Hernández-Mendo et al., 2012). Interobserver reliability was tested by having each of the three observers code a randomly selected segment of 15 min on three occasions, separated by 10 days each. The results yielded a mean kappa coefficient of 0.97, indicating almost perfect agreement. The reliability of the data was also guaranteed by applying the consensus agreement method (Arana et al., 2016), which is a qualitative method in which observers agree on how to code a particular item before it is included in the dataset. The three observers were trained for over 80 h over a 6-month period and recorded 15% of the total session content using the consensus agreement method.

Data Analysis

Because the first objective, which was quantitative in nature, consisted of identifying the verbal behavior of the participating teachers, the dataset of events recorded during each session was processed using lag sequential analysis. This data analysis technique, proposed by Bakeman (1978), and subsequently extended by Bakeman and Gottman (1986) and Bakeman and Quera (2011), has proven to be highly effective in diverse fields (Lapresa et al., 2013; Roustan et al., 2013), and is extremely useful for analyzing datasets compiled from direct and/or indirect observation that contain sequences of behaviors coded using an

ad hoc observation instrument. The first step in this analysis is to define our *criterion behaviors* (the starting point of any possible patterns detected) and to apply the time lags defined for the study. Observed probabilities were calculated for each of the lags using the binomial test; this test produces adjusted residuals (Allison and Liker, 1982), which show the strength of association between significantly associated categories (i.e., between criterion behaviors and the conditional behaviors with which they are associated). The level of significance was set at $p < 0.05$. Adjusted residuals are prospective when the lags are analyzed in a forward direction from the criterion behavior (lags +1, +2, etc.) and retrospective when they are analyzed in a backward direction (lags −1, −2, etc.). Adjusted residual values higher than 1.96 and lower than 1.96 are therefore statistically significant. In this study, we looked at two retrospective lags (−2, −1) and two prospective lags (+1 and +2). In other words, we looked at the two events that occurred immediately before the criterion behavior and the two events that occurred immediately afterwards.

We also performed a descriptive statistical analysis of the number and percentage of discursive strategies used during the two teaching modules analyzed (**Table 2**).

RESULTS

Descriptive Analysis

Table 2 shows the descriptive statistics for the discursive strategies observed for each teacher before and after participation in the action research program.

An increase in the frequency and variety of discursive strategies employed by the teachers was observed in the post-training phase, indicating that participation in the collaborative action research program provided the teachers with a greater
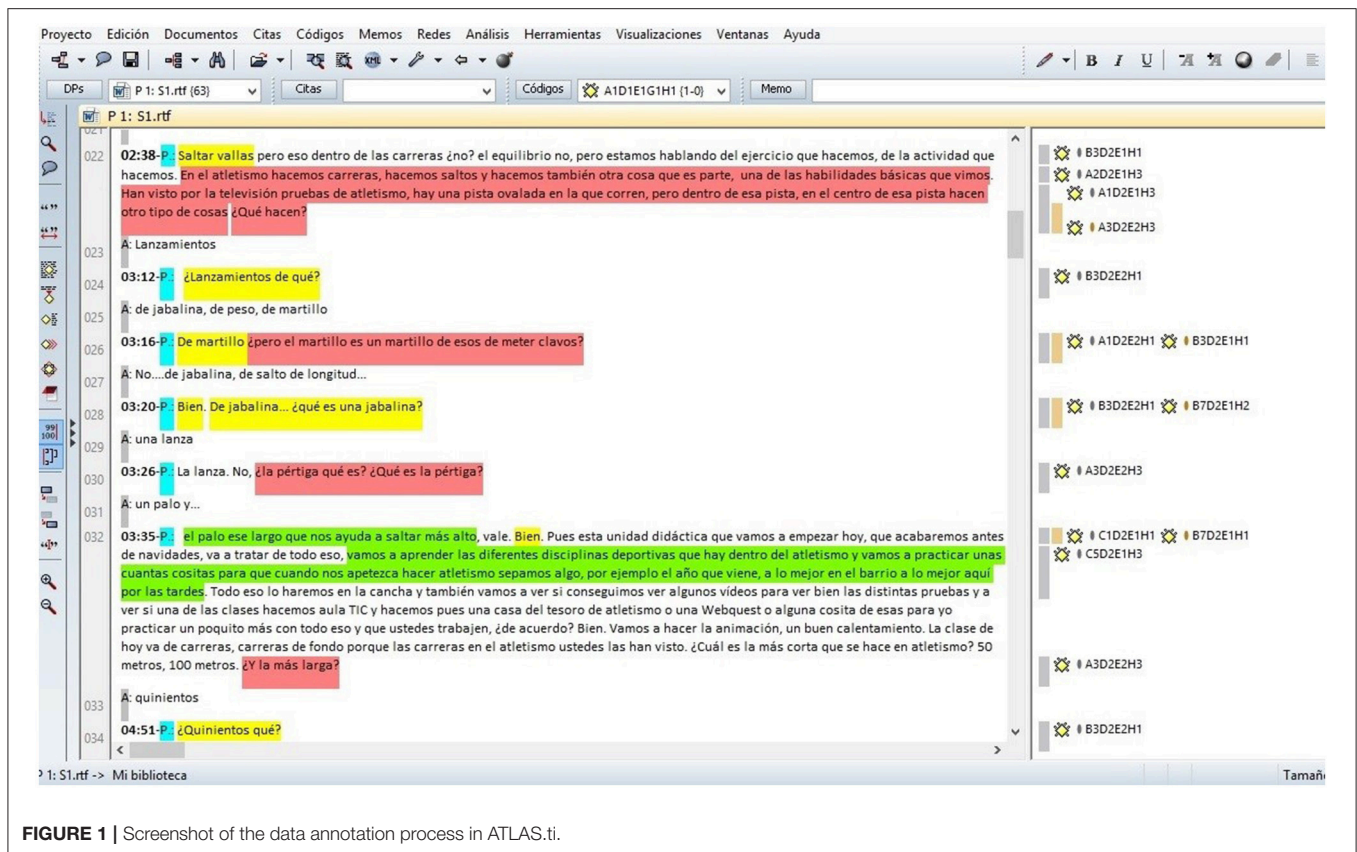
FIGURE 1 | Screenshot of the data annotation process in ATLAS.ti.

repertoire of discourse tools and resources with which to construct knowledge with their students.

Detection of Communication Patterns

Tables 3–**5** show the adjusted residual values for the retrospective lags (−1, −2) and the prospective lags (+1, +2) for teachers #1, #2, and #3, respectively, before and after participation in the collaborative action research program (pre- and post-training). The first cell in each row shows the criterion behavior, while the remaining cells show the respective conditional behaviors and the corresponding adjusted residuals.

For teacher #1 in the pre-training phase, a strong, stable association was observed between category A1 (social framework) and acknowledgment of acquired knowledge (B6, adjusted residual = 3, 87) and request for information (A3) at lag 2 (adjusted residual = 2.03) (**Table 3**).

Example:

Teacher: You have two weights and two discs over there, but be careful as it is very heavy. It's made of very hard rubber like the rubber on trucks (A1). You picked that up really well Jorge with your hands, opening your fingers (B6). How do you all think we can throw this weight? (A3).

This indicates that teacher #1 tends to ask questions immediately after making a comment linking the subject matter or task to everyday, social aspects. Requests for information (A3) were predominantly followed by incorporation of student

contributions into the teacher's discourse (B3, adjusted residual = 4.82) or re-elaboration of contributions (C1, adjusted residual = 2.49).

Example:

Teacher: What do we need to take into account in a race that lasts for a long time? (A3).
Student: Speed.
Teacher: Speed (B3). What do we do with speed Alba? (A3).
Student: Control it.
Teacher: Control it, spread out our energy (C1).

The above exchange shows a pattern formed by a question that triggers an answer, which is repeated and then elaborated on.

The pattern observed for teacher #2 (**Table 3**) was very similar, with requests for information strongly associated with incorporation of contributions (B3, adjusted residual = 8.81) and re-elaborations (C1, adjusted residual = 4.42).

Example:

Teacher: Sandra, tell me one way of warming up (A3).
Student: Heels back.
Teacher: Heels back (B3).

In this case, reference to the social framework (A1) was slightly more strongly associated with the use of meta-statements during task execution (B2, adjusted residual = 8.32), indicating that the teacher's strategy was to link the learning

TABLE 3 | Adjusted residuals for teacher #1 at the four lags analyzed before and after the collaborative action research program.

	Lag−2		Lag−1		Lag 1		Lag 2	
	Pre-training	Post-training	Pre-training	Post-training	Pre-training	Post-training	Pre-training	Post-training
A1	A2 2.88 C3 2.52		A2 2.81 B6 3.76		B6 3.87		A3 2.03 C2 3.7	B5 3.73
A2		B1 2.03 B6 2.83		B1 4.41 B6 2.75	A1 2.81 C3 3.51		A1 2.88	
A3	A1 2.03		B1 3.86	B3 2.27	B3 4.82 C1 2.49	B3 5.86 C1 6.08	B6 2.17 C1 2.44	
B1			B1 3.44		A3 3.86	A2 4.41	B3 3.56	A2 2.03 B4 3.14
B2		B6 2.88					C4 5.62	B7 3.25
B3	B1 3.56		A3 4.82	A3 5.86	B1 3.44 C1 2.79	A3 2.27		
B4		B1 3.14				B7 4.35		
B5		A1 3.73		B7 2.31				
B6	A3 2.17	B7 3.75	A1 3.87	B7 3.79	A1 3.76	A2 2.75 B7 2.65		A2 2.83 B2 2.88
B7		B2 3.25		B4 4.35 B6 2.65		B5 2.31 B6 3.79		B6 3.75
C1	A3 2.44	B3 2.6	A3 2.49 B3 2.79	A3 6.08			C4 3.29	
C2	A1 3.7							
C3			A2 3.51			B7 3.4	A1 2.52	
C4	B2 5.62 C1 3.29							

Adjusted residual values >1.96 implies p < 0.05.

objective to sociocultural aspects. The social framework was also associated, but to a lesser extent, with re-elaborations (C1, adjusted residual = 2.68).

Example:

Teacher: If the person holding it touches you, you have to stand on one leg, as if you were a stork (A1).
Student: Like in the red cross game.
Teacher: Yes, like in the red cross game but here you can be freed and saved (C1).

The teacher shows concern for establishing links between what the students already know and what is being taught. She links concepts from the animal world to the rules of the game to help the students to understand them. In the case of teacher #3, requests for information were also associated with re-elaborations (C1, adjusted residual = 3.06), showing a desire to explore and build on previous knowledge. Labeling (C2) was also associated with a literal incorporation of the students'

contributions into the discourse of teacher #3 (B3, adjusted residual = 2.03).

For teacher #1, the association observed in the pre-training phase between requests for information (A3) preceded by B3 (adjusted residual = 5.86) and C1 (adjusted residual = 6.08) was even stronger in the post-training phase, showing that the teacher continued to use this discursive pattern as a means of constructing knowledge (**Table 3**).
Example:

Teacher: They are practicing techniques. Which ones? (A3).
Student: Dodging.
Teacher: Dodging, dribbling, and feinting (C1).

The teacher constantly interacts with the students by asking them questions, acknowledging their answers, and then elaborating on them for the benefit of the group. We also observed a new association between the use of meta-statements (B1) and a specific framework (A2) at lag 1 (adjusted residual = 4.41) and lag 2 (adjusted residual = 2.03).

TABLE 4 | Adjusted residuals for teacher #2 at the four lags analyzed before and after the collaborative action research program.

	Lag−2		Lag−1		Lag 1		Lag 2	
	Pre-training	Post-training	Pre-training	Post-training	Pre-training	Post-training	Pre-training	Post-training
A1		B6 2.66		A2 2.4 C2 2.4	B2 8.32 C1 2.68	B1 2.25 B2 2.4 C2 2.25		B1 2.23
A2	C1 3.28 C4 5.17	B7 3.4	C4 5.21	B7 3.39		A1 2.4	C4 5.17	B1 3.58
A3		B3 2.68		C1 2.5	B3 8.81 C1 4.42	B3 5.61 C1 4.59		B5 2.15
B1		A1 2.23 A2 3.58		A1 2.25 B5 2.51			B3 2.87	B6 2.15
B2			A1 8.32	A1 2.4 C2 3.61		B7 2.19		B7 3.05
B3	B1 2.87	B5 2.11	A3 8.81	A3 5.61	A3 2.68			
B4								
B5		A3 2.15				B1 2.51		B3 2.11
B6		B1 2.45 B7 2.89 C3 3.5		B7 2.87				A1 2.66
B7		B2 3.05		B2 2.19		A2 3.39 B6 2.87		A2 3.4 B6 2.89
C1		C4 2.5	A1 2.68 A3 4.42	A3 4.59		A3 2.5	A2 3.28	
C2				A1 2.25		A1 2.4 B1 3.61		
C3						B2 2.16	A2 5.17	B6 3.5
C4	A2 5.17				A2 5.21			C1 2.58

Adjusted residual values >1.96 implies p < 0.05.

Example:

Teacher: Now we are going to learn how to pass the ball with the stick and to shoot. (B1). Does anyone remember how to hit the ball; we saw it yesterday? (A2).

The above example shows the use of a new discursive strategy involving commenting on the learning objective and linking it to a previous shared experience, thereby aiding comprehension. The teacher also incorporated the students' actions into his communication (B4) and combined this with praise (B7, adjusted residual = 4.35).

Example:

Teacher: Look how Carlos is holding the stick (B4). Good Miguel (B7), Good Luis (B7).

We also observed a recurrent pattern consisting of the prospective and retrospective interlinking of praise (B7) and recognition (B6), indicating concern for creating a positive learning climate.
Example:

Teacher: Nice Carlos (B7), good pass Dailos (B6).

The teacher also praised the students when comments were made by the group (B5, adjusted residual = 2.31). The above observations strongly suggest that participation in the collaborative action research program led teacher #1 to adopt new discursive strategies as a means of constructing knowledge.

The number of significant associations between the discursive strategies analyzed was also higher for teacher #2 in the post-training phase (**Table 4**). First, the social framework (A1) was strongly associated with labeling (C2, adjusted residual = 2.25).

TABLE 5 | Adjusted residuals for teacher #3 at the four lags analyzed before and after the collaborative action research program.

	Lag−2		Lag−1		Lag 1		Lag 2	
	Pre-training	Post- training	Pre-training	Post-training	Pre-training	Post-training	Pre-training	Post-training
A1							C1 2.36	
A2								
A3			B3 3.32 C2 1.9		C1 3.06	B3 6.28		
B1								
B2								
B3			A3 6.28		C2 2.03	A3 3.32		A3 2.37
B4								
B5								
B6								
B7			C3 2.73				C2 2.62	
C1	A1 2.36 C2 2.36		A3 3.06					
C2	B7 2.62	B3 2.03				A3 1.99	C1 2.36	
C3						B7 2.73		
C4								

Adjusted residual values >1.96 implies p < 0.05.

Example:

Teacher: It's shaped like Indian feathers (A1) but it's not a duster, it's called a shuttlecock or an *indiaca* (C2).

A1 was also associated with the use of meta-statements before (B1, adjusted residual = 2.25) and during the task (B2, adjusted residual = 2.4), as well as with A2 at lag −1 (adjusted residual = 2.4), showing that the teacher actively linked aspects of the task at hand to sociocultural content. The previously observed pattern between requests for information (A3) and incorporation of students' contributions (B3) and re-elaborations (C1) was still present but stronger (adjusted residual = 5.61 and adjusted residual = 4.59, respectively). Finally, incorporation of new referential expressions (C3) was associated with the use of meta-statements during the task (B2, adjusted residual = 2.16) indicating a concern for highlighting the important aspects of the task at hand.

Example:

Teacher: If you are going to shoot hard, stand away from the wall a little, look at the distance and think about how hard you are going to kick the ball (C3) and remember that we are practicing shooting and receiving in this task (B2).

The stronger associations observed between categories and the greater number of patterns suggest that this teacher intentionally incorporated a greater range of strategies into his teaching.

In the post-training stage, teacher #3 (**Table 5**) continued to use the communication pattern consisting of requests for information followed by incorporation of student contributions (B3, adjusted residual = 6.28).

Example:

Teacher: What do you know about baseball?
Student: You have to bat the ball.
Teacher: You have to bat the ball (B3). And what else? (A3).
Student: Be fast.
Teacher: Be fast (B3).

In the pre-training phase, there was a significant association between A3 and C1, while in the post-training phase; there was a significant association between A3 and B3. Fewer associations were observed between discursive strategies for this teacher than for teachers #1 and #2.

DISCUSSION

We have studied the verbal behavior of three teachers in their natural setting. Although each of these teachers is considered

as a "single case," they were monitored intensively over a series of sessions, resulting in the generation of large volumes of data, which, once converted into matrices of codes through annotation in ATLAS.it, were analyzed by lag sequential analysis to uncover patterns related to the use of social constructivist communication strategies. We are particularly interested in determining the extent to which single cases can reveal patterns that can then be merged, either partially or fully, to methodologically advance toward a multiple case, as proposed by Stake (2006) and Yin (2014).

We wished to investigate whether participation in a collaborative action research program would result in significant changes in the use of discursive strategies of a social constructivist nature by physical education teachers. Our analysis of these strategies by primary school physical education teachers shows a clear pattern composed of questions-answers-literal incorporation-re-elaboration of students' answers both before and after participation in a collaborative action research program designed to improve familiarity with and use of constructivist discursive strategies as a methodological resource. Such strategies encourage students to engage more actively in their learning, as claimed by Cazden (1988), Lemke (1990), Lobato et al. (2005) and Wright and Forrest (2007), who highlighted the importance of the triadic IRE dialogue pattern. The recurrent discursive pattern observed in our study (request for information (A3) + incorporation of students' contributions (B3), like request for information (A3) + re-elaboration of student contribution (C1), which is similar to the ERE pattern (Bowers and Nickerson, 2001), provides teachers with the means to guide their students toward the construction of significant meaning through the use of questions, reasoning, and argumentation. In this case, evaluation of students' answers leads teachers to take two decisions, i.e., to incorporate what the students say into their discourse and to re-elaborate when the answer is incomplete. Use of questioning to promote learning has been advocated by many authors (Wallian and Chang, 2007; Harvey and Light, 2015), who have shown that the use of open-ended questions in the classroom encourages reflective learning (Dyson et al., 2004; Mitchell et al., 2006; Hubball et al., 2007). Similarly, teachers who use closed questions to control construed meanings are better positioned to guide and elaborate on answers and to draw students' attention to the relevance or importance of certain learning points. Such strategies have been shown to play an important role in aiding understanding (Webster et al., 2012, 2013). In our analysis, just one change in the use of discursive strategies was observed for teacher #3 following his participation in the collaborative action research program. The observation of additional associations: meta-statements before task (B1) + request for information (A3) and incorporation of students' contributions (B3) + meta-statements before task (B1) in the pre-training phase for teacher #1 shows that this teacher was already using some of these strategies, even though he was not familiar with the theory behind them.

Participation in the collaborative action research program appears to have had a positive impact on teaching performance, as we detected an increase in the number and strength of associations observed in the post-training phase, suggesting that the use of new communication patterns was both intentional and strategic. The fact that the teachers recognized the usefulness of the strategies is evident through statements such as: "I can see that the kids are improving. I think that they are understanding things better and are doing the exercises with a greater understanding of why they are doing them and they are also making an effort to do things a little better, this gives me the strength to keep doing things and to keep trying. It's mutual reinforcement." They acknowledged the advantages of using constructivist techniques, probably because they feel that they will make their work easier and help their students to learn better. The patterns detected show that the teachers prefer to explore students' knowledge and reinforce correct answers rather than advance this knowledge to a more expert form; one exception is the use of re-elaborations of student contributions in the post-training phase. This greater tendency to explore and reinforce learning may be related to the short duration of the teaching modules analyzed. The identification of stable sequences in the forms of patterns as opposed to the use of isolated categories in the pre-training phase may indicate that the associations observed between categories from criteria 1 and 2 in the observation instrument reflects acquired practices, or habits, rather than an intentional, strategic use of strategies grounded in theoretical knowledge. The post-training results, by contrast, show that the teachers were familiar with the theory underlying the strategies they were incorporating into their instructional communication. It would therefore appear that participation in the collaborative action research program equipped the teachers with a greater repertoire of discursive strategies to actively engage students in the joint enterprise of learning.

We found that the three teachers all modified their use of discursive strategies after participation in the program. Particularly noticeable were improvements in the use of praise (B7), which was associated with incorporation of students' actions (B4) and acknowledgement of acquired knowledge (B6) as forms of recognition during task execution in the case of teacher #1. This observation reflects an increased interest in creating a positive learning climate. The continuous linking of previous knowledge is necessary to build knowledge, and in the post-training phase, teacher #2 intentionally used patterns linking social frameworks (A1) to other categories, such as meta-statements before task (B1) and meta-statements during task (B2). We also observed a significant relationship between characterization and labeling (C2) with explanations of tasks (B1).

Participation in the collaborative action research program also brought about changes in the way the teachers communicated with their students, as reflected in comments such as "I can see better results, I have saved time, and I feel that I am communicating better with my students. I can use discursive strategies to improve my teaching." The incorporation of new strategies also indicates the teachers' concern for improving both the teaching and learning process. Our findings support the usefulness of collaborative action research programs as an effective means of perfecting teaching performance.

Discursive strategies, which involve the conscientious use of language, should be used both strategically and intentionally in the classroom. Teachers need to know which form of language to use and when, and to see discursive strategies as a methodological resource rather than a means of support for their teaching activities. Teachers who use discursive strategies are thus effectively incorporating the potential of a scientific theory into their teaching practice and linking this to academic content. The integration of different formal and informal learning processes is particularly important in competence-based learning that aims to help students relate learning strategies to content and to use them effectively in different situations and contexts.

The limitations of our study are largely related to the difficulties associated with working with verbal behavior, as there is a risk of drawing inferences from the theoretical framework used as a reference for building the observation instrument.

The results of this study should bring us to reflect on the effectiveness of the methodological resources we conscientiously use in the classroom and on the functionality of the discursive strategies used by physical education teachers.

CONCLUSIONS

The teachers showed a consistent use of constructivist discursive strategies before and after participation in a research action program. The pattern detected consisted of requests for information followed by the incorporation of the students' contributions into their communication and the re-elaboration of their answers.

Following participation in this program, the teachers were seen to use more discursive strategies, generating new patterns.

By using lag sequential analysis, we were able to uncover hidden yet solid, meaningful patterns in the instructional communication of physical education teachers and to generate information of potential value for both teachers and researchers.

REFERENCES

Agbuga, B., Xiang, P., McBride, R. E., and Su, X. (2016). Student perceptions of instructional choices in middle school physical education. *J. Teach. Phys. Educ.* 35, 138–148. doi: 10.1123/jtpe.2015-0010

Allison, P. C. (1990). Classroom teachers' observation of physical education lessons. *J. Teach. Phys. Educ.* 9, 272–283. doi: 10.1123/jtpe.9.4.272

Allison, P. D., and Liker, J. K. (1982). Analyzing sequential categorical data on dyadic interaction: a comment on Gottman. *Psychol. Bull.* 91, 393–403. doi: 10.1037/0033-2909.91.2.393

Anguera, M. T. (2003). "Observational Methods (General)," in *Encyclopedia of Psychological Assessment,* Vol. 2, ed R. Fernández-Ballesteros (London: Sage), 632–637.

Anguera, M. T., Camerino, O., Castañer, M., Sánchez-Algarra, P., and Onwuegbuzie, A. J. (2017). The specificity of observational studies in physical activity and sports sciences: moving forward in mixed methods research and proposals for achieving quantitative and qualitative symmetry. *Front. Psychol.* 8:2196. doi: 10.3389/fpsyg.2017.02196

Anguera, M. T., and Izquierdo, C. (2006). "Methodological approaches in human communication. From complexity of situation to data analysis," in *From Communication to Presence. Cognition, Emotions and Culture towards the Ultimate Communicative Experience,* eds G. Riva, M. T. Anguera, B. K. Wiederhold, and F. Mantovani (Amsterdam: IOS Press), 203–222.

Anguera, M. T., Magnusson, M. S., and Jonsson, G. K. (2007). Instrumentos no estándar [No standard instruments]. *Av. Medición* 5, 63–82.

Anguera, M. T., Portell, M., Chacón-Moscoso, S., and Sanduvete-Chaves, S. (2018). Indirect observation in everyday contexts: concepts and methodological guidelines within a mixed methods framework. *Front. Psychol.* 9:13. doi: 10.3389/fpsyg.2018.00013

Arana, J., Lapresa, D., Anguera, M. T., and Garzón, B. (2016). *Ad hoc* procedure for optimising agreement between observational records. *Anal. Psicol.* 32, 589–595. doi: 10.6018/analesps.32.2.213551

Azzarito, L., and Ennis, C. D. (2003). A sense of connection: toward social constructivist physical education. *Sport Educ. Soc.* 8, 179–197. doi: 10.1080/13573320309255

Bakeman, R. (1978). "Untangling streams of behavior: sequential analyses of observation data," in *Observing Behavior. Vol. 2: Data Collection and Analysis Methods*, ed G. P. Sackett (Baltimore: University Park Press), 63–78.

Bakeman, R., and Gottman, J. M. (1986). *Observing Interaction. An introduction to Sequential Analysis*. Cambridge: Cambridge University Press.

Bakeman, R., and Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge: Cambridge University Press.

Blanco-Villaseñor, A., Losada, J. L., and Anguera, M. T. (2003). Analytic techniques in observational designs in environment-behavior relation. *Med. Amb. y Comp. Hum.* 4, 111–126.

Bowers, J. S., and Nickerson, S. (2001). Identifying cyclic patterns of interaction to study individual and collective learning. *Math. Think. Learn.* 3, 1-28. doi: 10.1207/S15327833MTL0301_01

Bunker, D., and Thorpe, R. (1982). A model for the teaching of games in secondary schools. *Bull. Phys. Educ.* 18, 5–8.

Carr, W., and Kemmis, S. (1986). *Teoría crítica de la Enseñanza [Toward a Critical Educational Science]*. Philadelphia, PA: Falmer Press.

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Canton, A., and Hileno, R. (2016). Goal scoring in soccer: a polar coordinate analysis of motor skills used by Lionel Messi. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Fernandes, T., and Hileno, R. (2017). Mastery in goal scoring, T-pattern detection and polar coordinate analysis of motor skills used by Lionel Messi and Cristiano Ronaldo. *Front. Psychol.* 8:741. doi: 10.3389/fpsyg.2017.00741

Castañer, M., Camerino, O., Anguera, M. T., and Jonsson, G. K. (2013). Kinesics and proxemics communication of expert and novice PE teachers. *Qual. Quant.* 47, 1813–1829. doi: 10.1007/s11135-011-9628-5

Cazden, C. (1988). *Classroom Discourse: The Language of Teaching and Learning*. Portsmouth, NH: Heinemann.

Chesebro, J. L., and Wanzer, M. B. (2006). "Instructional message variables," in *Handbook of Instructional communication*, eds T. P. Mottet, V. P. Richmond, and J. C. McCroskey (Boston, MA:Pearson/Allyn & Bacon), 89–116.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Coleman, M. M., and Mitchell, M. (2001). Assessing observation focus and conference targets of cooperating teachers. *J. Teach. Phys. Educ.* 20, 40–54. doi: 10.1123/jtpe.20.1.40

Coll, C., and Onrubia, J. (2001). Estrategias discursivas y recursos semióticos en la construcción sistemas de significados compartidos entre profesor y alumnos [Discourse strategies and semiotic resources in constructing systems of shared meanings between teacher and students]. *Invest. esc.* 45, 21–31.

Dyson, B., Griffin, L. L., and Hastie, P. (2004). Sport education, tactical games, and cooperative learning: theoretical and pedagogical considerations. *Quest* 56, 226–240. doi: 10.1080/00336297.2004.10491823

Eckrich, J., Widule, C. J., Shrader, R. A., and Maver, J. (1994). The effects of video observational training on video and live observational proficiency. *J. Teach. Phys. Educ.* 13, 216–227. doi: 10.1123/jtpe.13.3.216

Edwards, D., and Mercer, N. (1989). *Common Knowledge: The Development of Understanding in the Classroom*. London: Routledge.

Elliott, J. (1991). *Action Research for Educational Change*. Milton Keynes: Open University Press.

Haerens, L., Aelterman, N., Van den Berghe, L., De Meyer, J., Soenens, B., and Vansteenkiste, M. (2013). Observing physical education teachers' need-supportive interactions in classroom settings. *J. Sport Exerc. Psychol.* 35, 3–17. doi: 10.1123/jsep.35.1.3

Harvey, S., and Light, R. (2015). Questioning for learning in game-based approaches to teaching and coaching. *Asia-Pacif. J. Health, Sport and Phys. Educ.* 6, 175–190. doi: 10.1080/18377122.2015.1051268

Hernández-Mendo, A., López-López, J. A., Castellano Paulis, J., Morales Sánchez, V., and Pastrana Brincones, J. L. (2012). Hoisan 1.2: programa informático para uso en metodología observacional [Hoisan 1.2: software for observational methodology]. *Cuad. Psic. Dep.* 12, 55–78. doi: 10.4321/S1578-84232012000100006

Hubball, H., Lambert, J., and Hayes, S. (2007). Theory to practice: using the games for understanding approach in the teaching of invasion games. *Phys. Health Educ. J.* 73, 14–20.

Kirk, D., and MacPhail, A. (2002). Teaching games for understanding and situated learning: rethinking the bunker-thorpe model. *J. Teach. Phys. Educ.* 21, 177–192. doi: 10.1123/jtpe.21.2.177

Krippendorf, K. (2004). *Content Analysis: An Introduction to its Methodology, 2nd Edn.* Thousand Oaks, CA: Sage.

Lacy, A. C., and Darst, P. W. (1985). Systematic observation of behaviors of winning high school head football coaches. *J. Teach. Phys. Educ.* 4, 256–270. doi: 10.1123/jtpe.4.4.256

Lapresa, D., Arana, J., Anguera, M. T., and Garzón, B. (2013). Comparative analysis of the sequentiality using SDIS-GSEQ and THEME: a concrete example in soccer. *J. Sport. Sci.* 31, 1687–1695. doi: 10.1080/02640414.2013.796061

Lemke, J. (1990). *Talking Science: Language, Learning, and Values*. Norwood, NJ: Ablex.

Light, R. (2008). Complex learning theory – its epistemology and its assumptions about learning: implications for physical education. *J. Teach. Phys. Educ.* 27, 21–37. doi: 10.1123/jtpe.27.1.21

Light, R. (2014). "Positive pedagogy for physical education and sport: game sense as an example," in *Contemporary Developments in Games Teaching*, eds R. Light, J. Quay, S. Harvey, and A. Mooney (London: Routledge), 29–42.

Lobato, J., Clarke, D., and Ellis, B. (2005). Initiating and eliciting in teaching: a reformulation of telling. *J. Res. Math. Educ.* 36, 101–136.

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front. Psych.* 3:54. doi: 10.3389/fpsyg.2012.00054

Mitchell, S. A., Oslin, J. L., and Griffin, L. L. (2006). *Teaching Sport Concepts and Skills. A Tactical Games Approach, 2nd Edn.* Champaign, IL: Human Kinetics.

Morgan, K., and Kingston, K. (2008). Development of a self-observation mastery intervention programme for teacher education. *Phys. Educ. Sport Ped.* 13, 109–129. doi: 10.1080/17408980701345634

Oslin, J., and Mitchell, S. (2006). "Game-centred approaches to teaching physical education," in *The Handbook of Physical Education*, eds D. Kirk, D. Macdonald, and M. O'Sullivan (London: Sage), 627–651.

Portell, M., Anguera, M. T., Chacón-Moscoso, S., and Sanduvete-Chaves, S. (2015). Guidelines for reporting evaluations based on observational methodology. *Psicothema* 27, 283–289.

Powell, K. C., and Kalina, C. J. (2009). Cognitive and social constructivism: developing tools for an effective classroom. *Education* 130:241.

Rink, J. (1998). *Teaching Physical Education for Learning, 3rd Edn.* Boston: McGraw-Hill.

Roustan, M., Izquierdo Rodríguez, C., and Anguera Argilaga, M. T. (2013). Sequential analysis of an interactive peer support group. *Psicothema* 25, 396–401. doi: 10.7334/psicothema2012.93

Rovegno, I., and Dolly, J. P. (2006). "Constructivist perspectives on learning," in *The Handbook of Physical Education*, eds D. Kirk, D. Macdonald, and M. O'Sullivan (London: Sage), 242–261.

Sánchez-Algarra, P., and Anguera, M. T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: impact of recording and coding predominating perspectives. *Qual. Quant.* 47, 1237–1257. doi: 10.1007/s11135-012-9764-6

Stake, R. E. (2006). *Multiple Case Study Analysis*. New York, NY: Guilford Press.

Ussher, B., and Gibbes, C. (2002). Vygotsky, physical education and social interaction. *J. Phys. Educ. New Zeal.* 35, 76–86.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge: Harvard University Press.

Wallian, N., and Chang, C. W. (2007). Language, thinking and action: towards a semio-constructivist approach in physical education. *Phys. Educ. Sport Ped.* 12, 289–311. doi: 10.1080/17408980701610219

Webster, C. A. (2010). Increasing student motivation through teacher communication: six essential skills every physical educator should master. *J. Phys. Educ. Recreat. Dance* 81, 29–33. doi: 10.1080/07303084.2010.10598431

Webster, C. A., González, S., and Harvey, R. (2012). Physical education teachers' selfreported communication of content relevance. *Phys. Educ.* 69, 89–103.

Webster, C. A., Mîndrilǎ, D., and Weaver, G. (2011). The influence of state motivation, content relevance and affective learning on high school students' intentions to utilize class content following completion of compulsory physical education. *J. Teach. Phys. Educ.* 30, 231–247. doi: 10.1123/jtpe.30.3.231

Webster, C. A., Mîndrilǎ, D., and Weaver, G. (2013). Affective learning profiles in compulsory high school physical education. An instructional communication perspective. *J. Teach. Phys. Educ.* 32, 78–99. doi: 10.1123/jtpe.32.1.78

Wright, J., and Forrest, G. (2007). A social semiotic analysis of knowledge construction and games centred approaches to teaching. *Phys. Educ. Sport Ped.* 12, 273–287. doi: 10.1080/17408980701610201

Yin, R. K. (2014). *Case Study Research. Design and Methods, 5th Edn.* Los Angeles, CA: Sage.

Behavioral Patterns of Children Involved in Bullying Episodes

Carlos V. Santoyo[1]* and Brenda G. Mendoza[2]

[1] *Laboratorio de Desarrollo y Contexto del Comportamiento Social, División de Investigación y Posgrado, Universidad Nacional Autónoma de México, Delegación Coyoacán, Mexico,* [2] *Facultad de Ciencias de la Conducta, Universidad Autónoma del Estado de México, Toluca, Mexico*

This study applied a systematic observation strategy to identify coercive behavioral patterns in school environments. The aim was to describe stability and change in the behavioral patterns of children identified as victims of bullying. To this end, the following specific objectives were defined: (1) to identify episodes of bullying based on the frequency of negative behaviors received and power imbalances between bully and victim; (2) to describe stability and behavioral changes in student victims based on their social and academic conduct and the aggression they receive from peers and teachers; and (3) to describe the functional mechanisms responsible for the process of social organization (i.e., the Social Effectiveness, Social Responsiveness, and Social Reciprocity Indexes). The sample consisted of nine children identified as victims, nine classified as bullies, and nine matched controls, all elementary school students from the study developed at the National Autonomous University of Mexico files. A multidimensional/idiographic/follow-up observational design was used. Observational data describes asymmetry between victims and bullies based on microanalyses of the reciprocity of their behavioral exchanges. In addition, the behavioral patterns of victimized children were identified in relation to their academic activity and social relationships with peers. A model of coercive reciprocity accurately describes the asymmetry found among bullies, victims, and controls. A reduction in victimization was found to be related to: (1) responsiveness to the initiation of social interactions by peers and teachers; and (2) the time allocated to academic behavior during the study.

Keywords: bullying, behavioral patterns, children, victims, teachers

Studies of bullying have already identified serious detrimental effects, both short- and long-term, not only for victims (Lereya et al., 2015) but also for passive observers and the bullies themselves. Finally, it is clear that bullying negatively affects a school's social climate (Beaudoin and Roberge, 2015). Bullying is a type of aggression exhibited in a persistent manner via coercive behavior toward a person(s), with an existing power asymmetry between victim and aggressor (Olweus, 2001). Coercive behavior is expressed as the combination of functional events, bi-directional and generally asymmetric, where a person manipulates the conduct of others using the contingent presentation of aversive events that are removed when the others behavior takes the desired direction (Patterson, 1979).

In general, bullying has been observed using indirect measures (Olweus, 1993), as only a few studies 31 of 1,471 used observational methodologies that can provide novel empirical evidence for evaluating this phenomenon (Machado et al., 2015). Direct observation of patterns of social interaction facilitates identifying how social relationships are established, maintained, and modified (Cairns, 1979; Bakeman and Gottman, 1986; Espinosa, 2017), making this methodology suitable for the study of bullying. From an ecological perspective, this field takes into account the outcomes of how intra- and inter-individual dimensions relate over time (Modecki et al., 2014), based on multi-method research approaches.

Previous research on patterns of coercive behavior have used multi-method approaches that include behavioral data, self-reports and reports from peers and other adults, as well as sociometric indicators of behavior, status, and consensus (Santoyo et al., 2007b, 2008). The theoretical framework for this study corresponds to the sinthesys approach of Cairns (1979), and its implications to school settings (Cairns and Cairns, 1994), which is framed in the social organization processes, through which it is possible the study of social networks of the students through ecological social analysis, that has a sociometric technique called Social Cognitive Maps (SCM; Farmer and Cairns, 1991).

The analysis of social ecology in this perspective lies in studying the relevant elements of social links that children maintain at the school environment. SCMs, allow to identify the subgroups of students of a classroom, allowing to analyze their social interactions in the school setting.

For the analysis of the social links, it is necessary to use analyses of the *Functional Mechanisms* responsible for the process of social organization, and the use of the following social competency scales have been proposed: Social Effectiveness (SEI), Social Responsiveness (SRI), and Social Reciprocity Indexes (Santoyo, 1996).

As defined in Equation (1), the social effectiveness index describes the relative frequency of initiating acts by the target subject (TS) that result in a social episode (i.e., successful initiating acts), relative to the total number of initiating acts (with or without a peer response)

$$SEI = \frac{Succesful\ initiating\ acts}{Total\ initiating\ acts} \quad (1)$$

As Equation (2) shows, the Social Responsiveness Index describes the relative frequency of successful initiating acts directed toward the target that result in a social episode, relative to the total number of social initiating acts directed toward the target (with or without a target response).

$$SRI = \frac{Social\ response\ to\ acts}{Total\ initiating\ acts} \quad (2)$$

The third index reflects coercive reciprocity. It is useful for distinguishing behavioral symmetry between children who exhibit aggressive behavior and those who do not (Santoyo et al., 1996, 2007c; Espinosa, 2017). According to this index, aggressive children exhibit higher reciprocity in their coercive exchanges than a matched group (Santoyo et al., 2008, 2017). This kind of effect has also been found in the conflictive interaction of violent spouses (López and Santoyo, 2004): This relationship can be expressed using the coercive reciprocity model, as shown in Equation (3).

$$\frac{Nbe}{Nbe + Nbr} = \frac{Nsep}{Nsep + Nser} \quad (3)$$

The first part of this equation represents provocation events, where Nbe *(Negative behavior emitted)* indicates physical or verbal coercive behavior from the target toward a peer without the peer having addressed the target during the interval immediately preceding such behavior. Nbr *(Negative behavior received)* represents physical or verbal coercive behavior directed toward the target without provocation by her/him during the interval immediately preceding such behavior. The second part corresponds to the consequences of provocation: *Nsep (Negative social episodes produced)* represents negative social episodes initiated by the target, while *Nser (Negative social episodes received)* shows negative social episodes initiated by a peer. Negative social episodes are defined as physical and/or verbal behaviors between the focal subject and other persons that occur either simultaneously or successively, with mutual dependence on the participants' behavior.

Our review of current literature found no previous studies that attempted to identify these kinds of functional mechanisms of bullying based on dyadic interactions and assessments of mechanisms of effectiveness, responsiveness, and reciprocity.

In this study, the power imbalance or asymmetry between victim and bully was identified by analyzing the reciprocity mechanism (Equation 3) in light of the definition by Atlas and Pepler (1998), in which victims are defined as students who are targets of negative behaviors, and bullies as those who frequently initiate such behaviors. We adopted a synthesis approach based on implementing social ecology analysis (Cairns, 1979) that allows a better understanding of how both individual and social behavior among the members of a social network are regulated. This analysis is appropriate because studies have consistently shown that a risk factor for victimization by bullying is a lack of social links in the school environment (Salmivalli et al., 1996; Eslea et al., 2003; Mendoza and Maldonado, 2017). The present study provides evidence of social links based on the use of SCM (Farmer and Cairns, 1991), which could overcome the limitations of conventional sociometry, such as restricting nominations to only a few people, or naming a certain number of classmates when in reality no sustained relationship exists among them. The use of SCMs has the additional advantage of allowing the identification of existing sub-groups supported by statistical criteria.

Our study thus proposes an approach to the study of behavioral patterns exhibited by victims of bullying that employs an observational methodology in school settings.

This research is thus an extension of Study developed in school environments with a 3-year follow-up period (Santoyo, 2007; Santoyo and Colmenares, 2012), data are used to estimate

densities of individual and social activities, construct behavioral profiles of individuals, identify types and frequencies of social interactions, and integrate information on contextual and dyadic exchange. Thus, it is consistent with previous methodological strategies based on observing interactions and studies of social development in natural settings (Cairns et al., 1991).

Finally, our goal is to describe the behavioral patterns of children identified as victims of bullying in their social ecology. Achieving this entails addressing the following specific objectives: (1) identifying episodes of bullying based on the frequency of negative behaviors received and power imbalances between bully and victim; (2) describing stability and behavioral changes in student victims based on their social and academic conduct and the aggression they receive from peers and teachers; and (3) describing the functional mechanisms responsible for the process of social organization using the Social Effectiveness, Social Responsiveness, and Social Reciprocity Indexes. For this purpose, children in different grades of elementary school were selected. The study focuses on the behavioral and dyadic patterns of victims of bullying, matched controls, and bullies. Finally, a microanalysis of behavioral patterns was performed to obtain information on the conditional probabilities of behavioral acts by participants.

METHODS

Participants

The sample included 27 elementary school students aged six to nine. All subjects were attending a public school in Mexico City (first to third grade). The average number of children per classroom was 28. Nine of the children were identified as victims, nine others were matched as bullies, and nine were selected as matched controls. Written authorization to perform the research was obtained from school authorities. The project was approved previously by an ethics committee at the lead author's university. This work used behavioral observation methods which carefully preserved the identity of children and teachers. Observers never stood closer than 10 m from the children or interacted with them during the sessions. It should be pointed out that, at the time, observational records and notes were not considered as potentially damaging or harmful, given that they would never be made public or include personal identification of participants.

Design

A multidimensional/idiographic/follow-up observational design (Blanco-Villaseñor et al., 2003) was applied.

Selection Criteria

Victims Group

To be eligible to participate in the focal group, a student had to fulfill the following criteria:

1) Belong to the highest quartile for the frequency of negative behavior received, relative to the absolute frequency of positive and coercive behavior emitted by their peers.
2) Belong to the highest quartile for the absolute frequency of negative behavior received.

3) Coercive behavior emitted by the target could not exceed 6% of the student's total social behavior.

Only nine children were identified as victims. Then, other 18 children were selected as a member for each matched comparison group (control and bullies group).

Matched Control Group

For comparative purposes, for each child identified as a victim based on the selection criteria, a matched peer with similar characteristics of gender, age, group, and school grade was selected.

Bullies Group

In order to analyze the asymmetry between victims and bullies, and for purposes of comparison, nine students identified as bullies were included after verifying the following requirements:

1) Belonging to the highest quartile for the frequency of negative behavior emitted; and
2) Having coercive behavior that exceeded 6% of their total social behavior. This criterion identifies children that exhibit aggressive behavior, criterion established by Patterson (1982) and more recently supported and extended by empirical evidence by Cruz (2007) and Santoyo and Colmenares (2012) in school settings.

Setting

All observations took place in typical classrooms (during lessons) at a public elementary school in the south of Mexico City. The classrooms had adequate lighting and ventilation.

Instruments

Participants' behavioral data were collected based on the *Observational and Behavioral System of Social Interaction (OBSSI)* (Santoyo et al., 1994), which was designed specifically to study social interaction in school settings. The OBSSI makes it possible to identify events and situations that constitute behavioral patterns. It is an exclusive, exhaustive behavioral categories system based on 5-s intervals and constituted by representative behavioral categories for the actions that participants (previously designated as "target subjects") exhibit in educational settings. These categories are organized as follows:

a. Individual behavior: on-task, off-task, individual play
b. Social interaction:

  - Social actions initiated by a target child and directed at another person;
  - Social actions initiated by others and directed at a target child;
  - Such dyadic and group social interactions as: coercive behavior, group play, sharing, conversation, physical contact, etc.

c. Other responses:

- Behavior displayed by target child that is not covered by the other categories in the *Observational and Behavioral System of Social Interaction*.

The *Observational and Behavioral System of Social Interaction* thus generates an event-based, sequential record in which observers write the order of occurrence of events. Moreover, it allows the study of contextual factors at the site where a behavioral pattern emerges (i.e., classroom, playground, math lessons, Spanish lessons, etc.), and identifies the person who initiates an exchange. The use of this system made it possible to categorize participants' activities, the quality of social exchanges (coercive or prosocial), the social agents involved in social interactions (peers or teacher), and the direction and location of exchanges. Meanwhile, the contents of specific actions involving children are described by keywords or verbs that express the type of action emitted. For this study, the *Observational and Behavioral System of Social Interaction* categories used were: *Initiating acts, Response to acts, Social Interaction (identified as positive or negative), Academic activity (on-task behavior)* and *Other responses (off-task behavior)*.

Using *Observational and Behavioral System of Social Interaction*, researchers can identify three types of aggressive behavior: physical, verbal, and coercive or negative. Obviously non-aggressive behavior is also recorded. Previous studies obtained a 0.95 generalizability coefficient (Espinosa et al., 2006). This value indicates that the results from individuals and sessions can be reliably generalized based on the category system, number of participants, and the number of sessions programmed.

To describe social ecology, the sociometric technique called SCMs (Farmer and Cairns, 1991) adapted for use in Mexico was used (Santoyo and Espinosa, 2005). Here, students from the same class as the focal subjects were asked, individually, two questions: "Are there people in the class who hang around together a lot?" and "Are there people in the class who do not have a group?" When the interviewee did not include him/herself in any group, we asked: "What about you; do you have a group you hang around with at school? Based on these interviews, we generated a complete social network that identified the structure of relationships in the form of groups, sub-groups and isolated children. To identify social links and groups a co-occurrence matrices were designed, where each child was listed on the horizontal axis as a respondent and on the vertical axis as a nominee. The inclusion of a child in a sub-group was determined by a correlation equal to, or higher than, 0.40 (Farmer and Cairns, 1991). We classified weak and strong connections in the social network as those that showed significant correlations at the level of 0.01 and 0.005, respectively.

Procedure

A total of 84 children were observed during normal classroom lessons. The criteria applied identified nine children as victims, nine as bullies, and nine as matched controls. The SDIS-GSEQ program (Bakeman and Quera, 2011) was used to identify behavioral patterns (Santoyo et al., 2006). Each child was observed in focal samples for 90 min per year for 3 years.

Behavioral field data were collected using the *OBSSI*, followed by information from self-descriptive behavior and teachers' descriptions of the students in the class. This procedure was employed for 3 years, the data (cohorts) is derived from a study developed at the National Autonomous University of Mexico called Coyoacán Longitudinal Study, and allowed us to compare data stability throughout the follow-up period, as well as the behavioral patterns of participants in the three groups (victims, bullies, and controls). The sample was obtained on a *post hoc* basis (Elder et al., 1993), and during data collection neither the researchers nor the observers knew the status that had been assigned to the children (i.e., victims, bullies, or matched controls). In general, this strategy is based on a person-oriented approach (Cairns et al., 1998).

Pairs of trained observers collected the behavioral data, which satisfied the criterion of 80% reliability. In addition, a sample of the records (65%) was obtained and the Cohen's Kappa index (Cohen, 1960) was calculated, obtaining indexes of 0.81 for target subjects and 0.89 for matched controls. According to the parameters established by Fleiss (1981) and Bakeman and Gottman (1986), these data indicate excellent concordance.

Behavioral Sampling and Interviews

In order to compare behavioral patterns, the time allocated to different activities, and participants' social and behavioral preferences, a behavioral sampling was carried out *in situ* based on parameters tested in earlier studies (Santoyo et al., 2007b). For each participant, sampling entailed six 15-min sessions during classroom lessons. Efforts were made to conduct observation of each participant on consecutive days. Each subject was observed in the classroom for 90 min. To avoid interfering with the quality of the behavioral data, the interviews held to describe the social ecology, SCMs were implemented 1 week after completing the behavioral sampling. In this case, each student from the same class as the victim, bully and matched child was interviewed independently until the entire sample had been seen ($n = 84$).

Comparisons of the status and wave of measurement occasion were performed (Status = 3: victim, bully, matched control group; wave measurement occasions = 3: 1, 2, and 3). The sample was obtained on a *post hoc* basis, and during data collection neither the researchers nor the observers knew the status that had been assigned to the children—i.e., victims, bullies, or matched controls in a natural setting.

RESULTS

Characteristics of the Bullying Episodes

The first part of our results identifies the characteristics of the bullying episodes with the profiles of the students classified as bullies or victims. The outcomes of the power imbalance derive from one of the aforementioned equations of the *functional mechanisms* responsible for social interactions (Ec. 3, as outlined in the Introduction). One of the criteria used to distinguish bullying was the *frequency of negative behaviors received*. Student victims received negative behaviors more frequently than bullies or the matched control children. On average, victims received 13, 5, and 1%, respectively, of negative behavior from their peers

in the first, second and third waves of measurement. The latter figure represents a significant decrease, as shown by the results of a Tukey's test (first, 0.13 ± 0.01; second, 0.05 ± 0.01; third, 0.01 ± 0.008).
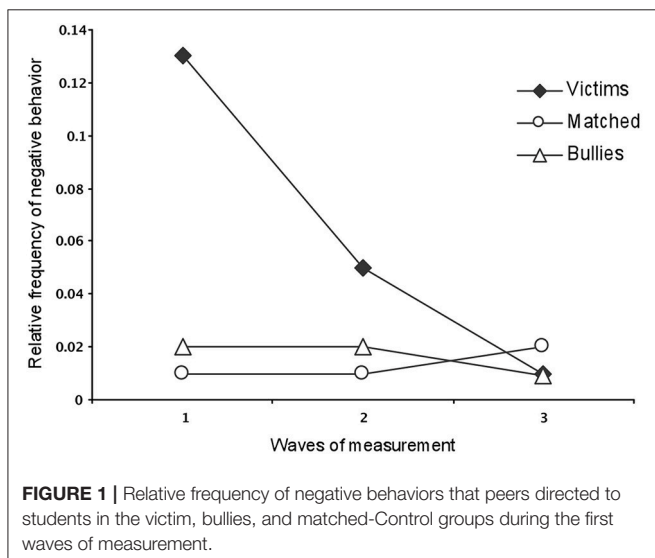
A Tukey's multiple comparison test also showed that victims received more negative behavior (0.10 ± 0.009) during the first wave of measurement than the bullies (0.02 ± 0.017; 0.02 ± 0.01; 0.009 ± 0.008) and the matched controls during the first, second and third waves of measurement (0.01 ± 0.17; 0.02 ± 0.01; 0.017 ± 0.008, respectively), $[F_{(4, 42 \text{d.f.})} = 5.67, p < 0.001]$.

Figure 1 shows the frequency of negative events received by the victims, bullies, and matched controls. The victims received a higher frequency (0.06 ± 0.006) of negative behaviors than bullies (0.02 ± 0.006) and controls (0.01 ± 0.006), and were the targets of over 5% of negative behaviors out of the total number of positive and negative behaviors directed at them $[F_{(2, 21 \text{d.f.})} = 17.81, p < 0.001]$.

The effect size was calculated, and a value of the effect size $f = 0.05$ was obtained, showing an effect size with a mean value (Cárdenas and Arancibia, 2014).

Finally, results show that in the third wave of measurement, 78% of the victims group exhibited a reduction in the relative frequency of harassment received, relative to the first wave (**Figure 1**).

Another criterion that distinguished bullying was the power imbalance, or asymmetry, between victim and bully. This relation was evaluated with the coercion reciprocity index derived from Equation (3). The reciprocity index was obtained for 100% of victims, bullies and controls. **Figures 2A,B** show the analysis of the reciprocity of coercive episodes for victims and bullies. The abscissa axis corresponds to relative provocations, the ordinate axis to their relative consequences. Values of 0.40–0.60 indicate high symmetry in provocation frequency; values above 0.60 indicate that the bullies consistently provoked conflicts; while values below 0.40 indicate that the conflicts were initiated by peers (Santoyo et al., 2007b).

FIGURE 1 | Relative frequency of negative behaviors that peers directed to students in the victim, bullies, and matched-Control groups during the first waves of measurement.

Figure 2A, shows that most of the bullies (six out of nine) consistently instigated conflicts (values above 0.60). The regression analysis based on Ec. 3 yielded an $r^2 = 0.72$, which confirms the bullies' coercive reciprocity.

The size of the effect was calculated and a value of the effect size $f = 2.5$ was obtained, showing an effect size with value denominated high (Cárdenas and Arancibia, 2014).

Figure 2B, shows that victims (eight out of nine) exhibited values below 0.60; that is, they were targets of coercive behaviors without provoking conflicts. The weak value of $r^2 = 0.25$ indicates that they tended not to respond symmetrically to the coercive behavior they received.

In summary, the power imbalance or asymmetry between bullies and victims was confirmed by the difference in the r^2-values and by the asymmetry in the location of the values obtained for these two groups based on the model of negative reciprocity (Ec. 3).

## Stability and Change Behavior: Victims

The second part of our results—shown below—describes the stability and change of behavior patterns identified for the victims, including social outcomes, academic behavior and aggression from teachers. The most striking result of **Figure 3** is that in the first wave of measurement the children in the victims group received a higher average frequency of negative behaviors from teachers (5 ± 0.99) than bullies (first, 2 ± 0.91; second, 1.8 ± 0.38; third, 0.62 ± 0.27) and the matched controls (first, 0.82 ± 0.99; second, 0.55 ± 0.38; third, 0.12 ± 0.27). These results were confirmed by a Tukey's multiple comparison test $[F_{(4, 42 \text{ d.f.})} = 4.85, p < 0.01]$.

The effect size was calculated, and a value of the effect size $f = 0.07$ was obtained, showing an effect size with medium value (Cárdenas and Arancibia, 2014).

We also observed a reduction in the average frequency of negative behaviors received by victims across the waves of measurement (5 ± 0.99; 0.50 ± 0.38; 0.12 ± 0.27, respectively, for the first, second and third waves), a finding consistent with the decrease in harassment that peers directed at victims.

Shown below are the results derived from the *functional mechanisms* responsible for the process of social preferences. With respect to the pattern of social behavior manifested by the student victims, **Figure 4** shows that they exhibited a greater increase in the value corresponding to the Social Responsiveness Mechanism (Equation 2) in the transition from the first (0.54 ± 0.03) to the third wave of measurement (0.72 ± 0.07). In the third wave, the matched control children established positive interactions with their peers at a rate of 47%; in contrast, the children in the victims group responded to 72% of their peers' attempt to initiate social interactions, and were able to establish positive interactions with them. This greater increase in the value corresponding to the social responsiveness index by the victims group is consistent with the decrease in negative behavior that victims received from their peers. These results were also confirmed by a Tukey's multiple comparisons test $[F_{(4, 42 \text{d.f.})} = 3.62, p < 0.05]$.
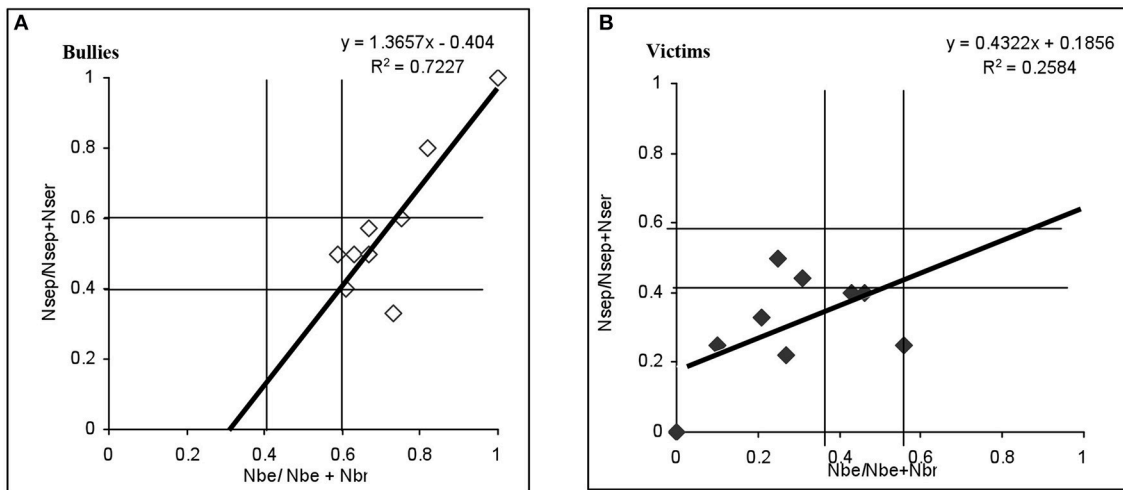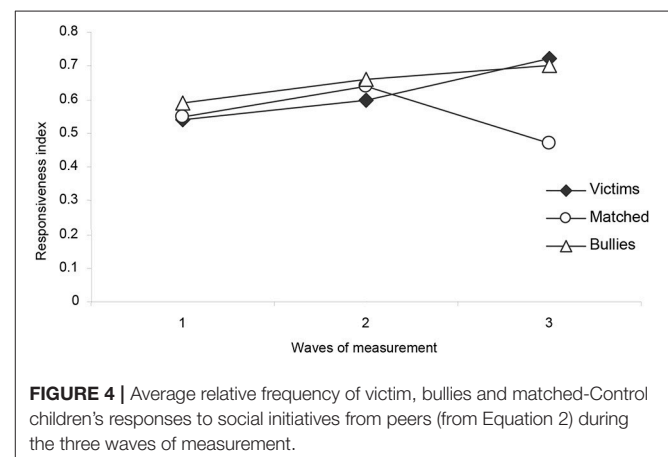
FIGURE 2 | Reciprocity of coercive events in children identified as bullies (A) and victims (B), from the three school grades, during the first wave of measurement (from Equation 3).
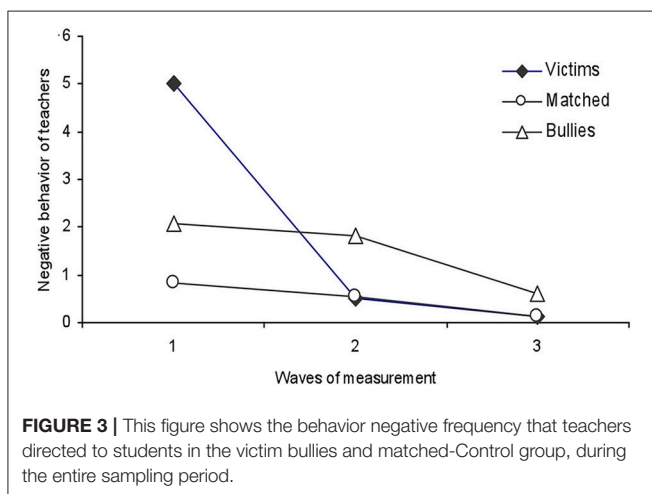
FIGURE 3 | This figure shows the behavior negative frequency that teachers directed to students in the victim bullies and matched-Control group, during the entire sampling period.

FIGURE 4 | Average relative frequency of victim, bullies and matched-Control children's responses to social initiatives from peers (from Equation 2) during the three waves of measurement.

The effect size was calculated, and a value of the effect size $f = 0.07$ was obtained, showing an effect size with medium value (Cárdenas and Arancibia, 2014).

Figure 5 shows that students from all three groups allocated 18% of their time to academic behavior during the first wave of measurement (0.18 ± 0.01), and that this increased to 25% in the second (0.25 ± 0.02), and 32% (0.32 ± 0.02) in the third. Hence, the amount of time assigned to academic work almost doubled in the transition from the first to the third waves. These results were confirmed by a Tukey's test [$F_{(2, 42 d.f.)} = 11.14, p < 0.001$].

The effect size was calculated, and a value of the effect size $f = 0.07$ was obtained, showing an effect size with medium value (Cárdenas and Arancibia, 2014).

Social Cognitive Maps (Social Ecology)

The description of the social ecology of the study setting is based on the SCMs of children in grades one two and three during

the first wave of measurement (see **Table 1**). We found that the matched control group had more links (44) than the children in the bullies (39) and victims groups (34). It is important to point out that two victims exhibited no links to their peers, while 77% of the children identified as bullies and matched controls, as well as 66% of the children identified as victims, had more than three links with their peers.

The negative behaviors of peers directed at the target children, and the corresponding responses of those children (bullies, victims and matched controls) were also examined (see **Table 2**), together with the target children's negative behavior directed toward others and the corresponding immediate responses of their peers (see **Table 3**). This analysis allowed us to identify whether or not the victims, bullies and matched control students became involved in negative social interactions in response to another classmate's negative behavior, or when they directed coercive behavior toward others. Results indicate that once victims received a coercive event, their probability of responding

was 0.46 (classified as "inhibitory" with an adjusted residual value of -2.5), in contrast to the matched controls whose probability of responding to a provocation was 0.74 (classified as "excitatory" with an adjusted residual value of 1.9). These results were confirmed by an X^2-value of (2 d.f.) $=7.77$ ($p < 0.05$) (see **Table 2**).

The effect size was calculated and a value of the effect size was obtained $f = 0.5$, showing an effect size with value denominated high (Cárdenas and Arancibia, 2014).

To extend the analysis, **Table 3** presents the results from the consequences of negative behavior emitted by the target children. This shows that the bullies group had a probability of 0.29 of receiving negative behavior when they directed physical or verbal coercive behavior toward other peers (inhibitory with an adjusted residual value of -2.7, in contrast to the victims group, whose probability of receiving negative behavior was 0.52; that is, excitatory with an adjusted residual value of 2.2 [$X^2_{(2\ d.f.)} = 6.99$ ($p < 0.05$); see **Table 3**]. The effect size was calculated and a value of the effect size was obtained $f = 0.45$, showing an effect size with value denominated high (Cárdenas and Arancibia, 2014).

DISCUSSION

Characteristics of the Bullying Episodes

In order to comply with the general objective of the study, we first employed systematic observation to analyze whether the aggressive events fulfilled the characteristics of bullying. In this case, the frequency of negative behavior directed toward victims,

FIGURE 5 | This figure shows the average time (in seconds) that students (victim, bullies, and matched-Control) allocate to academic behavior relative to the time spent on other behaviors during the three waves of measurement.

which has traditionally been measured using indirect instruments (Olweus, 1993).

Systematic observation used herein allowed us to overcome some of the difficulties associated with studying bullying at early ages, such as the distinction between aggressive behavior and bullying, identifying power asymmetry, and the persistence of negative behavior directed toward victims (Machado et al., 2015).

Stability and Change Behavior: Victims

Based on study data, we propose that in order for a child to be identified as a victim, 5% or more of the behavior that she/he receives from peers must be negative. This proposal extends the criterion suggested by Santoyo et al. (2007b) for identifying the coercive behavior of aggressive children with a sample of victims. Most social behavior is positive and some level of aggressive behavior is sometimes expected. In antisocial adolescents population (Patterson, 1982) risk boys shows more than 5% or more of negative behavior and control boys shows less than such percentage; for that, this criterion was proposed for student victims which are consistent with the adaptation proposed by Cruz (2007) with Mexican children in school settings. Our study further demonstrates that bullying is frequent in the school environment, a finding consistent with the results of Elgar et al. (2015). Thus, bullying is not necessarily "covert" (Olweus, 1993) or uncommon in the presence of adults (Landau and Swerdlik, 2005). Therefore, the systematic observation strategy proposed made it possible to identify the behavioral patterns of victims, bullies and matched children in school environments.

With respect to the general objective of the study, we were able to identify through our results that the behavior pattern of the children identified as victims evolved until they stopped "being so" (in the third cohort). This suggests that being able to cease being victimized is not related simply to the acquisition of social abilities but, rather, depends on several developmental and regulatory factors that do not develop in isolation, but appear to protect students who are at risk of being bullied. These results differ from those of Whitney and Smith (1993) and O'Moore et al. (1997), who indicate that the frequency of victimization remains constant in students aged seven to nine, decreases during the transition from elementary to middle school, and falls to zero when children reach the age of 16. It should be noted, however, that the sample in our study includes only a 3-year follow up, during elementary school education. Additional longitudinal studies are suggested which extend such follow up period.

These results suggest that a relationship exists between negative behavior that victims receive from their peers and
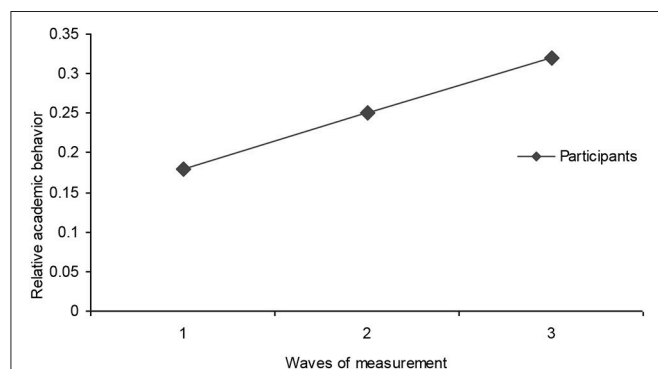
TABLE 1 | Total number of connections in the SCM by school grade as a function of the group to which students belong (victim, bully, and matched group); weak and strong connections (0.01 and 0.005 correlation coefficients, respectively).

Groups	Victims		Bullies		Matched		Total	
Level of Social Network	Low	High	Low	High	Low	High	Low	High
Links between peers	27	7	27	12	31	13	85	32
Sub total	34		39		44		117	

TABLE 2 | Conditional probability analysis of episodes of negative behavior received by SF (Nbr).

Events of negative behavior received by target (Nbr)	Negative social interactions resulted		
	Nser (ON)	No Nser (OFF)	Total
VICTIMS GROUP			
Frequency	41	47	88
Conditional probability	0.47	0.53	
Adjusted residual	−2.5*	2.5*	
BULLIES GROUP			
Frequency	31	15	46
Conditional probability	0.67	0.33	
Adjusted residual	1.7	−1.7	
MATCHED GROUP			
Frequency	20	7	27
Conditional probability	0.74	0.26	
Adjusted residual	1.9*	−1.9*	

FS getting involved (ON) or not getting involved (OFF), in negative interactions when they are the target of a provocation. *p < 0.05.

TABLE 3 | Conditional probability analysis of episodes of negative behavior emitted by SF (Nbe).

Events of negative behavior emitted by Target (Nbe)	Negative social interactions resulted		
	Nsep (ON)	No Nsep (OFF)	Total
VICTIMS GROUP			
Frequency	16	15	31
Conditional probability	0.52	0.48	
Adjusted residual	2.2*	−2.2*	
BULLIES GROUP			
Frequency	46	113	159
Conditional probability	0.29	0.71	
Adjusted residual	−2.7*	2.7*	
MATCHED GROUP			
Frequency	16	20	36
Conditional probability	0.44	0.56	
Adjusted residual	3.6*	3.6*	

FS getting involved (ON) or not getting involved (OFF), in negative interactions when FS emitted negative behavior to others. *p < 0.05.

from their teachers, findings that strengthens the evidence in studies developed with verbal reports, which indicate that victims of bullying could be also victims of their teachers (Mendoza, 2011, 2013). Next studies must search to the quality of social episodes with teachers. Also, this result is highly consistent with Wilson and Herrnstein (1985) theory about crime and human nature which highlights the role of strengthening non-aggressive behavior as a way of strengthening incompatible patterns of response with bullying and victimization, like prosocial behavior or academic behavior.

With regard to the objective of the study, and descriptions of the changes in the behavior patterns of child victims, our results suggest that the reduction in victimization is related to being highly responsive to the social initiatives of one's peers (i.e., a high social responsiveness index) and to academic motivation. Indeed, consistent with matching law (Wilson and Herrnstein, 1985), the groups of children (victims, bullies, and matched controls) became integrated, and a clear reduction in victimization was shown as the time assigned to academic behavior increased. These results represent a social and motivational index of great impact that may be related to decreased victimization in the classroom. This evidence also supports the findings of Turunen et al. (2017), who demonstrated that bullying interferes with both the victims' and the bullies' learning, coupled with the observation that coercive children show a low preference for academic activities (Cuenca and Mendoza, 2017; Santoyo et al., 2017), that being a victim is a multifactorial phenomenon, so for its attention and prevention it must be a comprehensive program that includes social skills, self-control, motivation for academic activities, also supervising the establishment of positive interactions with school authorities (teachers, managers, etc.).

Thus, results suggest the need to increase the value of academic behavior by implementing stimulating intellectual activities that will decrease the relative value of coercive behavior.

Social Cognitive Maps (Social Ecology)

To describe the stability of the behavior patterns of child victims (general objective) in relation to the social ecology of victims, bullies and matched controls, the findings from this study contrast with a rather large corpus of evidence which indicates that bullying is related to the absence of friendships (Eslea et al., 2003), lower levels of participation in social activities at school (Yüksel-Sahina, 2015), victims' isolation during activities (De Oliveira et al., 2016; Mendoza and Maldonado, 2017), and scant positive interaction with classmates (Mendoza and Maldonado, 2017). Based on the behavioral data and the socio-cognitive perspective (Farmer and Cairns, 1991), our results demonstrate that victims do not necessarily lack associations with their peers, since 66% of victims had three or more such links. This evidence is strengthened by the observation that, although victims are less responsive to their matched controls than the bullies, they do establish positive social interactions. It is important to point out that the methodological approach used to analyze the social ecology of the school setting (SCMs) has been recognized as a powerful predictor of social behavior, even more so than indirect psychometric measures (Santoyo et al., 2007a).

Another important result from this study is that bullies exhibit coercive behavior toward their classmates when there is a low probability that their attacks will be answered. This establishes a contingent relationship between the negative behavior and its consequences. Current findings (Santoyo et al., 2017) indicate that the violent behavioral pattern of some bullies is not regulated only by the relative reinforcement they receive, but also via negative reinforcement that victims receive (Sidman, 1989). Future studies should analyze the role of such regulatory mechanism in victims' behavior, incorporating the analysis of coercive interactions in the playground area with special

emphasis on the analysis of the consequences received within the school environment that would provide valuable information, not only on the direct consequences of coercive episodes, but also on which participants receive by non-coercive behavior (Wilson and Herrnstein, 1985).

Finally, in this study the power imbalance or asymmetry between victim and bully was evaluated by analyzing coercive social interactions. Thus, we confirmed, as indicated by Atlas and Pepler (1998), that victims become involved in conflicts initiated by other classmates, and that bullies initiate such conflicts. This serves to demonstrate that—based on an observational methodology in natural settings—the coercive reciprocity model is considered a sensitive strategy for identifying bullies, and an option that permits a more complete description and understanding of the asymmetry that exists in the dyadic relationships between victims and bullies.

The empirical evidence provided in this study shows that social competency indexes, such as the Effectiveness, Responsiveness, and Reciprocity indexes, established between peers in the school context reflect functional mechanisms and facilitate a more complete description and explanation of the social relationships involved in the phenomenon of bullying in natural settings.

One limitation of the present investigation was not to include in the study the analysis of interactions in areas outside the classroom, such as the playground during games and recess activities, which can be suggested for future research.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Mexican Society of Psychology, and

ethics committee of Facultad de Psicología de la Universidad Nacional Autónoma de México, and National Council for Science and Technology, with written informed consent from school authorities in accordance to children' parents. The protocol was approved by Ethics Committee of Facultad de Psicología, Universidad Nacional Autónoma de México. Such resolution is supported by: Art. 8.03 (1) Ethical principles of psychologist and code of conduct. American Psychological Association, Effective January 1, 2017. 8. Research and publication, 8.03 (1).

AUTHOR CONTRIBUTIONS

CS was the main researcher of this work; he provided de original idea and the organization of the research group. He organized the field work and the training of the research assistants; CS and BM wrote the first draft, by revising and rewriting the contents; Both CS and BM worked on the final draft and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

REFERENCES

Atlas, R., and Pepler, D. (1998). Observations of bullying in the classroom. *J. Educ. Res.* 92, 86–99. doi: 10.1080/00220679809597580

Bakeman, R., and Gottman, J. M. (1986). *Observing Interaction: An Introduction to Sequential Analysis.* New York, NY: Cambridge University Press.

Bakeman, R., and Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences.* Cambridge: Cambridge University Press.

Beaudoin, H., and Roberge, G. (2015). Student perceptions of school climate and lived bullying behaviours. *Proc. Soc. Behav. Sci.* 174, 321–330. doi: 10.1016/j.sbspro.2015.01.667

Blanco-Villaseñor, A., Losada, J. L., and Anguera, M. T. (2003). Data analysis techniques in observational designs applied to the environment-behavior relation. *Med. Ambient. Comport. Hum.* 4, 11–126.

Cairns, R. B. (1979). *Social Development: The Origins and Plasticity of Exchanges.* San Francisco, CA: Freeman.

Cairns, R. B., and Cairns, D. B. (1994). *Lifeliness and Risks: Pathways of Youth in Our Time.* Cambridge: Cambridge University Press.

Cairns, R. B., Kagan, J., and Bergman, L. (1998). "Lawfullness in individual development, " in *Methods and Models for Studying the Individual,* eds R. B. Cairns, L. R. Bergman, and J. Kagan (Thousand Oaks, CA: Sage Publications), 1–10.

Cairns, R. B., Santoyo, V. C., Ferguson, L., and Cairns, B. (1991). Integración de información interaccional y contextual: El procedimiento de las observaciones sincrónicas. *Rev. Mex. Anál. Cond.* 17, 105–120.

Cárdenas, M., and Arancibia, H. (2014). Statistical power and effect size calculating in G*Power: complementary analysis of statistical significance testing and its application in psychology. *Salud Soc.* 5, 210–224.

Cohen, J. A. (1960). Coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Cruz, C. A. (2007). *Estudio de los Patrones Conductuales de Niños Víctimas de Agresión Ejercida por Sus Pares.* Tesis de Licenciatura en Psicología, Universidad Nacional Autónoma de México, Ciudad de México.

Cuenca, V., and Mendoza, B. (2017). Comportamiento pro-social y agresivo en niños: Tratamiento conductual dirigido a Padres y Profesores. *Rev. Acta Invest. Psicol.* 7, 2691–2703. doi: 10.1016/j.aipprr.2017.03.005

De Oliveira, W. A., Silva, M. A., da Silva, J. L., de Mello, F. C., do Prado, R. R., and Malta, D. C. (2016). Associations between the practice of bullying and individual and contextual variables from the aggressors' perspective. *J. Pediatr.* 92, 32–39. doi: 10.1016/j.jped.2015.04.003

Elder, G. H., Pavalko, E. K., and Clipp, E. C. (1993). *Working with Archival Data: Studying Lives.* Newbury Park, CA: Sage Publications.

Elgar, F., McKinnon, B., Walsh, S., Freeman, J., Donnelly, P., Gaspar, M., et al. (2015). Structural determinants of youth bullying and fighting in 79 Countries. *J. Adolesc. Health* 57, 643–650. doi: 10.1016/j.jadohealth.2015.08.007

Eslea, M., Menesini, E., Morita, Y., O'Moore, M., Mora-Merchán, J. A., Pereira, B., et al. (2003). Friendship and loneliness among bullies and victims: data from seven countries. *Aggress. Behav.* 30, 71–83. doi: 10.1002/ab.20006

Espinosa, M. C. (2017). "Notas sobre el desarrollo de las preferencias sociales en niños de uno a seis años," in *Mecanismos Básicos de Toma de Decisiones:*

Perspectivas Desde las Ciencias del Comp Ortamiento y del Desarrollo, coord. C. Santoyo (México: UNAM/CONACYT 178383), 217–258.

Espinosa, M. C., Blanco-Villaseñor, A., Santoyo, C. (2006). "El estudio del comportamiento en escenarios naturales: El sistema de observación de las interacciones sociales," in *Desarrollo e Interacción Social: Teoría y Métodos de Investigación en Contexto,* eds C. Santoyo and M. C. Espinosa Arámburu (México: CONACYT/UNAM), 45–78.

Farmer, T. W., and Cairns, R. B. (1991). Social Networks and social status in emotionally disturbed children. *Behav. Disord.* 16, 288–298.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions.* New York, NY: Wiley.

Landau, S., and Swerdlik, M. (2005). Commentary: What you see is what you get: a commentary on school-based direct observation systems. *Sch. Psychol. Rev.* 34, 529–536.

Lereya, S. T., Copeland, W. E., Costello, E. J., and Wolke, D. (2015). Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries. *Lancet Psychol.* 2, 524–531. doi: 10.1016/S2215-0366(15)00165-0

López, C. E., and Santoyo, V. C. (2004). Asymmetry of the conflicting interaction of violent spouses: the test of a model. *Mex. J. Behav. Anal.* 30, 115–138. Available online at: http://rmac-mx.org/asimetria-de-la-interaccion-conflictiva-de-conyuges-violentos-la-prueba-de-un-modelo/

Machado, C., Madalena, A. E., Leite, C., Bertazzi, R., and Rossi, P. (2015). School bullying: a systematic review of contextual-level risk factors in observational studies. *Aggress. Viol. Behav.* 22, 65–76. doi: 10.1016/j.avb.2015.04.006

Mendoza, B. (2011). Bullying entre pares y el escalamiento de agresión en la relación profesor-alumno. *Psicol. Iberoam.* 19, 58–71.

Mendoza, B. (2013). "Tipos de adolescentes con respecto a la violencia escolar," in *Aristas y Perspectivas Múltiples de la Investigación Sobre Desarrollo e Interacción Social,* ed C. Santoyo (México: UNAM), 107–132.

Mendoza, B., and Maldonado, V. (2017). Acoso escolar y habilidades sociales en alumnado de educación básica. *Ciencia Ergo Sum* 24, 109–116. Available online at: https://cienciaergosum.uaemex.mx/article/view/7735

Modecki, K., Minchin, J., Harbaugh, A., Guerra, N., and Runions, K. (2014). Bullying prevalence across contexts: a meta-analysis measuring cyber and traditional bullying. *J. Adolesc. Health* 55, 602–611. doi: 10.1016/j.jadohealth.2014.06.007

O'Moore, A. M., Kirkham, C., and Smith, M. (1997). Bullying behavior in Irish schools: a nation wide study. *Irish J. Psychol.* 18, 141–169.

Olweus, D. (1993). *Bullying at School: What We Know and What We Can do.* Cambridge, MA: Blackwell Publishers.

Olweus, D. (2001). "Peer harassment: a critical analysis and some important issues," in *Peer Harassment in Schools: The Plight of the Vulnerable and Victimized,* eds J. Juvonen and S. Graham (New York, NY: The Guilford Press), 3–20.

Patterson, G. R. (1979). "A performance theory for coercitive family interaction," in *The Analysis of Social Interactions: Methods, Issues and Illustrations,* ed R. B. Cairns (Hilladale, NJ: Lawrence Erlbaum Associates), 119–162.

Patterson, G. (1982). *Coercive Family Process: A Social Learning Approach. Vol. 3.* Eugene Oregon, OR: Castalia Publishing Co.

Salmivalli, C., Lagerspetz, K. M. J., Björkqvist, K., Österman, K., and Kaukiainen, A. (1996). Bullying as a group process: participant roles and their relations to social status within the group. *Aggress. Behav.* 22, 1–15. doi: 10.1002/(SICI)1098-2337(1996)22:1<1::AID-AB1>3.0.CO;2-T

Santoyo, V. C. (1996). Behavioral assessment of social interactions in natural settings. *Eur. J. Psychol. Assess.* 12, 124–131. doi: 10.1027/1015-5759.12.2.124

Santoyo, V. C. (2007). *Estabilidad y Cambio de Patrones de Comportamiento en Escenarios Naturales: Un Estudio Longitudinal en Coyoacán.* México: UNAM/CONACYT 40242H.

Santoyo, V. C., Anguera, M. T., Colmenares, V. L., and Figueroa, B. N. (2007a). "Estudio longitudinal de patrones conductuales: Micro-análisis bajo una perspectiva de síntesis," in *Estabilidad y Cambio de Patrones de Comportamiento*

en Escenarios Naturales: Un Estudio Longitudinal en Coyoacán, ed C. Santoyo (México: UNAM/CONACYT 40242H), 181–213.

Santoyo, V. C., and Colmenares, V. L. (2012). "Investigación puente y de archivo: implicaciones para el estudio longitudinal de Coyoacán," in *Aristas y Perspectivas Múltiples de la Investigación Sobre Desarrollo e Interacción Social,* coord. C. Santoyo (México: UNAM/CONACYT 57327), 43–74.

Santoyo, V. C., Colmenares, V. L., and Figueroa, N. (2007b). "Estabilidad y cambio de patrones coercitivos: un enfoque de síntesis," in *Estabilidad y Cambio de Patrones de Comportamiento en Escenarios Naturales: Un Estudio Longitudinal en Coyoacán,* ed C. Santoyo (México: UNAM), 114–148.

Santoyo, V. C., Colmenares, V. L., Figueroa, N., Cruz, A., and López, E. (2008). Organización del comportamiento coercitivo en niños de primaria: un enfoque de síntesis. *Rev. Mex. Psicol.* 25, 71–87.

Santoyo, V. C., Colmenares, V. L., and Losada, J. L. (2007c). "Redes sociales: Modalidades, funcionalidad y continuidad en una disciplina de síntesis," in *Estabilidad y Cambio de Patrones de Comportamiento en Escenarios Naturales: Un Estudio Longitudinal en Coyoacán México,* ed C. Santoyo (México: UNAM/CONACYT 40242H), 81–113.

Santoyo, V. C., Colmenares, V. L., and Xicoténcatl, L. A. N. (2017). "Patrones de comportamiento extendidos en el tiempo: una estrategia integrativa," *Mecanismos Básicos de Toma de Decisiones: Perspectivas Desde las Ciencias del Comportamiento y del Desarrollo*, coord. C. Santoyo (México: CONACYT 178383/UNAM), 117–146.

Santoyo, V. C., and Espinosa, A. M. C. (2005). *Desarrollo e Interacción Social: Teoría y Métodos de Investigación y Contexto.* México: UNAM/CONACYT 40242H.

Santoyo, V. C., Espinosa, M. C., and Bachá, M. G. (1994). Extensión del sistema de observación conductual de las interacciones sociales: calidad, dirección, contenido, contexto y resolución. *Rev. Mex. Psicol.* 11, 55–68.

Santoyo, V. C., Espinosa, M. C., and Maciel, A. O. (1996). Reciprocidad y patrones coercitivos de niños preescolares. *Rev. Mex. Psicol.* 13, 63–74.

Santoyo, V. C., López, C. E., and Colmenares, V. L. (2006). "Estrategias para el estudio de perfiles y patrones de interacción social," in *Desarrollo e Interacción Social: Teoría y Métodos de Investigación en Contexto,* eds C. Santoyo and C. Espinosa (México: CONACYT 40242H/UNAM), 213–242.

Sidman, M. (1989). *Coercion and its Fallout.* Boston, MA: Authors cooperative.

Turunen, T., Poskiparta, E., and Salmivalli, C. (2017). Are reading difficulties associated with bullying involvement? *Learn. Instr.* 52, 130–138. doi: 10.1016/j.learninstruc.2017.05.007

Whitney, I., and Smith, P. K. (1993). A survey of the nature and extent of bullying in junior/middle and secondary school. *Educ. Res.* 35, 3–25. doi: 10.1080/0013188930350101

Wilson, J. W., and Herrnstein, R. J. (1985). *Crime and Human Nature.* New York, NY: Simon & Schuster, Inc.

Yüksel-Sahina, F. (2015). An examination of bullying tendencies and bullying coping: behaviors among adolescents. *Soc. Behav. Sci.* 191, 214–221. doi: 10.1016/j.sbspro.2015.04.415

T-Pattern Analysis and Cognitive Load Manipulation to Detect Low-Stake Lies: An Exploratory Study

Barbara Diana[1], Valentino Zurloni[1]*, Massimiliano Elia[1], Cesare Cavalera[2], Olivia Realdon[1], Gudberg K. Jonsson[3] and M. Teresa Anguera[4]

[1] Department of Human Sciences for Education, University of Milano-Bicocca, Milan, Italy, [2] Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy, [3] Human Behavior Laboratory, University of Iceland, Reykjavik, Iceland, [4] Faculty of Psychology, Institute of Neurosciences, University of Barcelona, Barcelona, Spain

Deception has evolved to become a fundamental aspect of human interaction. Despite the prolonged efforts in many disciplines, there has been no definite finding of a univocally "deceptive" signal. This work proposes an approach to deception detection combining cognitive load manipulation and T-pattern methodology with the objective of: (a) testing the efficacy of dual task-procedure in enhancing differences between truth tellers and liars in a low-stakes situation; (b) exploring the efficacy of T-pattern methodology in discriminating truthful reports from deceitful ones in a low-stakes situation; (c) setting the experimental design and procedure for following research. We manipulated cognitive load to enhance differences between truth tellers and liars, because of the low-stakes lies involved in our experiment. We conducted an experimental study with a convenience sample of 40 students. We carried out a first analysis on the behaviors' frequencies coded through the observation software, using SPSS (22). The aim was to describe shape and characteristics of behavior's distributions and explore differences between groups. Datasets were then analyzed with Theme 6.0 software which detects repeated patterns (T-patterns) of coded events (non-verbal behaviors) that regularly or irregularly occur within a period of observation. A descriptive analysis on T-pattern frequencies was carried out to explore differences between groups. An in-depth analysis on more complex patterns was performed to get qualitative information on the behavior structure expressed by the participants. Results show that the dual-task procedure enhances differences observed between liars and truth tellers with T-pattern methodology; moreover, T-pattern detection reveals a higher variety and complexity of behavior in truth tellers than in liars. These findings support the combination of cognitive load manipulation and T-pattern methodology for deception detection in low-stakes situations, suggesting the testing of directional hypothesis on a larger probabilistic sample of population.

Keywords: deception detection, cognitive load manipulation, kinesics, analysis of observational data, T-patterns

INTRODUCTION

Deception is a matter of everyday life, as several studies have underlined. Turner et al. (1975) estimated that 62% of the statements in everyday general conversations could be somehow classified as deceptive. DePaulo et al. (1996) employed a 1-week diary study to record people's everyday communication, specifically deceptive communication. Their results suggest that people tell approximately two lies per day on average and that approximately 20 to 33% of our daily interactions are deceptive. Serota et al. (2010) report an average of 1.65 lies in a 24-h period. These data have been supported in two other empirical studies (Hancock et al., 2004; George and Robb, 2008). With a similar methodology, Hancock et al. (2004) observed 26% of our everyday communication to be involving some form of deception, while George and Robb (2008) estimated that 22–25% of our daily communication might be deceptive.

Even though percentages slightly change among different studies, we can affirm deception is a ubiquitous phenomenon which has evolved to become a fundamental aspect of human interaction (O'Sullivan, 2003; Trivers, 2011). Despite prolonged efforts across a broad array of contexts and disciplines, a diagnostic cue to deception has not been found yet (DePaulo et al., 2003; Sporer and Schwandt, 2007; Vrij, 2008).

Two cognitive lie detection approaches emerge from the literature, both relying on the classic Cognitive Capacity Theory (Kahneman, 1973) and adapted from the Cognitive Load Theory (Van Merrienboer and Sweller, 2005). The "mere cognitive load approach" and the "imposing cognitive load approach."

The first approach ("mere cognitive load"), assumes that the act of lying itself generates observable signs of cognitive load (intrinsic cognitive load). This is also known as the traditional cognitive lie detection approach, based on the work by Zuckerman et al. (1981). Several authors agree in stating that some aspects of lying contribute to the increased mental load (e.g., DePaulo and Kirkendol, 1989; Buller and Burgoon, 1996; DePaulo et al., 2003; Kassin and Norwick, 2004; Kassin, 2005; Vrij et al., 2006; Walczyk et al., 2013).

Lying is not always more cognitively demanding than truth telling (McCornack, 1997). Differences between liars and truth-tellers may be relatively small, and perhaps not readily discernable by observers (Zuckerman et al., 1981; DePaulo et al., 2003), especially when considering low-stake lies. Low-stake lies include pedagogic or white lies, day-to-day polite lies (DePaulo et al., 1980), as well as different kinds of concealment, omission, and evasive messages. On these occasions deceivers and truth-tellers are assumed either to have little to gain (or lose) by being judged deceptive by the addressee or to feel little fear of being caught telling these lies as required by a polite society (Frank and Ekman, 1997). The liar can be at ease in these contexts and does not need particular cognitive demands in generating this kind of deceptive message (Anolli et al., 2002). Conversely, high-stakes lies may have serious effects and consequences for both the deceiver and the deceived. Generally, these kinds of lies are likely to happen in complicated relational situations and in conflicting or face-threatening contexts, such as police

interrogations, customs inspections, and high-stake poker games (Frank and Ekman, 1997; Anolli et al., 2002). The speaker has to face up to high cognitive demands since, telling the truth or telling a lie, he has to fabricate a message with the lowest risk of penalty. Many studies focused on the cognitive design of deception in high-stake contexts (Tsiamyrtzis et al., 2007; for an overview see Vrij et al., 2017), while low-stake ones have been scarcely investigated.

The second approach ("imposing cognitive load"), is based on the manipulation of extraneous cognitive load and, for this reason, we think it is particularly suited for low-stakes lies detection. In this perspective, an additional cognitive demand is imposed on individuals to highlight the observable differences between lying and truth telling. Two variations of this method can be identified: in some studies, cognitive demand was increased by making the lie task harder (e.g., Vrij et al., 2008, 2011), while other ones adopted the dual-task paradigm (Baddeley and Hitch, 1974; Baddeley, 1992). For this study, we chose the second option.

Techniques under the heading of dual-task paradigm seek to induce cognitive load selectively on liars not making the lie task harder but rather by altering other aspects of the examination procedure or context. In dual-task paradigm experiments on deception, researchers ask subjects to carry out a secondary task while lying. Because of the additional resources needed for fabricating and telling the lie, people should find dual task more cognitively difficult when they lie than when they tell the truth, and as a result, they should perform worse when they lie. A powerful framework for understanding multi-task interference effect is the Adaptive Executive Control (AEC), which claims that the major sources of interference are in the competition between concurrent tasks for the same perceptual or motor response systems and the executive process performing one task before another (due to its higher priority given the performer's goals) (Meyer and Kieras, 1997; Meyer et al., 2002). The second task, therefore, will be as more effective as its execution activates the same underlying mechanisms, loading on the same systems used for creating, processing and telling the specific kind of lie being investigated.

As it happens in standard communication, liars are able to arrange a set of different signaling systems to communicate and make their communicative intentions effective, like language, the paralinguistic system, the face and gestures system, gaze, proxemics and the haptic system, or the chronemic system.

Since no diagnostic cue to deception occurs, it could be that a diagnostic pattern does arise when a combination of cues is taken into account (Vrij, 2008). Several studies showed that multimodal data collection could be effective in deception detection. Vrij (2008) claims that, with a combination of four different variables (illustrators, hesitations, latency period, and hand/finger movements) he was able to classify correctly 84.6% of liars and 70.6% of true tellers (Vrij et al., 2000). Jensen et al. (2010), focused on cues extracted from audio, video and textual data, with the aim of building a paradigm for deception detection via a multi-layered model. They reached a classification accuracy of 73.3%, claiming that deception indicators are subtle,

dynamic, and transitory, and often elude a human's conscious awareness. Other studies have shown that between 71 and 78% of correct classifications were made when the researchers investigated a cluster of behaviors (Heilveil and Muehleman, 1981; Vrij et al., 2004; Davis et al., 2005). In other words, more accurate truth/lie classifications can be made if a cluster of non-verbal cues is examined rather than each of these cues are treated separately.

Of course, people can easily control only those patterns that are manifest and have a macroscopic nature, easily readable from the outside time by time. However, patterns in behavior are frequently hidden from the consciousness of those who perform them as well as to unaided observers (Magnusson, 2006). As Eibl-Eibesfeldt (1970) argued, "behavior consists of patterns in time. Investigations of behavior deal with sequences that, in contrast to bodily characteristics, are not always visible." When the order of events is the only variable considered, the main challenge is to detect the pattern without being distracted by background noise from other events. T-pattern analysis was developed by Magnusson (2000, 2005, 2006) to find temporal and sequential structure in behavior. The term T-pattern stands for temporal pattern; this approach focuses on determining whether arbitrary events sequentially occur within a specified time interval at a rate greater than that expected by chance. In this way, it detects repeated patterns of behavior units coded as events on one-dimensional discrete scales. Temporal pattern analysis and its related software THEME (Magnusson, 2000) have been applied to a great number of research experiments in very different fields. Patterns have been used to describe, interpret and understand phenomena such as deceptive communication (Anolli and Zurloni, 2009; Zurloni et al., 2013, 2016; Diana et al., 2015), animal and human behavior (Kerepesi et al., 2005; Casarrubea et al., 2015), patient–therapist communication in computer assisted environments (Riva et al., 2005), a wide variety of observational and sports studies, such as analysis of soccer team play (Camerino et al., 2012; Cavalera et al., 2015; Diana et al., 2017), motor skill responses in body movement and dance (Castañer et al., 2009) and deception detection in doping cases (Zurloni et al., 2015).

Moreover, patterns of this kind may often be hard or impossible to detect with the well-known statistical methods that are found in major statistical program packages and behavior research software, such as The Observer (Noldus, 1991; Noldus et al., 2000) or GSEQ (Bakeman and Quera, 1995).

Basing on the literature discussed above and results from previous studies (Vrij, 2008; Zurloni et al., 2013, 2016; Burgoon et al., 2014, 2015; Diana et al., 2015), this work proposes an approach to deception detection combining cognitive load manipulation and T-pattern methodology with three objectives: (a) testing the efficacy of dual task-procedure in enhancing differences between truth tellers and liars in a low-stakes situation; (b) exploring the efficacy of T-pattern methodology in discriminating truthful reports from deceitful ones in a low-stakes context; (c) setting the experimental design and procedure for future and follow-up research.

METHOD AND DATA ANALYSIS

Participants

The convenience sample was initially composed by 46 students, (50% male and 50% female), aged from 21 to 31, born in Italy and living in the same geographic area (inclusion criteria). They were volunteers, contacted through the University's online study recruiting platform. Due to the exploratory aim of this study, the recruiting lasted 2 months, until 23 males and 23 females fitting the inclusion criteria signed for the experiment. Six students never showed up for the experiment, restricting our final sample to 40.

Instructions on the recruiting platform informed candidates they would participate in an experiment involving their communicative abilities and working memory. At the beginning of the experiment, all participants signed an informed consent to both audio and video recording and authorized the use and processing of personal data; they were also explicitly informed of the possibility to withdraw from the experiment at any time. To increase motivation further, we guaranteed a restitution of results, giving information about their communicative skills.

Instruments and Materials

– *post hoc* controls and manipulation check instruments, based on cognitive abilities and possible arousal interferences for the participants involved in the experiment (anxiety levels too high or cognitive capacities too distant from the average). For this purpose, we used the State-Trait Anxiety Inventory (STAI; Spielberger, 1983) to measure trait and state anxiety at different times of the experiment. We also used a digit span test and a Corsi test (Mammarella et al., 2008) for short term memory (phonological and spatial).

– a short video clip was made for the experiment, divided in two segments, which were showed separately (see procedure for information about the trials order). The whole video is a no-dialog story about four characters organizing a bank robbery and committing it. The video is silent because we wanted to avoid having participants use words from the characters and to let them free in the interpretation and accounting of the story (no violent acts or emotionally involving events are depicted, in order to avoid arousal interference). The video was cut approximately in half, so that each segment's length would be about the same as the other and contain the same number of events.

– dual task procedure. In the experimental condition, participants were asked, on two occasions, to memorize a series of four sentences. The two series are different, balanced to contain the same number of words and the same syntactic structure. Participants were told to remember these sentences since they could be asked to recall them at any time during the experiment. This extra memory task was added to produce interference in visual-spatial memory.

– observation instrument: the behavior categories we identified refer to self-contact gestures (ADAPTOR in the coding grid), illustrator gestures (ILLUSTRATOR), rhythmic gestures (RHYTHMIC), emblems or symbols (SYMBOL), leg movements (LEG), feet movements (ONLY_FOOT), finger movements (ONLY_FINGERS). These behaviors have been chosen according to what emerged in our previous study (Zurloni et al., 2013, 2016), in other studies found in literature (Caso et al., 2006), and also according to what was affirmed by Anolli (2012) concerning the scarcity of distal movements compared to proximal ones during deception, as well as the results obtained by Burgoon et al. (2014) on kinesic T-patterns related to deception. The objective was to create a coding grid that would satisfy the need to simplify interpretation of patterns while identifying specific components in patterns that could be more relevant, when combined, than the configurations tested until now.
– Research Randomizer -Version 4.0- computer software (Urbaniak and Plous, 2013), to randomize the order of trials;
– Behavior Coder software, for manually coding non-verbal behaviors from video sources;
– SPSS 22 software, for statistical analysis of behavior frequencies and pattern data.
– Theme 6.0[1] software for T-pattern detection. This software detects the temporal structure of data sets, revealing repeated patterns (T-patterns) that regularly or irregularly occur within a period of observation. It allows the detection of repeated temporal patterns even when multiple unrelated events occur in between components of the patterns.

Procedure

This study was carried out at the University of Milano-Bicocca in an audio-isolated laboratory room equipped with four cameras, set to video-record participants' full-lengths and close-ups. The cameras were connected to a 2-channel quad device (split-screen technique).

The participants, males and females, were assigned to conditions using a procedure (control condition for the first male participant, experimental for the second and so on; the same with the female group) designed to have the same number of participants, balanced for gender, per group (as near as possible).

The whole experiment, in both conditions, lasted 40–50 min. After signing the informed consent form, all participants filled in the STAI_T inventory. After this, the experimenter administered them a digit span memory test and a Corsi test for spatial memory.

The experiment, in general, consisted of watching two segments of a video (see Instruments and Materials) and then report to a confederate the truth about the content of one segment and lie about the content of the other. The order of trials has been randomized within the two conditions with a randomization software (Research Randomizer, Urbaniak and Plous, 2013), so

that half of the participants would lie about the first video, while the other half about the second one (see **Table 1**).

In the control condition, before the first video, participants are asked to watch and pay attention to every detail, explaining that the next part of the experiment will concern that particular video. After watching the video, the examiner asks them to fill out a STAI_S inventory, to measure state anxiety (and verify possible changes in arousal provoked by watching the video or other conditions). Then, depending on the trial, they are asked to report what they saw in the video or to lie about what they saw in the video to an interlocutor who, to their knowledge, does not have prior information about its content. Participants are given a list of things they can use in their report (if they have to lie, there are examples of details that can be changed, such as the number or gender of the characters, their features, their actions, etc.). They are given 5 min alone to recall and organize their report (mentally); after this, the confederate enters the room and is introduced to the participant as a fellow participant. The examiner exits the room and starts the recording. An audio signal cues the participant to start telling his/her report to the interlocutor, who does not participate in the conversation. After this part, the examiner ends the recording, goes back to the room and restarts the same routine with the second video (second part of the story). Participants are given the same instructions as before, adding the information that the interlocutor does not know which part of their story is made up but knows that one part is. This was necessary so they would not have to justify inconsistencies with their first report, since they are expected.

In the experimental condition (cognitive load manipulated), the procedure is the same but participants have to perform a dual task; before starting each of their reports, they are given a list of 4 numbered sentences to memorize and are told that they will have to recall them (when asked by the examiner) at random times during their report. They do not know when they will be interrupted or which sentence will be asked to recall, nor how many times this will happen. To enhance interference, instructions suggest to keep recalling the sentences mentally during the whole report. After the instructions, participants are given two extra minutes to memorize the sentences. Then, like in the control condition, the interlocutor enters the room and, after an audio signal, participants start their report.

At the end of the two segments, the examiner tells participants that the experiment is over and answers any question they might have about the procedure or the study.

Data Analysis

The memory span and anxiety tests we used for exclusion criteria did not show outliers. Nonetheless, 3 participants out of the initial

[1]http://patternvision.com/

TABLE 1 | Number of participants per trial within conditions.

	Control	Experimental
Trial 1*	9	10
Trial 2**	9	9

*Trial 1: Truthful report for the first video and deceptive for the second. **Trial 2: Deceptive report for the first video and truthful for the second.*

40 had to be excluded for technical problems with the recordings, restricting our sample to 37 participants and 74 reports (data from the 37 participants included in the analysis were available for all the measures considered, see **Table 1**). The videos were coded on Behavior Coder software by two coders, using a blind coding procedure. The occurrences of each event-type within the selected observation period form the so called T-dataset'.

To assess inter-rater reliability of the T-dataset, Cohen's Kappa was calculated on 10% of the encodings. Although differing through categories, inter-coder reliability was found to be good to satisfactory (ranging from 0.78 to 0.90; $p < 0.05$). When disagreements were identified or the agreement was not perfect, the specific cases were discussed and agreed on by both coders.

Single Cues

We carried out a descriptive analysis of the behaviors' frequencies coded through Behavior Coder, using SPSS (22). The aim was to show shape and characteristics of the distributions. Next, we carried out Mann-Whitney and Wilcoxon Signed Rank Tests, as a guideline for interpreting data and exploring differences between groups.

T-Pattern Detection

T-datasets were then analyzed with Theme 6.0[2] for T-pattern detection. A T-pattern is essentially a combination of events where the events occur in the same order, with the consecutive time distances between consecutive pattern components remaining relatively invariant, regardless of the occurrence of any unrelated event in between them (Magnusson, 2005).

---

[2]http://patternvision.com/

The 74 datasets were analyzed with THEME software to search for patterns and describe behavior structure and complexity (number of unique T-patterns, mean T-patterns' lengths and levels) in the truth and deception data, exploring differences between groups in the control and experimental conditions. The software allows to set statistical parameters for the pattern detection, according to research aims and scope. For this study, the threshold pattern significance was set to $p < 0.005$ and the minimum number of pattern occurrences was set to 2 (chosen based on mean length of the observation period, 2 min; Zurloni et al., 2015).

A descriptive analysis on T-pattern frequencies was carried out to show shape and characteristics of the distributions; Mann–Whitney and Wilcoxon Signed Rank Tests as a guideline for interpreting data and exploring differences between groups.

An in-depth analysis on more complex patterns was performed to get qualitative information on the behavior structure expressed by the participants. We chose to consider the more complex patterns because they represent the highest level of organization expressed by the participants in the two conditions.

RESULTS

Single Cues

Table 2 shows descriptive statistics for the difference between the deceptive and truthful reports for the single cues within the control condition. **Table 3** shows the corresponding statistics within the experimental condition. **Table 4** and **Figure 1** show descriptive statistics of single cues in truthful and deceptive reports within the control condition, while

TABLE 2 | Descriptive statistics for the difference between the deceptive and truthful reports for the single cues within the control condition.

Control condition	Mean	Mean standard error	Median	Standard deviation	Minimum	Maximum	Int.Q range	First quartile	Third quartile	Skewness	Kurtosis
ADAPTOR_LT	1.67	6.34	−1.50	26.91	−79.00	51.00	130.00	−8.50	22.25	−1.21	4.38
ILLUSTRATOR_LT	−6.78	3.54	−7.00	15.02	−26.00	32.00	58.00	−16.75	0.25	0.97	1.43
LEG_LT	0.17	1.22	−0.50	5.16	−9.00	10.00	19.00	−4.00	3.75	0.21	−0.56
FINGER_LT	1.44	0.94	0.50	3.99	−5.00	10.00	15.00	−1.25	5.00	0.36	−0.29
FEET_LT	−1.33	1.92	−1.50	8.15	−21.00	21.00	42.00	−4.25	0.25	0.51	4.26
RHYTHMIC_LT	3.33	1.69	2.50	7.19	−7.00	18.00	25.00	−2.00	7.25	0.65	−0.22
SYMBOLIC_LT	−0.39	0.51	0.00	2.15	−5.00	4.00	9.00	−1.25	1.00	−0.39	0.71

TABLE 3 | Descriptive statistics for the difference between the deceptive and truthful reports for the single cues within the experimental condition.

Experimental condition	Mean	Mean standard error	Median	Standard deviation	Minimum	Maximum	Int.Q range	First quartile	Third quartile	Skewness	Kurtosis
ADAPTOR_LT	4.10	3.22	4.00	14.03	−35.00	24.00	59.00	−5.00	16.00	−0.99	2.03
ILLUSTRATOR_LT	−3.47	3.33	2.00	14.53	−36.00	16.00	52.00	−17.00	7.00	−0.88	−0.14
LEG_LT	0.63	2.02	0.00	8.83	−24.00	15.00	39.00	−1.00	5.00	−1.01	2.67
FINGER_LT	−0.79	1.00	0.00	4.37	−8.00	7.00	15.00	−5.00	2.00	0.09	−0.48
FEET_LT	−0.26	0.75	0.00	3.28	−8.00	7.00	15.00	0.00	1.00	−0.57	1.86
RHYTHMIC_LT	4.37	1.92	3.00	8.35	−8.00	23.00	31.00	−4.00	11.00	0.26	−0.38
SYMBOLIC_LT	−0.21	0.36	0.00	1.58	−4.00	2.00	6.00	−1.00	1.00	−1.39	1.98

TABLE 4 | Descriptive statistics for the single cues in truthful and deceptive reports within the control condition.

Control condition	Mean	Mean standard error	Median	Standard deviation	Minimum	Maximum	Int.Q range	First quartile	Third Quartile	Skewness	Kurtosis
ADAPTOR_LIE	27.61	4.75	18.50	20.17	2.00	78.00	29.75	14.50	44.25	1.05	0.65
ADAPTOR_TRUTH	25.94	5.25	21.50	22.26	6.00	101.00	18.50	10.50	29.00	2.51	7.64
ILLUSTRATOR_LIE	20.94	1.95	20.00	8.29	2.00	36.00	10.00	16.50	26.50	−0.07	0.81
ILLUSTRATOR_TRUTH	27.72	3.22	27.50	13.66	4.00	54.00	24.25	16.00	40.25	−0.02	−0.54
LEG_LIE	9.39	1.14	9.50	4.85	3.00	22.00	7.50	4.75	12.25	0.80	1.14
LEG_TRUTH	9.22	1.11	8.50	4.72	3.00	19.00	7.50	5.00	12.50	0.74	−0.30
FINGER_LIE	6.44	1.01	6.00	4.29	1.00	19.00	5.25	3.00	8.25	1.30	3.41
FINGER_TRUTH	5.00	0.85	4.00	3.60	1.00	13.00	6.25	2.00	8.25	0.88	−0.13
FEET_LIE	5.39	2.25	1.50	9.55	0.00	33.00	5.00	0.00	5.00	2.39	4.95
FEET_TRUTH	6.72	1.53	5.00	6.48	0.00	22.00	10.25	1.75	12.00	1.12	0.60
RHYTHMIC_LIE	17.22	2.17	16.50	9.19	6.00	41.00	13.75	8.75	22.50	0.95	1.05
RHYTHMIC_TRUTH	13.89	1.62	12.50	6.88	4.00	25.00	11.50	8.50	20.00	0.31	−1.00
SYMBOLIC_LIE	1.67	0.39	1.00	1.64	0.00	6.00	1.25	0.75	2.00	1.50	2.24
SYMBOLIC_TRUTH	2.05	0.49	1.00	2.10	0.00	7.00	2.25	0.75	3.00	1.20	0.68

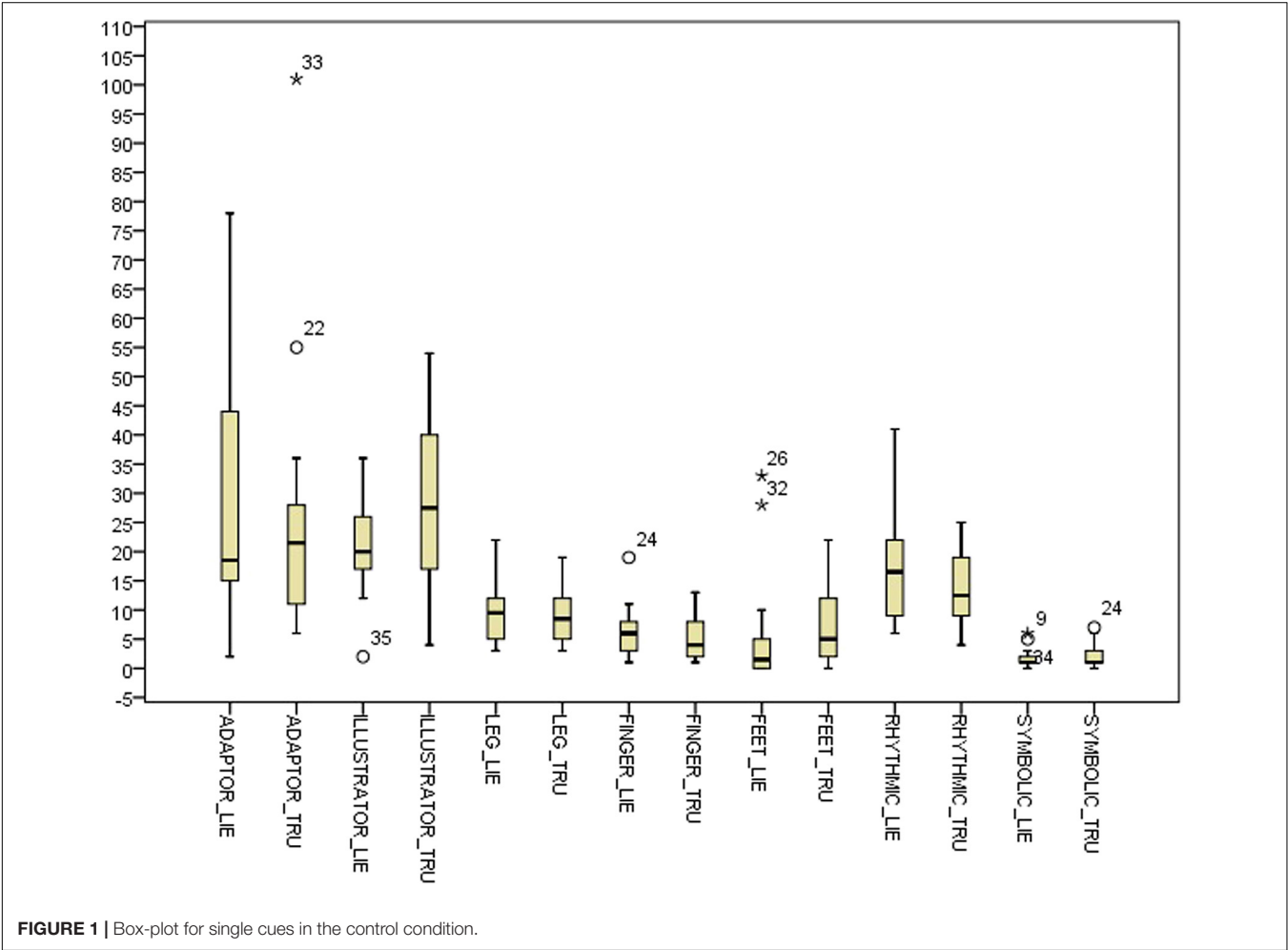FIGURE 1 | Box-plot for single cues in the control condition.

Table 5 and Figure 2 show descriptive statistics of single cues in truthful and deceptive reports within the experimental condition.

Results highlight some differences between conditions, especially regarding the distribution of ADAPTOR gestures. In deception data, the median increases from 18.50 in the

TABLE 5 | Descriptive statistics for the single cues in truthful and deceptive reports within the experimental condition.

Experimental condition	Mean	Mean standard error	Median	Standard deviation	Minimum	Maximum	Int.Q range	First quartile	Third quartile	Skewness	Kurtosis
ADAPTOR_LIE	31.16	4.49	33.00	19.56	2.00	75.00	15.00	22.00	37.00	0.62	0.85
ADAPTOR_TRUTH	27.05	4.25	28.00	18.54	0.00	78.00	21.00	13.00	34.00	0.95	1.92
ILLUSTRATOR_LIE	21.50	2.81	23.00	12.25	2.00	40.00	21.00	11.00	32.00	−0.10	−1.29
ILLUSTRATOR_TRUTH	24.95	4.17	24.00	18.20	0.00	76.00	23.00	10.00	33.00	1.10	2.26
LEG_LIE	9.84	1.83	10.00	7.99	0.00	23.00	12.00	3.00	15.00	0.49	−1.11
LEG_TRUTH	9.21	2.00	8.00	8.73	0.00	38.00	9.00	3.00	12.00	2.09	6.02
FINGER_LIE	4.16	0.79	3.00	3.47	0.00	11.00	7.00	1.00	8.00	0.57	−0.93
FINGER_TRUTH	4.95	0.78	5.00	3.41	0.00	12.00	6.00	2.00	8.00	0.24	−0.69
FEET_LIE	3.89	1.37	1.00	6.00	0.00	18.00	5.00	0.00	5.00	1.78	1.94
FEET_TRUTH	4.16	1.42	1.00	6.18	0.00	19.00	8.00	0.00	8.00	1.53	1.28
RHYTHMIC_LIE	18.95	2.89	17.00	12.59	0.00	51.00	14.00	10.00	24.00	0.92	1.01
RHYTHMIC_TRUTH	14.58	2.22	14.00	9.68	0.00	29.00	17.00	6.00	23.00	−0.05	−1.45
SYMBOLIC_LIE	1.16	0.29	1.00	1.26	0.00	4.00	2.00	0.00	2.00	1.16	0.91
SYMBOLIC_TRUTH	1.37	0.36	1.00	1.57	0.00	5.00	3 .00	0.00	3.00	0.94	−0.11
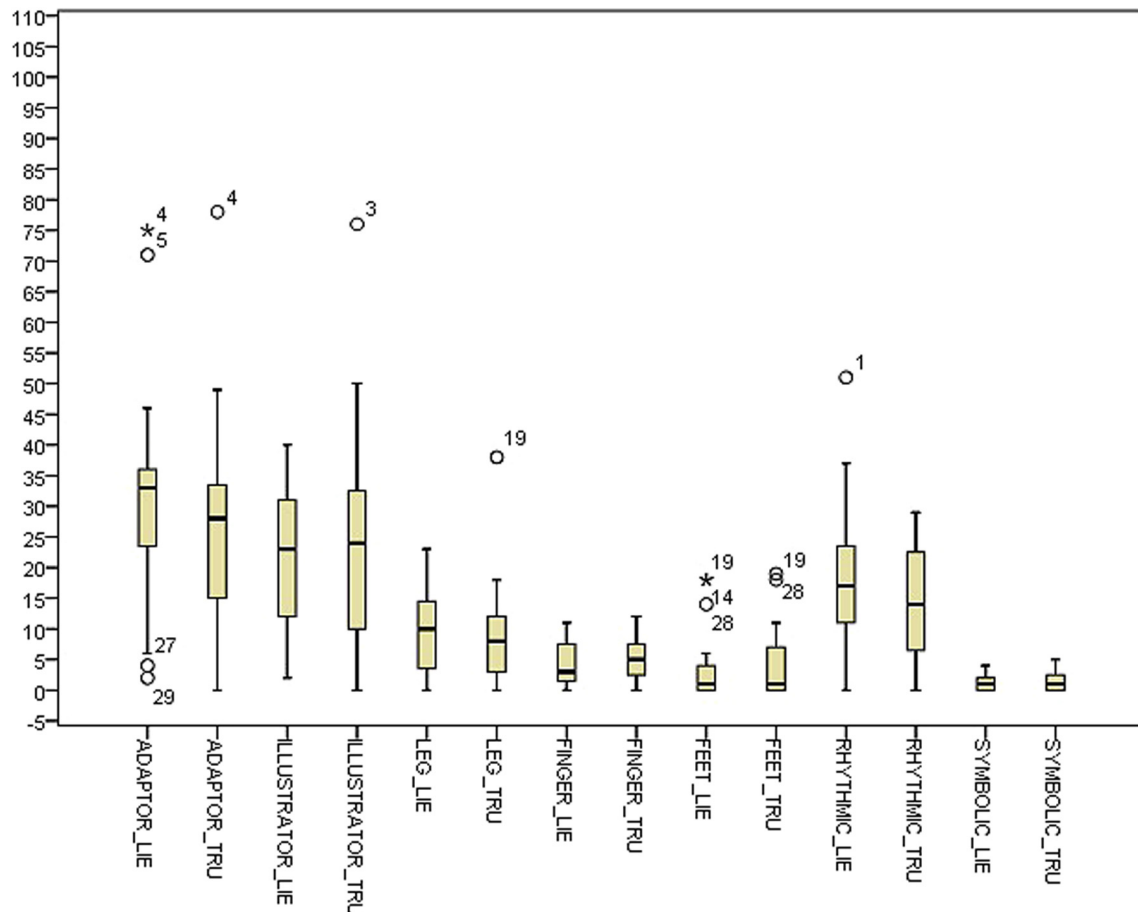
FIGURE 2 | Box-plot for single cues in the experimental condition.

control condition to 33.00 in the experimental one, while the data dispersion decreases (inter-quartile range: 29.75 in the control condition to 15.00 in the experimental one). Considering data distribution within the control condition, ILLUSTRATOR gestures were more frequently observed in truthful reports (median 27.50) rather than deceptive ones (median 20.00).

In the experimental condition, RHYTHMIC gestures were more frequent in deceptive reports (median 17.00), rather than in truthful ones (median 14.00). Independent samples Mann–Whitney *U* Test seems to show no differences in distributions across the two conditions for all the considered indexes. Differences between truth tellers and liars were explored within groups using related samples Wilcoxon Signed Rank Test (Wilcoxon, 1945). Results in the control condition seem to show no differences except for the ILLUSTRATOR behavior ($p = 0.047$), which appeared to be less present in deception than in truthful reports (see **Table 5**).

Results in the experimental condition show a difference in the distributions of RHYTHMIC gestures ($p = 0.033$), more present in deceptive reports than in sincere ones (see **Table 5**).

T-Pattern Analysis

Descriptive statistics and distributions of T-patterns in the control condition are presented in **Table 6**, while the experimental condition data are presented in **Table 7**.

Descriptive statistics and distributions of T-patterns of truthful and deceptive reports in the control condition are presented in **Table 8** and **Figure 3**, while the experimental condition data are presented in **Table 9** and **Figure 4**. Independent samples Mann–Whitney *U* Test suggest a difference in the distributions of unique patterns between the two conditions ($p = 0.026$). In fact, while the control condition shows no differences between truthful and deceptive data (in number of unique T-patterns, mean T-patterns' lengths and levels), the experimental condition shows a difference in terms of unique patterns between truth tellers and liars (Wilcoxon's test, $p = 0.036$). In particular, the number of unique patterns is substantially higher in truthful reports than in deceptive ones with a less dispersion in data distribution (see **Table 9**).

The most distinctive patterns for both conditions have also been qualitatively analyzed. In the control condition, there are few quantifiable differences in detected T-patterns, but it is possible to notice a trend for half of the participants, showing less

complex patterns in deceitful accounts and more complex ones in truthful accounts. An example is shown in **Figure 5**[3], relative to a sincere report in control conditions: it is a complex pattern, characterized by 6 levels, 11 event-types for its length, compared to an average in control conditions of 1.6 levels ($SD = 0.5$) and 2.86 ($SD = 0.76$) event-types. It is composed of rhythmic, illustrator gestures, feet and leg movements. It occurs twice during the whole observation period. Some "blocks" (or sub-patterns) included in this T-pattern are also identified by the software (singularly) and appear next to the complete pattern, occurring often and involving illustrator gestures, mostly.

A deception-related T-pattern for the control condition by the same participant is shown in **Figure 6**: it is a complex pattern (3 levels, 6 events, with a mean in control deceptive reports of 1.6 levels and 2.8 events), characterized by an alternation of rhythmic and self-contact gestures. It occurs twice during the observation period, toward the end, although different combinations of the same behaviors occur in earlier sections of the observed period.

In experimental conditions, the qualitative evaluation of patterns confirms the lack of richness suggested by the exploratory analysis in deception: T-patterns are generally simple; in some cases, the most complex one is only made by 2 different events. In more complex cases, repetition of gestures of the same category are found, linked in sub-patterns of this kind. Rhythmic gestures are identified in many deception-related patterns.

The truth condition shows variety in T-pattern compositions, with a general trend toward a complex and varied non-verbal

---

[3]How to read the pattern tree graph: the left box of **Figures 5–8** shows the events occurring within the pattern, listed in the order in which they occur within the pattern. The first event in the pattern appears at the top and the last at the bottom. The lower right box shows the frequency of events within the pattern, each dot means that an event has been coded. The pattern diagram (the lines connecting the dots) shows the connection between events. The number of pattern diagrams illustrates how often the pattern occurs. Sub-patterns also occur when some of the events within the pattern occur without the whole of the pattern occurring. The upper box illustrates the real-time of the pattern. The lines show the connections between events, when they take place and how much time passes between each event.

**TABLE 6 |** Descriptive statistics for T-pattern data in the control condition.

Control condition	Mean	Mean standard error	Median	Standard deviation	Minimum	Maximum	Int.Q range	First quartile	Third quartile	Skewness	Kurtosis
LENGTH_LT	5.65	0.25	5.51	1.06	4.50	8.71	1.08	4.82	5.90	1.63	3.18
LEVEL_LT	3.27	0.17	3.15	0.73	2.50	3.15	0.91	2.71	3–61	1.34	1.65
UNIQUEPAT_LT	19.83	2.62	18.00	11.12	4.00	45.00	13.50	12.00	25.50	0.96	0.34

TABLE 7 | Descriptive statistics for T-pattern data in the experimental condition.

Experimental condition	Mean	Mean standard error	Median	Standard deviation	Minimum	Maximum	Int.Q range	First quartile	Third quartile	Skewness	Kurtosis
LENGTH_LT	5.33	0.32	5.19	1.41	2.00	7.97	1.94	4.33	6.27	−0.22	0.66
LEVEL_LT	3.03	0.21	3.14	0.91	1.00	4.40	1.44	2.33	3.78	−0.36	−0.27
UNIQUEPAT_LT	17.84	4.96	14.00	21.62	1.00	100.00	13.00	6.00	19.00	3.33	12.72

Control condition	Mean	Mean standard error	Median	Standard deviation	Minimum	Maximum	Int.Q range	First quartile	Third quartile	Skewness	Kurtosis
LENGTH_LIE	2.79	0.16	2.55	0.70	2.00	4.43	1.07	2.31	3.39	0.99	0.06
LENGTH_TRUTH	2.86	0.18	2.66	0.76	2.00	5.42	0.67	2.43	3.10	2.41	7.54
LEVEL_LIE	1.62	0.12	1.40	0.50	1.00	2.78	0.71	1.31	2..02	0.93	0.13
LEVEL_TRUTH	1.65	0.11	1.56	0.47	1.00	3.15	0.47	1.38	1.86	1.87	5.49
UNIQUEPAT_LIE	9.83	1.65	9.50	7.01	2.00	29.00	8.25	4.75	13.00	1.51	2.48
UNIQUEPAT_TRUTH	10.00	1.71	8.00	7.24	2.00	26.00	7.25	5.00	12.25	1.35	0.85
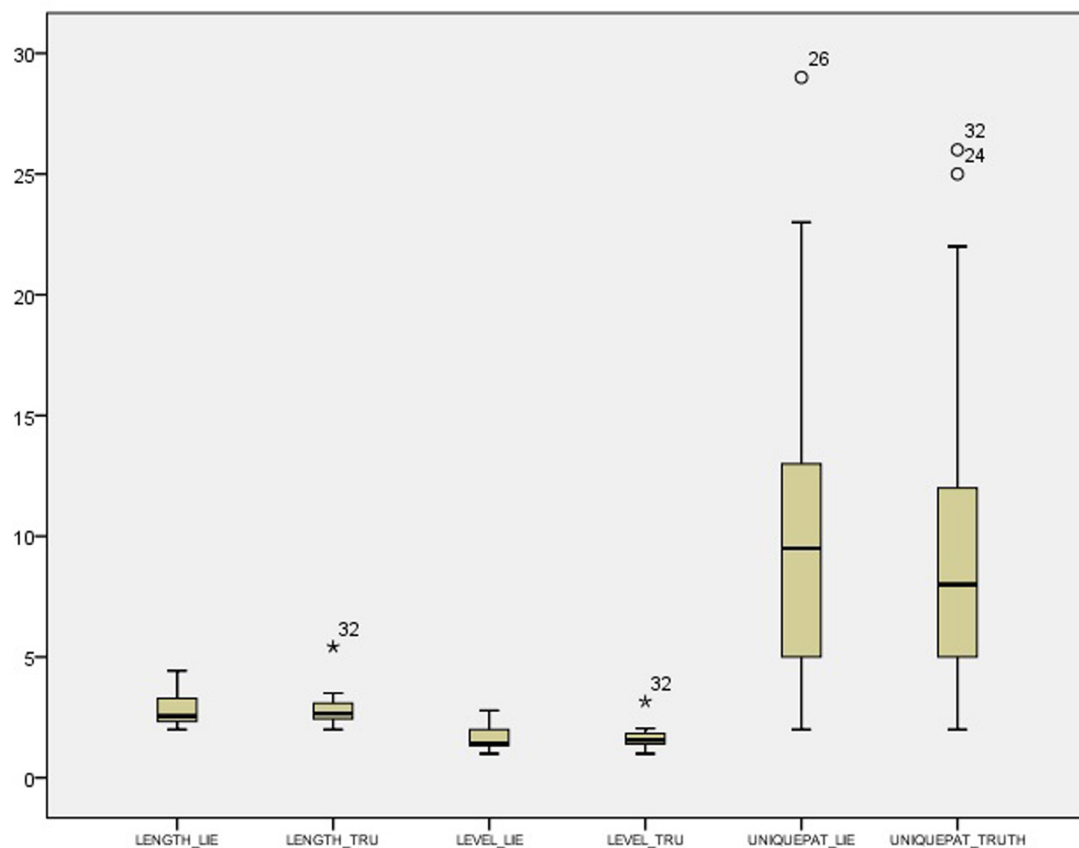
FIGURE 3 | Box-plot for T-pattern data in the control condition.

behavior, similar to the control condition patterns. An example is shown in **Figure 7**: a complex pattern, 2 levels and 4 event-types, with a general mean in its group of 1.53 levels (*SD* = 0.74) and 2.71 event-types (*SD* = 1.21). It is made of two sub-patterns including self-contact gestures, finger movements and illustrator gestures.

A T-pattern related to deception reports by the same participant is shown in **Figure 8**: it is a very simple one (1 level, 2 event-types, with a general mean of 1.5 levels and 2.6 events), being made of an alternation of self-contact and rhythmic gestures. It occurs 10 times during the observation period.

DISCUSSION AND CONCLUSION

Research on deception detection has been focused for a long time on the identification of single unmasking cues, while there have been few studies where deceptive behavior has been observed in a temporal and sequential structure perspective (Vrij, 2008; Burgoon et al., 2015). T-pattern analysis allowed to identify repeated patterns of behavior with different qualities and quantities between deception and truth.

The frequencies observed for single cues suggest that cognitive load manipulation did not affect the occurrence of specific behaviors, except for the adaptor gestures, which

TABLE 9 | Descriptive statistics for T-pattern data in truthful and deceptive reports within the experimental condition.

Experimental condition	Mean	Mean standard error	Median	Standard deviation	Minimum	Maximum	Int.Q range	First quartile	Third quartile	Skewness	Kurtosis
LENGTH_LIE	2.62	0.16	2.50	0.69	2.00	4.59	1.00	2.00	3.00	1.56	2.59
LENGTH_TRUTH	2.71	0.28	2.61	1.22	0.00	5.47	1.27	2.00	3.27	0.61	2.04
LEVEL_LIE	1.50	0.11	1.40	0.49	1.00	2.78	0.86	1.00	1.86	1.12	1.01
LEVEL_TRUTH	1.53	0.17	1.46	0.74	0.00	3.24	1.00	1.00	2.00	0.57	1.13
UNIQUEPAT_LIE	6.05	1.39	5.00	6.07	1.00	27.00	6.00	2.00	8.00	2.49	7.76
UNIQUEPAT_TRUTH	11.79	4.61	7.00	20.10	0.00	92.00	8.00	3.00	11.00	3.90	16.15
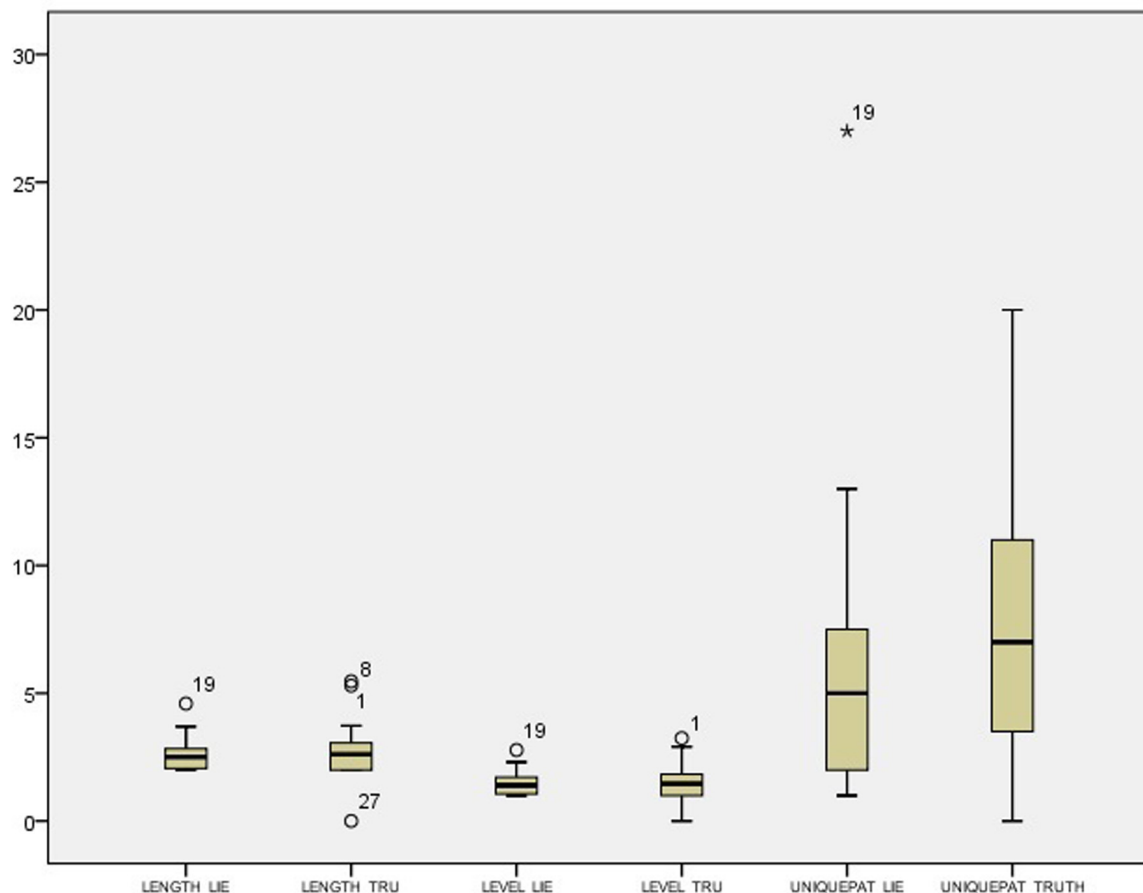
FIGURE 4 | Box-plot for T-pattern data in the experimental condition.

clearly decrease in the experimental condition, especially when lying. The decrease in data dispersion could be an effect of cognitive load manipulation. Adaptors (self-contacts and self-manipulations) are self-regulating gestures, that can increase with the increase in emotional or cognitive load (e.g., Vrij et al., 2008). Exploring differences within conditions, illustrator gestures seem to occur more in the truth excerpts of the control group (the less cognitively demanding setting). The decrease of illustrator gestures occurrences during deception has been discussed and linked to the intrinsic cognitive load increase by DePaulo et al. (2003), and by Sporer and Schwandt (2007) in their

meta-analyses; our findings seem to confirm that direction. In the experimental condition, rhythmic gestures were found to be more present in deceptive reports than in truthful ones. It is interesting to underline how this supports what was observed by Caso et al. (2006) in a study which used an experimental variable such as the rise of suspicion level, operationalized in a phase of the interview where the interlocutor directly accused the participant to be lying. The cognitive load manipulation condition used in our study could have produced the same result of the invasive interview used in the study by Caso et al. (2006). In conclusion, the behaviors identified in this analysis have already been found
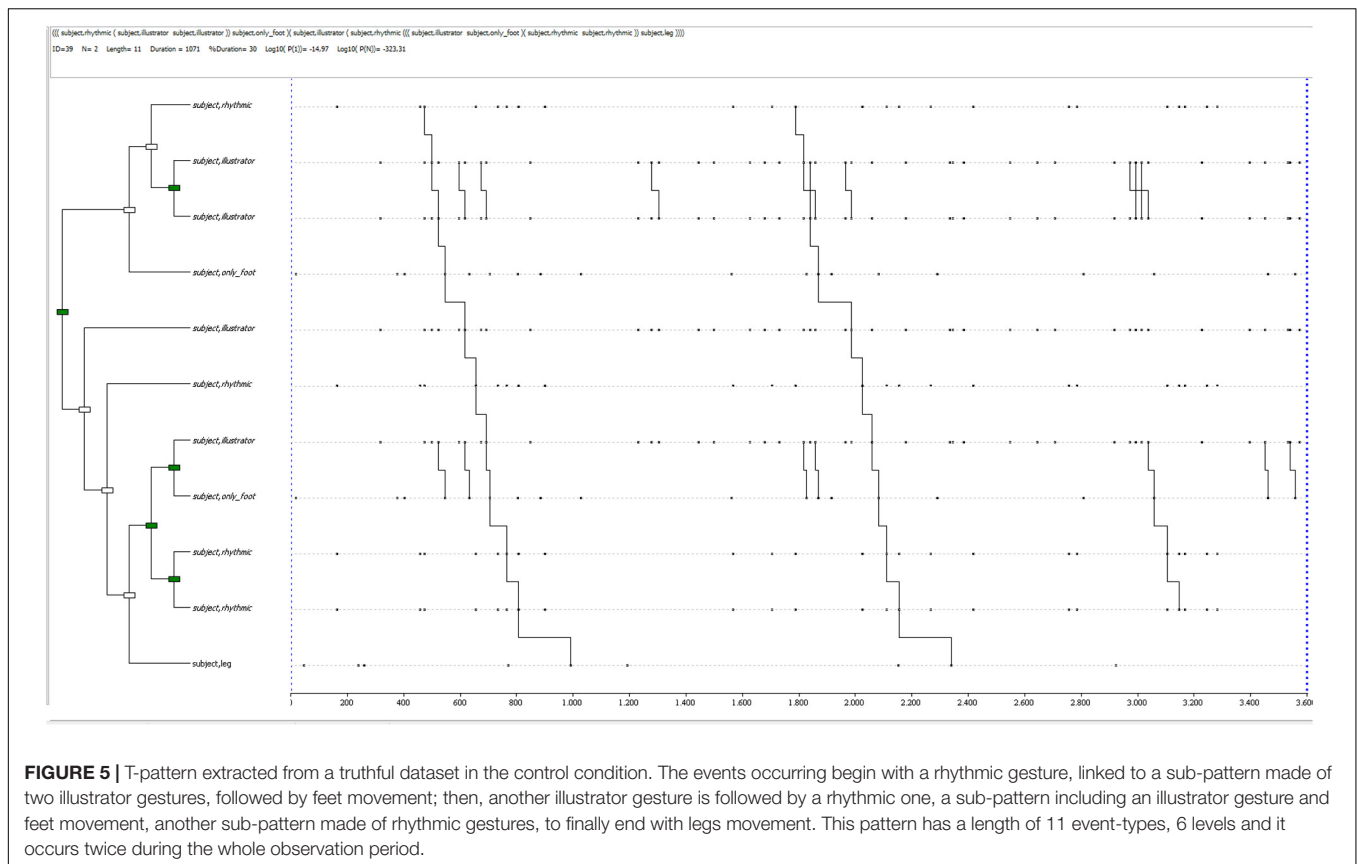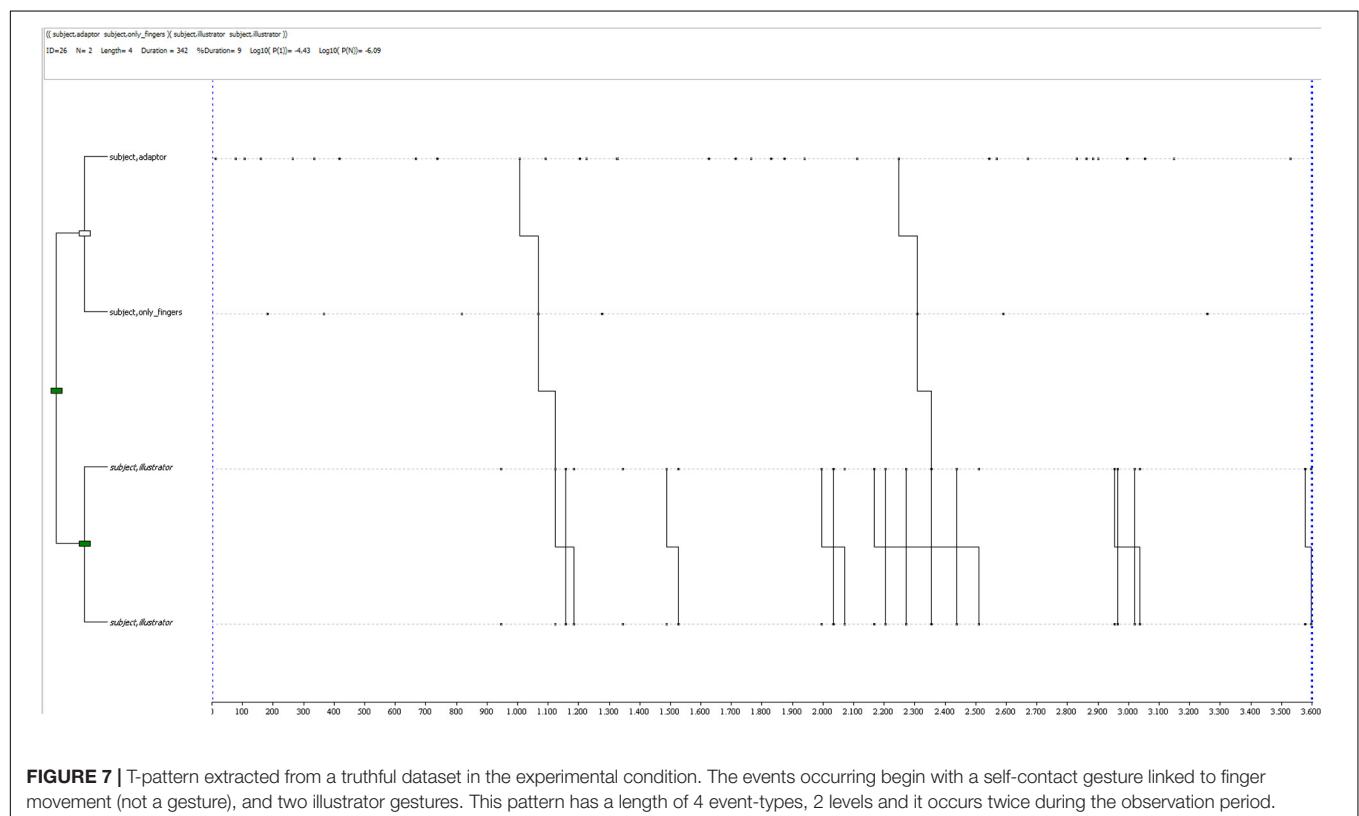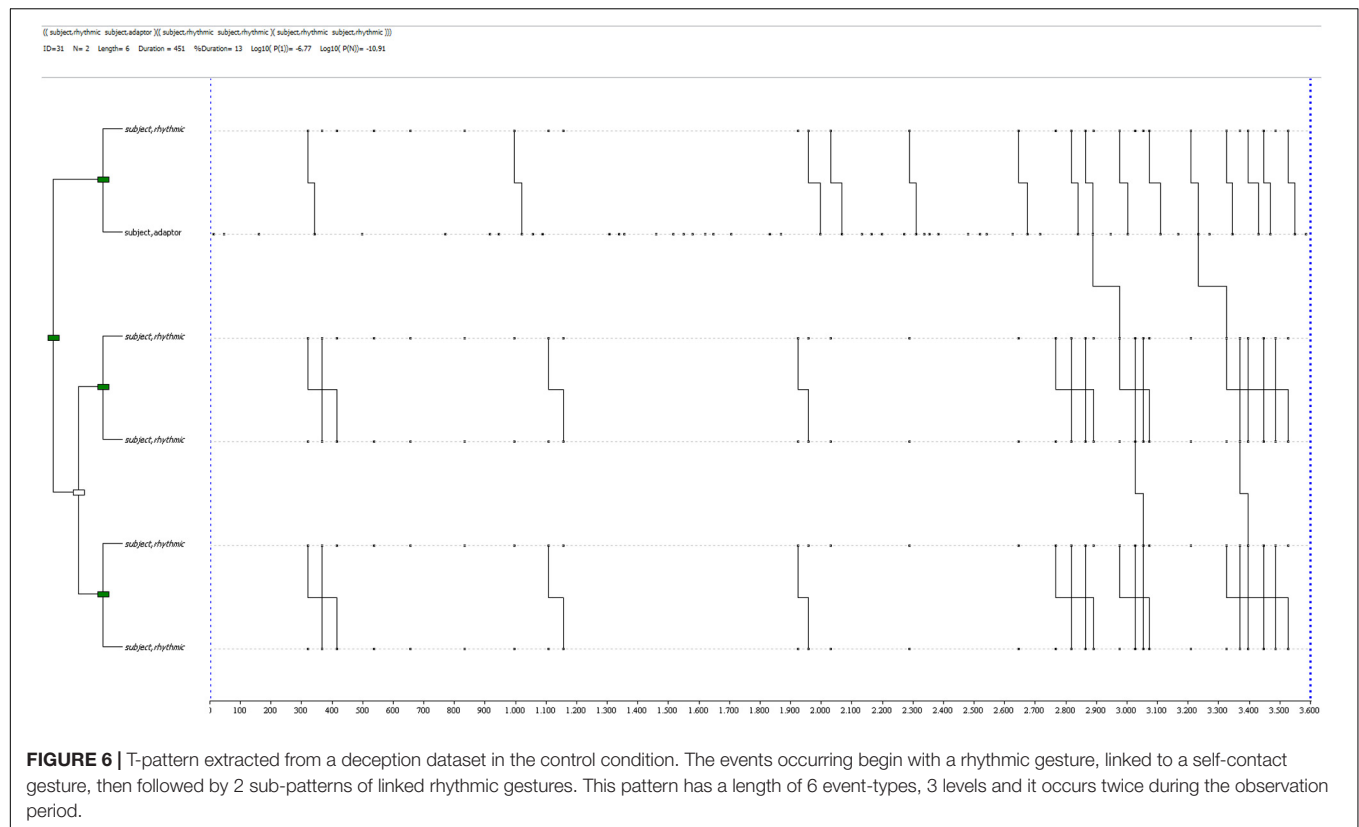
FIGURE 5 | T-pattern extracted from a truthful dataset in the control condition. The events occurring begin with a rhythmic gesture, linked to a sub-pattern made of two illustrator gestures, followed by feet movement; then, another illustrator gesture is followed by a rhythmic one, a sub-pattern including an illustrator gesture and feet movement, another sub-pattern made of rhythmic gestures, to finally end with legs movement. This pattern has a length of 11 event-types, 6 levels and it occurs twice during the whole observation period.
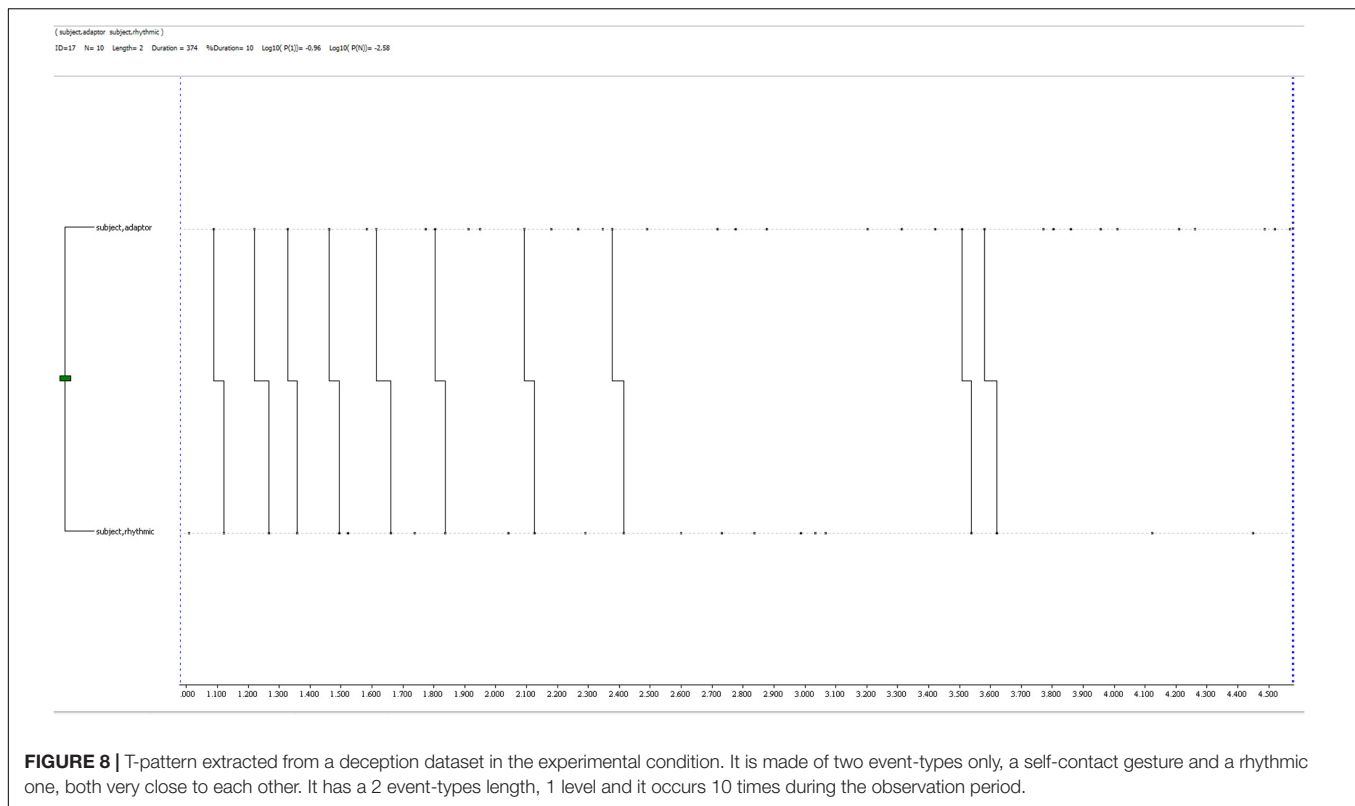
to be intrinsic of a cognitively demanding condition; in fact, these could be the effects of the increased arousal related to cognitive overload.

T-pattern frequencies show that the distribution of unique patterns seems to be affected by cognitive load manipulation. These data have to be confirmed by further studies (since the *p-value* was here used as a guideline), but it is interesting to notice that the dual-task procedure interfered on the behavior structure and variety rather than on single behavior's occurrences. As for the control condition, our data did not show any relevant difference between truth and deception in terms of unique patterns, mean lengths and levels. This result, although to be confirmed, is in accordance with what emerged from literature and supports using techniques that interfere with cognitive load; in fact, the cognitive effort demanded by "low stakes" lying could have been not enough to produce observable effects on non-verbal behavior. In the experimental condition, instead, a difference seems to exist between truth and deception in terms of unique T-patterns, with a higher number of them in truthful reports than in deceptive ones. No differences were found in the number of levels and lengths of T-patterns, although we think that this result might be affected by the length of the observation period. In general, it appears that the cognitive effort related to dual task affected non-verbal behavior variability and richness, making it more stereotypical and "mechanical" (Zurloni et al., 2015).

Qualitative in-depth analysis of detected T-patterns has shown a wide range of behaviors exhibited in patterns of all conditions. It is clear that, in control conditions, differences in non-verbal behavior between lying and truth-telling are subtle or very hard to interpret. For some participants (less than half of the sample), deception and truth-telling were characterized by similar patterns, while for others there is a decrease in complexity while lying (described by the lower mean of lengths, levels, and the number of unique patterns). These differences are not systematic enough to be ascribed to a general rule, although we do not expect them to be, since the huge amount of studies carried out in recent years produced similar results (DePaulo et al., 2003). We can speculate that the lying task might have elicited different responses because of different factors, among them a stronger or weaker inclination to deception (Anolli, 2012), or a more specific advantage or disadvantage caused by the cognitive capabilities used to lie (such as working memory, Baddeley, 1992). As suggested by findings above, in the control condition, many truth-related patterns are characterized by one or more illustrator gestures. In literature, illustrator gestures have been linked to sincerity and rhythmic gestures to deception (DePaulo et al., 2003; Caso et al., 2006). Finding these data in the structures detected by THEME contributes to support its usefulness in observational studies on deception. In the experimental condition, patterns detected by the software clearly indicate the difference between truth telling and lying conditions. In the latter one, in fact, structures of minimal complexity are

FIGURE 6 | T-pattern extracted from a deception dataset in the control condition. The events occurring begin with a rhythmic gesture, linked to a self-contact gesture, then followed by 2 sub-patterns of linked rhythmic gestures. This pattern has a length of 6 event-types, 3 levels and it occurs twice during the observation period.

FIGURE 7 | T-pattern extracted from a truthful dataset in the experimental condition. The events occurring begin with a self-contact gesture linked to finger movement (not a gesture), and two illustrator gestures. This pattern has a length of 4 event-types, 2 levels and it occurs twice during the observation period.

FIGURE 8 | T-pattern extracted from a deception dataset in the experimental condition. It is made of two event-types only, a self-contact gesture and a rhythmic one, both very close to each other. It has a 2 event-types length, 1 level and it occurs 10 times during the observation period.

detected, often as a chain of occurrences of the same kind of gesture. In many cases, patterns related to deception include rhythmic gestures. Caso et al. (2006), in their study on sincerity and deception based on the gestures' frequencies, have observed the same increase in rhythmic gestures, and we believe a cognitive load increase could explain both results. Patterns related to truth-telling in the experimental condition show a certain variety in composition, similar in its abnormal distribution of frequencies to patterns related to the control condition. These patterns are still less complex than the ones detected in the control condition, but for some participants they are similar to the first ones, including numerous illustrator gestures.

This exploratory study gave results which are in line with findings from our previous studies (Zurloni et al., 2013, 2015), and with the picture described by analyzing existent literature (DePaulo et al., 2003; Sporer and Schwandt, 2007; Vrij, 2008; Burgoon et al., 2014, 2015). Overall, THEME proves to be an effective tool for discriminating truth-telling and deceptive reports in manipulated cognitive load conditions, and the dual-task procedure seems to be effective in this sense. Differences that are not detectable in terms of single cues emerged within the structure of behavior, which, as we have discussed, seems to result less rich and more stereotypical in deception.

Limits and Future Directions

This work had its main limitations in size and characteristics of the sample but it allows us to proceed in further research, aiming at testing directional hypotheses and confirming these findings. A stratified sampling and a longer observation period

will be crucial to enhance the validity of all data extracted by THEME software. The procedure used to assign participants to conditions was meant to make the groups as equivalent as possible, despite the small sample size, but may have provoked biases according to personal characteristics of participants, for instance their motivation level (generally, the participants who enroll first to an experiment are also the most motivated). Manual coding could be an obstacle due to time-consuming practices and may have had negative effects on data quality. New technologies can help with this issue, with automatic data quality control integrated within the observation software (e.g., Castañer et al., 2017) or automatic extraction of relevant data from the source [even motor information, such as facial expressions, gestures and movement, etc. (Zhang, 2012)]. Common examples are motion capture devices like Kinect (e.g., Yu et al., 2011) wearable sensors for the extraction of biofeedback data or unobtrusive techniques like thermal imaging (Pavlidis et al., 2002; Dcosta et al., 2015). Machine learning algorithms, for example (Bartlett et al., 2005) can extract information from video or audio sources and process them through advanced algorithms that can automatically code facial expressions, body movement, typical gestures, emotions, glance direction and tone of speech. All this cues, if collected in a systematic manner, give access to a large-scale analysis, both from an observational and a statistical point of view. THEME software can work with all kind of data or events detected in a particular moment in time, making the potential applications to include a large range of sources. The experimental procedure could be improved, for example adding a naïve interlocutor and considering the interaction as moderator of behaviors.

The observation instrument built for this study could be enriched with behavior cues from other systems, such as head movements or facial expressions; although, it is fundamental to keep a balance between the exhaustivity of observed behaviors and the interpretability of results.

Impact

Deception is a ubiquitous phenomenon, regulated by the same processes used in "standard" communication. Furthermore, lying implicates an additional use of cognitive resources which, in "low stakes" conditions (low risk, low gain), common to daily life and to which humans are more "practiced," is an insignificant (or at least, not currently measurable) amount. We believe that, with cognitive approaches, a step forward was made in studying deception as a standard communication phenomenon: manipulating cognitive load allows unveiling differences otherwise inaccessible to analysis, because they are usually intrinsic to communication processes. In our experiment, THEME software was able to detect differences in behavior structure when the cognitive load experienced from participants was the highest. Although these findings need to be confirmed (a physiologic and a self-report measure of the cognitive effort could be important), it would be interesting to explore if the intrinsic higher cognitive load characterizing high-stakes deception could be enough to allow a detection without manipulation. If proved, the transferability and application of this methodology in real life contexts could be easier and potentially include several research and interventions areas, such as public security monitoring (frontiers, airports, stations, etc., Burgoon et al., 2014) or the detection of illegal and/or dangerous behaviors, for instance doping in sport (Zurloni et al., 2015).

ETHICS STATEMENT

At the time we collected data for this study, the IRB approval, in our institution, was not mandatory for minimum-risk studies. For this reason, the present study was conducted according with the ethical principles stated by the Association of Italian Psychologists and the general principles stated in The Federal Policy for the Protection of Human Subjects (or the "Common Rule"), defining "minimum" the risk in which "the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests" (45 CFR 46.102). The literature agrees in considering low-stakes deception to be a daily-life phenomenon.

AUTHOR CONTRIBUTIONS

BD and VZ contributed in method development, study designing, data analysis, and paper writing. ME contributed in data acquisition and coding, data analysis, and paper writing. CC contributed in study designing and data acquisition and coding. OR contributed in study designing and data analysis. GJ contributed in method development and data analysis. MTA contributed in method development and paper writing. All authors made suggestions and critical reviews to the initial draft and contributed to its improvement until reaching the final manuscript, which was read and approved by all authors.

ACKNOWLEDGMENTS

REFERENCES

Anolli, L. (2012). *Fondamenti di Psicologia della Comunicazione*, 2nd Edn. Bologna: Il Mulino.

Anolli, L., Balconi, M., and Ciceri, R. (2002). Deceptive miscommunication theory (DeMiT): a new model for the analysis of deceptive communication. *Emerg. Commun.* 3, 73–100.

Anolli, L., and Zurloni, V. (2009). Cooperative and non cooperative lies in close and casual relationships. *Stud. Commun. Sci.* 9, 7–25.

Baddeley, A. (1992). Working memory: the interface between memory and cognition. *J. Cogn. Neurosci.* 4, 281–288. doi: 10.1162/jocn.1992.4.3.281

Baddeley, A. D., and Hitch, G. (1974). Working memory. *Psychol. Learn. Motiv.* 8, 47–89. doi: 10.1016/S0079-7421(08)60452-1

Bakeman, R., and Quera, V. (1995). *Analyzing Interaction: Sequential Analysis with SDIS and GSEQ*. Cambridge: Cambridge University Press.

Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., and Movellan, J. (2005). "Recognizing facial expression: machine learning and application to spontaneous behavior," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2 (Piscataway, NY: IEEE), 568–573.

Buller, D. B., and Burgoon, J. K. (1996). Interpersonal deception theory. *Commun. Theory* 6, 203–242. doi: 10.1111/j.1468-2885.1996.tb00127.x

Burgoon, J. K., Proudfoot, J. G., Schuetzler, R., and Wilson, D. (2014). Patterns of nonverbal behavior associated with truth and deception: illustrations from three experiments. *J. Nonverb. Behav.* 38, 325–354. doi: 10.1007/s10919-014-0181-5

Burgoon, J. K., Schuetzler, R., and Wilson, D. W. (2015). Kinesic patterning in deceptive and truthful interactions. *J. Nonverb. Behav.* 39, 1–24. doi: 10.1007/s10919-014-0190-4

Camerino, O. F., Chaverri, J., Anguera, M. T., and Jonsson, G. K. (2012). Dynamics of the game in soccer: detection of T-patterns. *Eur. J. Sport Sci.* 12, 216–224. doi: 10.1080/17461391.2011.566362

Casarrubea, M., Jonsson, G. K., Faulisi, F., Sorbera, F., Di Giovanni, G., Benigno, A., et al. (2015). T-pattern analysis for the study of temporal structure of animal

and human behavior: a comprehensive review. *J. Neurosci. Methods* 239, 34–46. doi: 10.1016/j.jneumeth.2014.09.024

Caso, L., Maricchiolo, F., Bonaiuto, M., Vrij, A., and Mann, S. (2006). The impact of deception and suspicion on different hand movements. *J. Nonverb. Behav.* 30, 1–19. doi: 10.1007/s10919-005-0001-z

Castañer, M., Barreira, D., Camerino, O., Anguera, M. T., Fernandes, T., and Hileno, R. (2017). Mastery in goal scoring, T-pattern detection, and polar coordinate analysis of motor skills used by Lionel Messi and Cristiano Ronaldo. *Front. Psychol.* 8:741. doi: 10.3389/fpsyg.2017.00741

Castañer, M., Torrents, C., Anguera, M. T., Dinušová, M., and Jonsson, G. K. (2009). Identifying and analyzing motor skill responses in body movement and dance. *Behav. Res. Methods* 41, 857–867. doi: 10.3758/BRM.41.3.857

Cavalera, C., Diana, B., Elia, M., Jonsson, G. K., Zurloni, V., and Anguera, M. T. (2015). T-pattern analysis in soccer games: relationship between time and attack actions. *Cuad. Psicol. Deporte* 15, 41–50. doi: 10.4321/S1578-84232015000100004

Davis, M., Markus, K. A., Walters, S. B., Vorus, N., and Connors, B. (2005). Behavioral cues to deception vs. topic incriminating potential in criminal confessions. *Law Hum. Behav.* 29, 683–704. doi: 10.1007/s10979-005-7370-z

Dcosta, M., Shastri, D., Vilalta, R., Burgoon, J. K., and Pavlidis, I. (2015). "Perinasal indicators of deceptive behavior," in *Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Vol. 1, (Ljubljana: IEEE), 1–8.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 979–995. doi: 10.1037/0022-3514.70.5.979

DePaulo, B. M., and Kirkendol, S. E. (1989). "The motivational impairment effect in the communication of deception," in *Credibility Assessment*, ed. J. C. Yuille (Dordrecht: Kluwer Academic), 51–69.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychol. Bull.* 129, 74–118. doi: 10.1037/0033-2909.129.1.74

DePaulo, B. M., Zuckerman, M., and Rosenthal, R. (1980). Detecting deception: modality effects. *Rev. Pers. Soc. Psychol.* 1, 125–162. doi: 10.1016/j.ijpsycho.2016.04.010

Diana, B., Elia, M., Zurloni, V., Elia, A., Maisto, A., and Pelosi, S. (2015). "Multimodal deception detection," in *Proceedings of the ACM on Workshop on Multimodal Deception Detection*, (New York, NY: ACM Press), 21–28. doi: 10.1145/2823465.2823466

Diana, B., Zurloni, V., Elia, M., Cavalera, C. M., Jonsson, G. K., and Anguera, M. T. (2017). How game location affects soccer performance: T-pattern analysis of attack actions in home and away matches. *Front. Psychol.* 8:1415. doi: 10.3389/fpsyg.2017.01415

Eibl-Eibesfeldt, I. (1970). *Ethology. The Biology of Behavior*. New York, NY: Holt, Rinehart and Winston Inc.

Frank, M. G., and Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *J. Pers. Soc. Psychol.* 72, 1429–1439. doi: 10.1037/0022-3514.72.6.1429

George, J. F., and Robb, A. (2008). Deception and computer-mediated communication in daily life. *Commun. Rep.* 21, 92–103. doi: 10.1080/08934210802298108

Hancock, J. T., Thom-Santelli, J., and Ritchie, T. (2004). "Deception and design: the impact of communication technology on lying behavior," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (New York, NY: ACM), 129–134.

Heilveil, I., and Muehleman, J. T. (1981). Nonverbal clues to deception in a psychotherapy analogue. *Psychother. Theory Res. Pract.* 18, 329–335. doi: 10.1037/h0088382

Jensen, M. L., Meservy, T. O., Burgoon, J. K., and Nunamaker, J. F. Jr. (2010). Automatic, multimodal evaluation of human interaction. *Group Decis. Negot.* 19, 367–389. doi: 10.1007/s10726-009-9171-0

Kahneman, D. (1973). *Attention and Effort*. Englewood, CO: Prentice-Hall.

Kassin, S. M. (2005). On the psychology of confessions: does innocence put innocents at risk? *Am. Psychol.* 60, 215–228. doi: 10.1037/0003-066X.60.3.215

Kassin, S. M., and Norwick, R. J. (2004). Why people waive their Miranda rights. *Law Hum. Behav.* 28, 211–221. doi: 10.1023/B:LAHU.0000022323.74584.f5

Kerepesi, A., Jonsson, G. K., Miklósi, Á, Topál, J., Csányi, V., and Magnusson, M. S. (2005). Detection of temporal patterns in dog–human interaction. *Behav. Processes* 70, 69–79. doi: 10.1016/j.beproc.2005.04.006

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Beha. Res. Methods Instrum. Comput.* 32, 93–110. doi: 10.3758/BF03200792

Magnusson, M. S. (2005). "Understanding social interaction: discovering hidden structure with model and algorithms," in *The Hidden Structure of Interaction: From Neurons to Culture Patterns*, eds L. Anolli, G. Riva, S. Duncan, and M. S. Magnusson (Amsterdam: IOS Press), 3–22.

Magnusson, M. S. (2006). "Structure and communication in interaction," in *From Communication to Presence: Cognition, Emotions and Culture Towards the Ultimate Communicative Experience. Festschrift in Honor of Luigi Anolli*, eds G. Riva, M. T. Anguera, B. K. Wiederhold, and F. Mantovani (Amsterdam: IOS Press), 127–146.

Mammarella, I. C., Toso, C., Pazzaglia, F., and Cornoldi, C. (2008). *BVS-Corsi. Batteria per la valutazione della memoria visiva e spaziale. Con CD-ROM*. Trento: Edizioni Erickson.

McCornack, S. (1997). "The generation of deceptive messages: laying the groundwork for a viable theory of interpersonal deception," in *Message Production: Advances of Communication theory*, ed. J. O. Greene (Mahwah, NJ: Erlbaum), 91–126.

Meyer, D. E., and Kieras, D. E. (1997). A computational theory of human multiple-task performance: the EPIC information-processing architecture and strategic response deferment model. *Psychol. Rev.* 104, 1–65.

Meyer, D. E., Kieras, D. E., Lauber, E., Schumacher, E. H., Glass, J., Zurbriggen, E., et al. (2002). "Adaptive executive control: flexible multiple-task performance without pervasive immutable response-selection bottlenecks," in *Cognitive Modeling*, eds T. A. Polk, A. Thad, and C. M. Seifert (Cambridge, MA: MIT Press), 101–128.

Noldus, L. P., Trienes, R. J., Hendriksen, A. H., Jansen, H., and Jansen, R. G. (2000). The Observer video-pro: new software for the collection, management, and presentation of time-structured data from videotapes and digital media files. *Behav. Res. Methods Instrum. Comput.* 32, 197–206. doi: 10.3758/BF03200802

Noldus, L. P. J. J. (1991). The Observer: a software system for collection and analysis of observational data. *Behav. Res. Methods Instrum. Comput.* 23, 415–429. doi: 10.3758/BF03203406

O'Sullivan, M. (2003). The fundamental attribution error in detecting deception: the boy-who-cried-wolf effect. *Pers. Soc. Psychol. Bull.* 29, 1316–1327. doi: 10.1177/0146167203254610

Pavlidis, I., Eberhardt, N. L., and Levine, J. A. (2002). Human behaviour: seeing through the face of deception. *Nature* 415, 35–35. doi: 10.1038/415035a

Riva, G., Zurloni, V., and Anolli, L. (2005). "Patient-therapist communication in computer assisted environment," in *The Hidden Structure of Interaction. From Neurons to Culture Patterns*, eds L. M. Anolli, S. Duncan, M. S. Magnusson, and G. Riva (Amsterdam: IOS Press), 159–177.

Serota, K. B., Levine, T. R., and Boster, F. J. (2010). The prevalence of lying in America: three studies of self-reported lies. *Hum. Commun. Res.* 36, 2–25. doi: 10.1111/j.1468-2958.2009.01366.x

Spielberger, C. (1983). *Manual for the State-Trait Anxiety Inventory STAI (form Y)*. Palo Alto, CA: Consulting Psychologists Press.

Sporer, S. L., and Schwandt, B. (2007). Paraverbal indicators of deception: a meta-analytic synthesis. *Appl. Cogn. Psychol.* 20, 421–446. doi: 10.1002/acp.1190

Trivers, R. (2011). *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life*. New York, NY: Basic Books.

Tsiamyrtzis, P., Dowdall, J., Shastri, D., Pavlidis, I. T., Frank, M. G., and Ekman, P. (2007). Imaging facial physiology for the detection of deceit. *Int. J. Comput. Vis.* 71, 197–214. doi: 10.1007/s11263-006-6106-y

Turner, R. E., Edgley, C., and Olmstead, G. (1975). Information control in conversations: honesty is not always the best policy. *Kans. J. Soc.* 11, 69–89. doi: 10.17161/STR.1808.6098

Urbaniak, G. C., and Plous, S. (2013). *Research Randomizer (Version 4.0) [Computer software]*. Available at: http://www.randomizer.org/ Retrieved on [accessed June 22, 2013].

Van Merrienboer, J. J., and Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Edu. Psychol. Rev.* 17, 147–177. doi: 10.1007/s10648-005-3951-0

Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities*. Chichester: John Wiley & Sons.

Vrij, A., Akehurst, L., Soukara, S., and Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Hum. Commun. Res.* 30, 8–41. doi: 10.1111/j.1468-2958.2004.tb00723.x

Vrij, A., Edward, K., Roberts, K. P., and Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *J. Nonverbal Behav.* 24, 239–263. doi: 10.1023/A:1006610329284

Vrij, A., Fisher, R., Mann, S., and Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends Cogn. Sci.* 10, 141–142. doi: 10.1016/j.tics.2006.02.003

Vrij, A., Fisher, R. P., and Blank, H. (2017). A cognitive approach to lie detection: a meta-analysis. *Legal Criminol. Psychol.* 22, 1–21. doi: 10.1111/lcrp.12088

Vrij, A., Granhag, P. A., Mann, S., and Leal, S. (2011). Outsmarting the liars: toward a cognitive lie detection approach. *Curr. Dir. Psychol. Sci.* 20, 28–32. doi: 10.1177/0963721410391245

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., and Bull, R. (2008). Increasing cognitive load to facilitate lie detection: the benefit of recalling an event in reverse order. *Law Hum. Behav.* 32, 253–265. doi: 10.1007/s10979-007-9103-y

Walczyk, J. J., Igou, F. P., Dixon, A. P., and Tcholakian, T. (2013). Advancing lie detection by inducing cognitive load on liars: a review of relevant theories and techniques guided by lessons from polygraph-based approaches. *Front. Psychol.* 4:14. doi: 10.3389/fpsyg.2013.00014

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biom. Bull.* 1, 80–83. doi: 10.2307/3001968

Yu, X., Wu, L., Liu, Q., and Zhou, H. (2011). "Children tantrum behaviour analysis based on Kinect sensor," in *Proceedings of the 3rd Conference on Intelligent Visual Surveillance*, (Beijing: IEEE), 49–52.

Zhang, Z. (2012). Microsoft kinect sensor and its effect. *Multimed. IEEE* 19, 4–10. doi: 10.1109/MMUL.2012.24

Zuckerman, M., DePaulo, B., and Rosenthal, R. (1981). "Verbal and nonverbal communication of deception," in *Advances in Experimental Social Psychology*, Vol. 14, ed. L. Berkowitz (New York, NY: Academic Press), 1–57.

Zurloni, V., Diana, B., Cavalera, C., Argenton, L., Elia, M., and Mantovani, F. (2015). Deceptive behavior in doping related interviews: the case of Lance Armstrong. *Psychol. Sport and Exerc.* 16, 191–200. doi: 10.1016/j.psychsport.2014.02.008

Zurloni, V., Diana, B., and Elia, M. (2013). "Imposing cognitive load to unmask prepared lies a temporal pattern detection approach," in *Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, (Shanghai: IEEE), 1–6. doi: 10.1109/FG.2013.6553801

Zurloni, V., Diana, B., Elia, M., and Anolli, L. (2016). "Imposing cognitive load to detect prepared lies: a T-pattern approach," in *Discovering Hidden Temporal Patterns in Behavior and Interaction: T-Pattern Detection and Analysis with THEME*, eds M. S. Magnusson, J. K. Burgoon, and M. Casarrubea (Basel: Springer International Publishing), 63–82.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a shared affiliation, though no other collaboration, with one of the authors CC.

Behavior Patterns of Antisocial Teenagers Interacting with Parents and Peers: A Longitudinal Study

Francisco J. P. Cabrera[1]*, Ana del Refugio C. Herrera[2], San J. A. Rubalcava[3] and Kalina I. M. Martínez[1]

[1] *Laboratorio de Interacción Social, Departamento de Psicología, Universidad Autónoma de Aguascalientes, Aguascalientes, Mexico,* [2] *Psicología, Departamento de Ciencias Sociales, Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, Mexico,* [3] *Laboratorio de Investigación sobre Desarrollo y Contexto del Comportamiento Social, Facultad de Psicología, Universidad Nacional Autónoma de México, Ciudad de México, Mexico*

Antisocial behavior may begin during childhood and if maintained during adolescence, is likely to continue and escalate during adulthood. During adolescence, in particular, it has been established that antisocial behavior may be reinforced and shaped by exchanges between the teenager and his parents and peers, although the molecular process of these relations is as yet unknown. This paper explores the patterns of social interaction established by adolescents with and without the risk of engaging in antisocial behavior in order to understand the exchanges of them with their most important social groups, during 2 years. The study involved a sample of 70 adolescents classified into these two groups (with risk of antisocial behavior and control group). They were video-recorded interacting with one of their parents and one of their peers, independently. The interaction was done about the negotiation of conflictive conversational topics. Those video-records were registered by pairs of trained observers, using an observational catalog with nineteen behavioral categories, to know about the molecular interactional patterns characteristics. Thirty participants were evaluated only once, 30 were evaluated two times, and the other 10 were evaluated three times, the evaluations were performed annually. It was found that a higher occurrence of eye contact and use of open questions and elaborate answers appears to act as a protective factor for engaging in antisocial behavior.

Keywords: social interaction, teenagers, antisocial behavior, parenting, friendship

INTRODUCTION

Antisocial behavior has been defined as kind of behavior that is directed against other people, their property or breaks social rules (Garaigordobil and Maganto, 2016; Jalling et al., 2016; Garaigordobil, 2017). This type of behavior takes various forms (with different seriousness) such as lying, risky sexual practices, rule-breaking, illegal substance use and disruptive behavior such as theft, destruction, fraud, engaging in aggression (either physical or verbal), and vandalism (Patterson, 1982; Kazdin, 1987; Arce et al., 2011; Torry and Billick, 2011; Pears et al., 2016).

This range of behavior makes it a problem whose severity and frequency are a matter of concern (Pedroza, 2006, unpublished). It is usually maintained during adolescence and adulthood in individuals who displayed behavioral problems in childhood (Robins, 1986; Barkley et al., 1990; Campbell, 1991; Gaeta and Galvanovskis, 2011; Snyder et al., 2012; Alink and Egeland, 2013; Rhee et al., 2013; Çelik et al., 2016). The influence of a number of risk

factors linked to the development of this type of behavior has been found, including family environment and involvement with antisocial peers (Alcázar and Bouso, 2008; Antolín et al., 2009).

As for the family environment, risk factors associated with the development of antisocial behavior include marital conflict, family stress, parental authoritarianism, parental criminality, domestic violence, social marginalization and coercive social interaction between parents and children (Loeber, 1990; Frías-Armenta et al., 2003; Quiroz et al., 2007; Antolín et al., 2009; López and Rodríguez-Arias, 2012; Rhee et al., 2013; Çelik et al., 2016).

The term "coercive" refers to the use of aversive utterances by a member of the dyad regarding the behavior of a third party, intended to modify the behavior of the latter (Patterson et al., 1989, 1992; Santoyo and López, 1990). Coercion is the result of escape contingencies and positive reinforcement of aversive events in interactions between parents and their offspring (Smith et al., 2014).

This become clear in Braga et al. (2017) meta-analysis, that included 33 studies centered in antisocial behavior in youth, authors this kind of behaviors associated with suffering aggressiveness in home; they also found that withdrawal/neglect propitiate antisocial enrolment from the adolescent.

These results coincide with Slattery and Meyers's (2014) work; they evaluated parental monitoring, deviant peer's association, aggression in the social environment, and behavioral problems in 503 adolescents. They found that antisocial behavior is correlated positively with the association with antisocial peers. While parental monitoring is negatively associated with behavioral problems in general, and mediates the influence of the aggression in the social environment.

In Centro-America, this kind of result has been also found in a 1599 youth sample using self-reports. It was found that the parental monitoring, conflictive family interactions, and low intimacy, were related with alcohol and drugs consumption (Obando et al., 2014).

Smith et al. (2014) got similar findings in children. They conduct a tree years' study whit 731 dyad parent-toddler observing teaching task, playing, and preparing/eating time, finding a strong relationship between coercive interactions among adult and child and the escalation and generalization of the aggressive behavior in the toddler.

Other work has also found that low quality in family relationships, lack of emotional expression (in parents) and intrusive parental practices can lead social isolation during adolescence (Rovis et al., 2015), this isolation, by itself implies an important risk to be a bullying victim (Aguilera et al., 2013), and rule breaking, vandalism and alcohol consumption (Ettekal and Ladd, 2015).

This may be due to the fact that parents of antisocial teenagers and children often lack the necessary skills to respond appropriately to their children's behavior, using aggressive techniques to modify undesirable behavior, modeling and reinforcing these behaviors over prosocial ones (Torry and Billick, 2011), and thereby establishing patterns of coercive interaction that are replicated in other settings (Patterson et al., 1992; Quiroz et al., 2007; Bowker et al., 2016).

On the other hand, it had been found that positive affect, parental monitoring and responsiveness will facilitate the linking with prosocial peers (Bowker et al., 2016), while adolescents whose parents use harsh discipline tend to link with antisocial peers, as Li et al. (2015) found in their study, in which participated 993 same sex twin's pairs (13.72-year-old) and their parents. In this work the researcher evaluated peer filiation, discipline and parent's negative emotional expression through self-report (questionnaires) from both parents and adolescents. They found that adolescents, whose parents use harsh discipline, tend to link with antisocial peers. This became important taking in account that another important risk factor, especially during adolescence, is the teenagers' interaction with peers, particularly when the latter are antisocial (Snyder et al., 2012), since antisocial behavior in these groups is modeled, and positively reinforced (Ayala et al., 2002; Snyder and Stoolmiller, 2002).

In terms of the peers' influence, a 9 years' longitudinal work, with 998 adolescents has found an increase in deviant behavior with the social reinforcement (of this kind of behavior) provided by peers (Dishion et al., 2012), they also fund that isolated individuals will tend to connect with antisocial peers during adolescence.

Ettekal and Ladd (2015) got similar outcomes in a longitudinal study whit 383 children followed 5 to 14 years old (in nine waves), parents, professor and the children respond questioners about rule breaking, disruptive/aggressive behavior and peer deviant friendship, respectively, while classmates provide information about peer rejection. This work shows that disruptive/aggressive behavior during childhood is a risk for prosocial peer rejecting and later for deviant peers linking and early adolescence rule breaking behavior and aggressive behavior. Consistent with this work, Carlo et al. (2014), using a questionnaire with 666 adolescents, found that prosocial peer filiation were negatively related to antisocial social behavior; Van Ryzin and Dishion (2014), make a 7 years' longitudinal study, found a strong relationship between drug use and affiliation with deviant peers.

Although these factors have been identified, most of the work has been conducted using questionnaires so the molecular interaction process that encourages the presence and maintenance of antisocial behavior remain unknown (Reid et al., 2002). A molecular approach would allow to assess the triple relationship of contingency in terms of antecedents, behavior and consequences that strengthen or weaken the conduct, as molecular studies are characterized by a detailed comportment analysis, behavior by behavior, moment to moment, while molar analysis show us general aspects of behavior (Anguera, 2003; Anguera et al., 2007; Pellón, 2013).

Given the importance of these two major spheres of interaction, it is essential to understand the micro process that takes place during teenagers' interaction with parents and peers, and to determine whether there are differences in interaction patterns at different stages of development. Taking in account the findings in the literature, the hypothesis of this work is that there will be differences between the social interactions held in parent-teen in risk of antisocial behavior and parent-teen without risk. It is also expected that those differences will found in

adolescent-peer interactions; the adolescents in risk would have more conflictive and less responsive social exchanges in both cases.

One method that makes it possible to identify these processes is the direct observation of behavior, since it makes it possible to pinpoint both the coercive process and the development of the trajectories of antisocial behavior (Anguera et al., 2007). This study therefore focused on understanding the interaction patterns of adolescents reported as engaging in any form of antisocial behavior by their teachers as well as possible differences from teenagers matched for age, sex and school year, through the implementation of an interactive task, in a longitudinal study.

MATERIALS AND METHODS

Participants

A total of 70 high school students participated voluntarily, in the period from 2011 to 2013. Participants had a mean of 13 years old ($SD = 8$ months) at the start of the study, and 24 students was men and the rest were women. Of the total sample, 35 students were reported by their teachers as being at risk for antisocial behavior (risk group, RG) of which 15 participated in a single evaluation, 15 in two evaluations and the remainder in three evaluations. The evaluations were performed annually.

The other 35 evaluated students were chosen as controls using two criteria: (1) were matched with participants in at risk-group by age, sex and school year, and (2) they had to be students who were not reported by teachers in the behavioral risk categories. The edge and gender was the same in risk and control groups. Because the study suffered experimental death, different numbers of participants in RG were evaluated during each wave of evaluation, then, each subject in the control group was evaluated so many times as was the person in the RG to which was matched.

Additionally, in this study participated, in each evaluation period, one of the teenagers' parents and one of their friends, which participated voluntarily. The parents were constant if the evaluation occurred two or three times, but in some cases the teenager select a different friend in the next evaluation of teenager-peer interaction.

Scenario

In order to obtain videos on dyadic social interaction, three different scenarios were used: (1) The Psychology Care and Research Unit at the Autonomous University of Aguascalientes, in a 3×4 meter cubicle with a table and two armchairs; (2) an area in high school facilitated by authorities, the scholar auditorium; and (3) the teenager's parents' house, where the activities were done in a room in which was collocated a table and chairs. In each setting, we ensured adequate lighting and privacy.

Materials and Instruments

- Informed consent. Printed document explaining the purpose of the research, as well as the activities carried out, and requesting parent's authorization for the inclusion of their teenager son/daughter to participate in the study. This document

lets the parents know about the possibility that they need to participate either.

- Observer XT® version 9. Computer program designed to record observational data and permits the use of behavioral categories to code behavioral variation in the course of social interaction.

- List of topics. List of 37 conversation topics and a final section in which other topics may be added and classified as conflict or no-conflict, made *ex profeso* for this investigation. The topics were relation with family members, relation with parents, home rules, time of arrival home, permissions, friends, best friend(s), boyfriend/girlfriend, school teachers, school mates, school grades, subjects failed, homework, future studies, eating habits, study habits, reading habits, sporting activities, personal belongings cleaning, room cleaning, housework, dressing way, way of talking, weekends, movie theater, party with friends, television programs, outings with the family, music, vacations, drugs, alcohol, money, work, sexuality, religion, politics.

- Behavioral Catalog for Teachers. Check list of 21 behaviors, based on diagnostic criteria for antisocial behavior from the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders in its fourth edition (American Psychiatric Association [APA], 2002).

- Catalog of Direct Observation of Negotiation (Pedroza et al., 2012). Comprising 19 categories providing a detailed account of eye contact and the content of the utterances made by each member of the dyad during the negotiation process (for access to full catalog, see reference). The categories were evaluated by experts in the area of social exchange and some of them were re-defined after its use in the classification of social exchanges occurred in son of the videos. Experts approve the final version with 100% of concordance.

- Theme 5®.- Computer program designed to locate behavioral patterns hidden from the naked eye; fed with continuous observational data.

- SPSS 20. -Software that makes it possible to process statistical data.

Procedure

The researchers contacted the directors of a public middle school, requesting their authorization to undertake the research activities. Teachers were informed of the project, after which the Behavioral Catalog for Teachers was applied. Once the data had been obtained, students who were eligible for participation were identified. These students and their parents were contacted through the institution, the objectives of the research were explained, and they were asked to sign the informed consent form.

The evaluations spend around 30 min, in which each dyad (teenager-parent or teenager-peer) have two activities, first they were asked to make the classification of the issues of the List of topics as conflictive or no conflictive item of talking. In the second activity, the dyad was asked to negotiate about of the most conflictive issues previously classified, for 20-min. The evaluations were performed two more times, annually.

These interactions were video recorded, and after that were behavioral-categorized using the Catalog of Direct Observation

of Negotiation, through XT® OBSERVER the observations were done for two observers previously training for such labor, and that obtained more than 0.70 of concordance inter-observers three consecutive times, evaluated with the statistic Kappa de Cohen. Finally, the obtained behavioral sequence was used to search for hidden patterns through THEME 5®. A statistical analysis of the data was undertaken through SPSS 20 version, considering the teenager's experimental group, school year and time of assessment, and noting the rate of displays of aggressive behavior and average duration.

RESULTS

In order to analyze the data obtained in the first assessment, we analyzed the behavior displayed by the teenagers in their interaction with their peers, considering whether they belonged to the RG or the matched-control group (CG) and the students' school year, using the Kruskal–Wallis Test. Through these analyses, significant differences were found in *agreement* behavior, which, in the case of the teenagers, was displayed by the control group in the first grade, with an average rate of 0.03 ($SD = 0.01$). Regarding interaction with peers, no significant differences were observed in the population analyzed for state behaviors.

Regarding the rest of the behaviors analyzed in the teenagers' interaction with their parents ($p < 0.05$), a number of differences were observed when they engaged in *debating* behavior (verbalizations that explain or justify facts). A multiple analysis showed that the differences occurred between teenagers in the first and second grade, regardless of the experimental group to which they belonged, with those in the second grade displaying this behavior to the greatest extent, with $\overline{X} = 0.72$ ($SD = 0.1$) in RG, and $\overline{X} = 0.65$ ($SD = 0.12$) in CG, whereas in the first grade, the results were as follows: $\overline{X} = 0.29$ ($SD = 0.12$) in RG, and $\overline{X} = 0.13$ ($SD = 0.21$).

In the case of the students' parents, significant differences were found regarding *negative verbal behavior, containment,* and *hostile containment*. Multivariate analysis revealed that these

differences were due to the utterance rate of parents of RG teenagers in second grade. In particular, this group of parents engaged in these three behaviors more frequently than parents of those in the control group, in second grade. In the case of containment behavior, differences were also observed with regard to the parents of teenagers in the control group in the first grade of middle school, with *p < = 0.05* in all cases.

During the second evaluation, Kruskal–Wallis analysis showed that the utterance rate of RG and CG students in interaction with their peers differed significantly as regards *negative verbal behavior about third parties* and *termination*, which involves verbalizing disapproval of personal or situational behaviors of various individuals outside the dyad in interaction, and a specific request to change topic from the conflictive issue being discussed. Multivariate analysis showed that differences were observed between the data on the teenagers in RG in second grade and those in CG in the second and third grades, the latter, who expressed greater disapproval of third parties with an average rate of 0.37 ($SD = 0.45$) and 0.30 ($SD = 0.28$) utterances, respectively, compared with an average of 0.09 ($SD = 0.07$) utterances by adolescents in the at-RG in second grade. As regards *termination* behavior, a difference was only observed between students in the at-RG in third grade and students in the control group in the second grade of middle school, with the latter displaying the highest rate of requests to change topic. Data on the mean and standard deviation of both behavioral categories are given in **Table 1**.

A comparison of the data on the behaviors displayed by teenagers during the interaction with their parents, considering the school year, yielded significant differences in *clarification* behavior, engaged in at a higher rate of occurrence per minute by teenagers in the control group in second grade ($=1$, $DE = 0.68$) compared to those in first grade RG ($\overline{X} = 0.31$, $SD = 0.21$); second grade RG ($\overline{X} = 0.28$, $SD = 0.15$); third grade RG ($\overline{X} = 0.39$, $SD = 0.27$); first grade CG ($\overline{X} = 0.71$, $SD = 0.43$) and the third grade group CG ($\overline{X} = 0.31$, $SD = 0.28$).

Regarding the time spent by teenagers on issues outside the experimental task, significant differences were found when the experimental group was considered. RG spent more time on this

TABLE 1 | Descriptive statistics of negative verbal and termination behaviors engaged in by teenagers in interaction with peers.

Behavior	Experimental group	School year	Mean	Standard deviation	N
NEGATIVE VERBAL COMMENTS ABOUT THIRD PARTIES	At-risk group	First grade	0.00	0.00	2
		Second grade	0.09	0.07	9
		Third grade	0.13	0.10	10
	Control group	First grade	0.37	0.45	4
		Second grade	0.30	0.28	14
TERMINATION	At-risk group	First grade	0.81	0.71	2
		Second grade	0.40	0.57	9
		Third grade	0.26	0.26	10
	Control group	First grade	0.64	0.15	4
		Second grade	0.62	0.36	14
		Third grade	0.28	0.23	4

The mean reported is about the rate of emission by minute.

category with an average time per minute of 0.05 ($SD = 0.07$) as opposed to the mean of 0.008 ($SD = 0.03$) by the control group. It was also found that teenagers in the control group spent more time per minute in *eye contact* ($\overline{X} = 0.66$, $SD = 0.31$) than those in the at-RG ($\overline{X} = 0.41$, $SD = 0.28$), with an alpha < 0.01.

As for the parents, the second evaluation revealed differences in the number of utterances per minute involving *hostile containment*. The application of multivariate analysis showed that the differences were due to the scores of parents of RG teenagers in third grade when compared with parents of GI adolescents in second grade and those of teenagers in the control group in the second grade. This last group displayed the lowest emission rate per minute with 0.01 and 0.03, respectively.

The time spent on verbal behavior also varied among the parents of the adolescents described in the preceding paragraph, particularly as regards *conversation topic* and *being silent*. Differences were due to the average duration of the utterances of parents of RG teenagers in second grade, who verbalized the topics in the task a third of the time, whereas the other participants showed the reverse pattern. By comparing the data that only considered the experimental group, it was found that the parents of RG teens spent less time addressing conflictive issues ($\overline{X} = 11$ min) than the parents of CG teens ($\overline{X} = 16$ min), $p < 0.005$, but spent longer in silence, $p < 0.01$ (RG $\overline{X} = 8.5$ min; and CG $\overline{X} = 4$ min).

With regard to the third application, Kruskal–Wallis's non-parametrical statistical test was used to determine whether students in the two experimental groups in second and third grade showed significant differences. No significant differences were found in adolescents' behavior in their interaction with peers. It is worth noting that behaviors involving a positive assessment of the other person with whom one is interacting *(positive verbal)* were only observed in participants in the control group; the opposite happened when people absent at the time of the negotiation were evaluated. On the other hand, behaviors that involve offering to make changes (*concession*), giving in to the demands of the other person (*hostile concession*) and requesting a change in the other person coercively (*hostile containment*) were not observed in teenagers during their interaction with their peers. Moreover, although the differences were not significant, a trend of increased time spent on addressing the issues in the experimental task and eye contact was detected in teenagers in the control group, who used it just over half the time in the former behavior and about 80% in the latter, compared with 30 and 50% in the at-RG for *conversation topic and eye contact,* respectively.

As for interaction with parents, teenagers in the control group were the only ones who engaged in *positive verbal behavior* with an average of approximately one utterance per video in the case of participants in second grade and two utterances per video in those in third grade. *Positive verbalization about third parties* occurred with the same average although in this case, it was engaged in by second graders in the at-RG and by third graders in the control group. *Hostile concession* and *agreement behaviors* were not displayed by second graders in either experimental group. The Kruskal–Wallis statistical test only revealed significant differences in the *termination* behavior, with the at-RG displaying this behavior to a greater extent.

A multivariate analysis revealed differences between students in the at-RG in third grade and students in the same experimental group but in second grade of middle school and those in the control group in second grade, with a probability of error of less than 0.05.

Although no significant differences emerged regarding state behaviors, it was found that participants spent more time without making verbal utterances, approximately 80% of the time available for interaction, though eye contact is more common in adolescents in the control group and present in over 60% of the time of interaction as opposed to 35% in the at-RG.

Regarding parents, only parents of third grade adolescents in the at-risk third group engaged in *hostile concession*, with an average utterance rate of 0.02 ($SD = 0.03$). Conversely, *hostile containment behavior* was only displayed by parents of GI adolescents in the second grade ($\overline{X} = 0.15$, $SD = 0.10$) and CG third grade teenagers ($\overline{X} = 0.03$, $SD = 0.04$). *Agreement* behavior was not displayed by these participants.

Although no significant differences occurred in *questioning, clarifying and simple answer* behaviors, involving requests for information, verbalizations that provide extensive information at the request of the other person and brief verbalizations (even monosyllables) given in response to the request for information, it was observed that these behaviors display the highest rate of utterances per minute, as shown in **Table 2**.

Lastly, hidden interactions patterns were sought throughout the evaluations. This analysis yielded an average of 20 patterns per group of observations submitted. This paper therefore only describes the most significant patterns found for each analysis undertaken. It should be noted that all the results obtained through this method have a $p < = 0.0001$.

Regarding all the observations and the participants separated only by experimental group (at-risk/matched), the following behavioral patterns were identified in the interactions between adolescents at risk for antisocial behavior and their parents: the first pattern (**Figure 1A**) shows that there is a high probability that once a parent belonging to group 1 begins talking about the topic of discussion (conversation topic), s/he will also utter expressions of disapproval about the teenager's behavior (negative verbal behavior); subsequently, the teenager will begin discussing the topic in hand.

The second (**Figure 1B**) and the third (**Figure 1C**) relate to the teenager's behavior when s/he begins discussing the subject. There is a high probability that once the teenager begins to

TABLE 2 | Utterance rate per minute of ask, clarify, and simple answer behaviors.

Participant	Ask		Clarify		Simple answer	
	$\overline{X}$	*SD*	$\overline{X}$	*SD*	$\overline{X}$	*SD*
Teenager in interaction with parent	0.45	0.45	0.55	0.51	0.90	0.75
Father	1.43	1.00	0.34	0.29	0.46	0.34
Teenager in interaction with peer	1.97	1.32	0.16	0.28	0.13	0.16
Peer	1.06	0.67	0.47	0.47	0.66	0.56
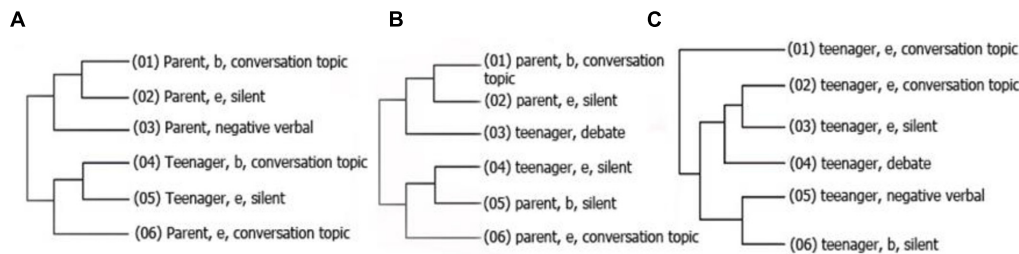
FIGURE 1 | Graphic representation of three hidden patterns (A–C) of interaction occurred between RG teenagers and their parents. Patterns shows the sequence of presentation of the behaviors which were found associated significantly, moreover, the lines connect the conducts that occur more frequency relates with each other. Numbers in parenthesis indicate the sequence of the patter and the letters b and e were used to indicate if the behavior presented after was beginning (b) or ending (e), and appear only if the behavior was a state behavior.

address the conflictive issue, s/he will begin to engage in *debating* behavior (which involves justifying acts or behavior), whereby s/he ends his or her verbal behavior and the parent takes up the conversation.

Three significant patterns emerged in the comments made by adolescents engaged in interaction with their parents. The first (**Figure 2A**) shows that when the teenager is silent, the father will seek information (*ask*), which triggers an elaborate response made by the teenager (*clarify*), which leads to a resumption of the conversation.

The second pattern (**Figure 2B**) shows the link between the expressions of disapproval by the parent and the justifications by the teenager. Another important pattern found (**Figure 2C**) shows the parent's use of expressions designed to justify his or her own behavior prior to the request for a change in the teenager's behavior.

Regarding the patterns found in the different cohorts, in the interactions between first grade adolescents and parents in the at-RG during the first evaluation, a pattern emerged of expressions of disapproval (negative verbal behavior) by the parent, followed by a search for information, which only elicited a short response from the teenager (**Figure 2D**). Conversely, in the control group, patterns were found that included a search for information by the parent (**Figure 3A**) which elicited a short response from the teenager, followed by the use of expressions designed to present arguments related to the topic in hand.

As for the dyads in the control group in second grade in the first evaluation, it was found that the parents expressed disapproval, they continued searching for information to which the teenager gave short replies (**Figure 3B**). It was also found that after s/he began discussing the topic, the teenager uttered expressions designed to explain her or his behavior (**Figure 3C**).

During the 2nd year of assessment, participants from the first, second and third grade were included. In the interactions between second grade adolescents and their peers in the control group, a pattern emerged showing the path of the conversation (**Figure 4A**). In the at-RG, this group contains fewer elements (**Figure 4B**) and only involved one of the participants (this pattern was found in both the adolescent and the peer).

During the 3rd year of evaluation, second and third grade students participated. An important pattern found among second-grade participants in the at-RG interacting with their parents included silence by both the parent and the teenager after the topic had been mentioned by the other person (**Figure 4C**). Another pattern involved the process of negotiation, whereby the teenager *debates*, after beginning to discuss a particular conversation topic.

At the same time, second-grade participants in the control group interacting with their parents showed a pattern whereby the parent interrupted eye contact only after it had been interrupted by the teenager. Another pattern shows that the utterance of expressions of disapproval by the parent (negative
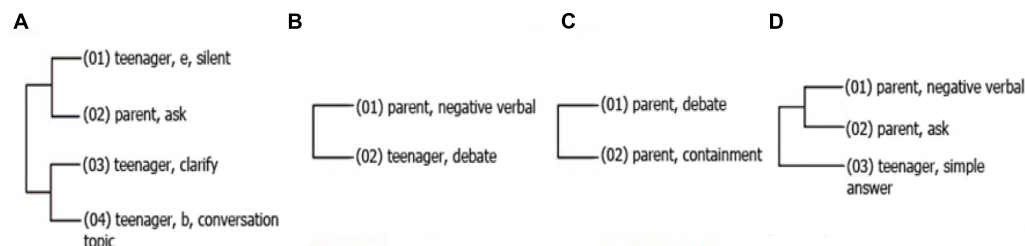
FIGURE 2 | Graphic representation of hidden interaction patterns in exchanges between CG teenagers and parents. This figure presents four patterns called **(A–D)**. Patterns shows the sequence of presentation of the behaviors which were found associated significantly, moreover, the lines connect the conducts that occur more frequency relates with each other. Numbers in parenthesis indicate the sequence of the pattern and the letters b and e were used to indicate if the behavior presented after was beginning (b) or ending (e), and appear only if the behavior was a state behavior.
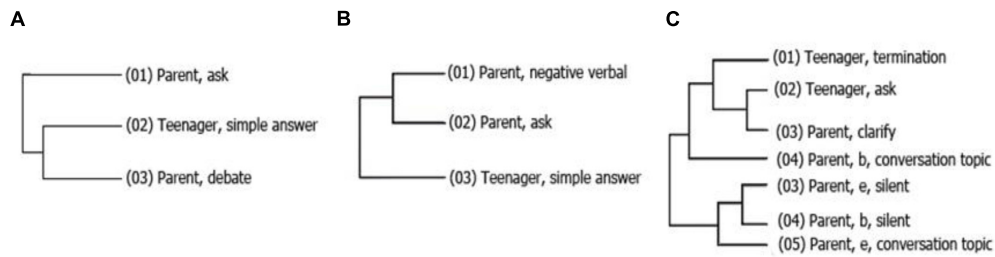
FIGURE 3 | Hidden patterns found in CG (A,C) and RG **(B)** interactions with parents related to the way they ask and give for information. The patterns of adolescents of control group interacting with parents shows a clear interchange of bidirectional information, that is, they used consistently the behavior of ask, while the risk group shows that the parents use aversive events (negative verbal) before asking for information, and the behavior of ask for information didn't appear as part of the teenager's patterns.
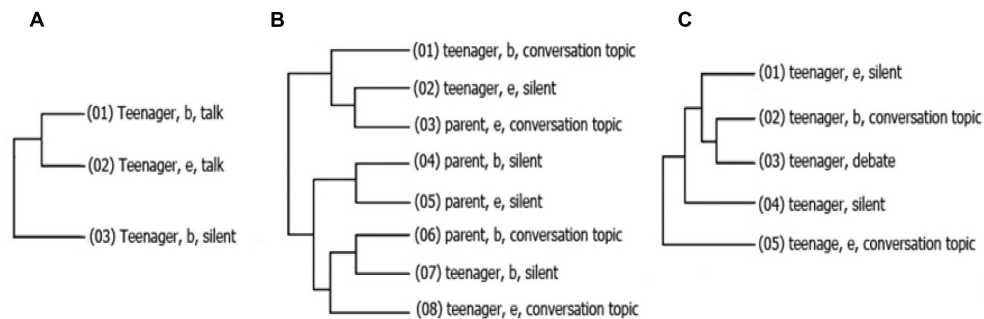
FIGURE 4 | Hidden patterns found in exchanges between RG teenagers and parents. Even though the patterns were found in interactional situations, the patterns **(A,C)** shows that teenagers behavior is consistent itself, regardless the parents behavior. The pattern **(B)** shows that the talking of one of the dyad members was associated with the end of talking of the other.

verbal) occurs after the parent changes the topic of conversation (termination).

Lastly, among the participants belonging to the at-RG who were in third grade during the third wave of evaluation, a total of 16 interaction patterns were found. Of these, the most important shows that in response to negative verbal behavior by the parent, the teenager begins to engage in debate. Another pattern shows that participants ask questions and provide clarification immediately afterward. It is worth noting that both patterns occur in both directions.

In relation to the patterns of interaction found using Theme in the interactions of adolescents with their peers, 72 patterns

were found, of which the following are the most complex for each group.

A pattern of **Figure 5A** shows the changes in Eye contact; where the adolescent begins *eye contact*, that was follow by the peer beginning eye contact, later. **Figure 5B** shows the use of *clarify* by the teenager, follow by a question form the peer that lead to a *simple answer* by the teenager. The **Figure 5C** displays a pattern that has a topic transition (*finish*) made by de teenager, that lead a question by peer; that question (*ask*) is responded by a *simple answer*.

In the adolescents CG interacting with peers (**Figure 6**) a first pattern (**Figure 6A**), shows, the changes in the central topic of
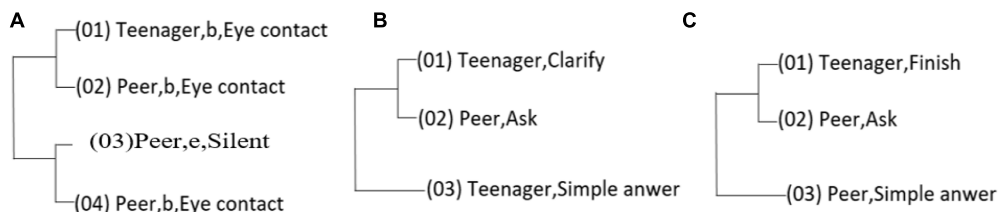
FIGURE 5 | Hidden patterns found in RG adolescent-peer interactions using the software THEME. Numbers in parenthesis indicate the sequence of the pattern and the letters *b* and *e* were used to indicate if the behavior presented after was beginning (b) or ending (e), and appear only if the behavior was a state behavior. The **(A–C)** were used to indicate each one of the patterns.

discussion, where change were made by de peer (finish), using a question as topic changer (ask). The **Figure 6B** contains, a pattern of the using of word, that shows that once de teenager begins (b) to talk about a conversation topic, the peer would start silence (e [end] Conversation topic). Meanwhile, **Figure 6C**, shows that the beginning of *eye contact* from the teenager, leads to *eye contact* by the peer.

DISCUSSION

The mean objective of the present study was to analyze social interaction patterns of teenagers at a micro level, to try of explain how differ the behaviors that occur in teenagers with risk of antisocial behavior versus teenagers without such risk, when they interact with their parents and friends, and how this difference affect the maintenance of teenager's behavior.

The differences between adolescents who are at-risk or not at-risk of antisocial behavior in the interactions with their parents and their peers are reflected in the time each of the groups spends discussing relevant issues (conversational topics), with control group participants spending longer on this behavior. The same occurs with eye contact. The higher rates of occurrence of these behaviors indicate greater exchange in the dyads in the control group.

With regard to interactions between adolescents in both groups and their peers, the fact that no significant differences have been found in the utterance rates may indicate that the greatest difference lies in the type of activities in which they jointly engage (prosocial or antisocial), which are therefore reinforced and modeled (Reid et al., 2002; Snyder and Stoolmiller, 2002). Moreover, the hidden patterns identified in teenager-peer interaction show a little complexity, that is, they talked about the conversational topics using answers and given answers, but didn't used complex behaviors as a constant, such as *debating* or *negative verbal* or categories related to negotiation processes (*concession*, *containment,* or *agreement*).
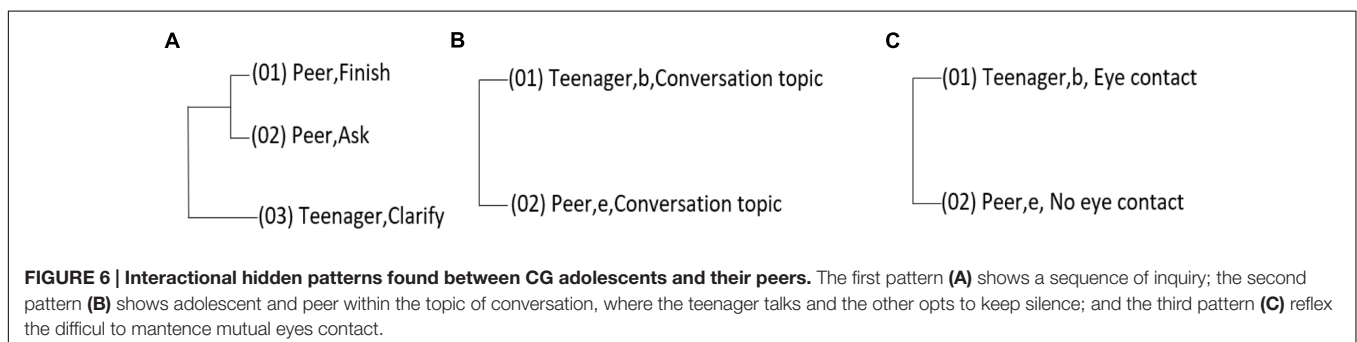
Meanwhile, the micro-analysis of data in terms of the search for hidden patterns showed that the interaction that occurred between adolescents and their peers and parents is qualitatively different, according to previous studies (Ayala et al., 2002; Ettekal and Ladd, 2015). In this regard, observations during the 1st year of evaluation showed that regardless of their school year or experimental group, interaction between adolescents and their peers shows a pattern in which addressing issues is associated with the request for information and subsequently obtaining this information, followed in the case of at-risk students, by a change of subject. This pattern may imply an exchange limited to obtaining data without questioning them, meaning that the exchange of views is not reciprocal. In other words, in addressing various issues, each student questions the other about a different issue, meaning that it is not necessary to share views on a topic when they diverge. If this strategy avoids conflict, it decreases the likelihood of the breakdown of a relationship, which is an indicator of coercion (Reid et al., 2002; Van Ryzin and Dishion, 2014).

These results about teenager-peers interaction give partial confirmation to our hypothesis. On the one hand, more time is spends discussing relevant issues and having visual contact by CG, but, on the other hand, the king of contend of such discussion did not differ significantly from RG. We had argued that the difference could lies in the type of activities they engage together. However, more interaction time of teenagers in CG with their peers let a mayor possibility of practice another behavior that could be efficient in negotiation situations. Future research on older adolescents is necessary to corroborate whether negotiation behaviors improve.

An analysis of the period when the adolescents were evaluated revealed changes in the rates of occurrence of a particular behavior, particularly at-risk family conditions, reflected in the rates of *negative verbal, containment and hostile containment* behavior in parents of RG teenagers in second grade during the 1st year of assessment, when students had been exposed to anti-social peers for longer (Frías-Armenta et al., 2003; Antolín et al., 2009). These rates were higher than those displayed by parents of CG students in first and second grade, which may suggest that parents of matched adolescents use less rigid discipline.

In the case of the at-RG of both first and second grade students in social interaction with parents, addressing conflictive issues follows a behavioral pattern in which parents present an aversive stimulus (*negative verbal*) that is associated with a subsequent request for information, which makes it possible to obtain *short answers* from the teenager, whereas in the case of the control group, the request for information by the parent is not associated with the prior presentation of an aversive stimulus and elicits elaborate answers from the teenager. Although the control group shows a behavioral pattern in which *asking* on

FIGURE 6 | Interactional hidden patterns found between CG adolescents and their peers. The first pattern **(A)** shows a sequence of inquiry; the second pattern **(B)** shows adolescent and peer within the topic of conversation, where the teenager talks and the other opts to keep silence; and the third pattern **(C)** reflex the difficul to mantence mutual eyes contact.

the part of the parent is associated with a *simple answer* from the teenager, in this pattern, the use of a *simple answer* is associated with subsequent engagement in *debating behavior*. These differences in the exchange structure may indicate the use of escape contingencies by the at-RG of teenagers (Reid et al., 2002; Snyder and Stoolmiller, 2002; Smith et al., 2014), while in the control group it represents the continuity of discussion, that had more posibiliti of *agreement,* as data indicated.

During the 2nd year of evaluation, a striking feature is the use of the disapproval of third parties' behavior by RG teenagers in the third grade when interacting with their peers, which may suggest that the issue of aggression toward others plays an important role in RG teenagers' exchanges with their peers and, therefore their affiliation with antisocial peers, which constitutes a risk factor for engaging in antisocial behavior (Snyder et al., 2012).

As for teenager-parent interactions in the second evaluation, it is important to highlight the use of elaborate responses (*clarify*) by the parents of CG adolescents in second grade, which was significantly higher than the rate of RG participants in all grades who participated in this evaluation. This situation may indicate the use of good parenting skills expressed through the use of successful information seeking strategies by parents and low rates of aggressive requests (*hostile containment*) and the availability to interact and low use of avoidance strategies by adolescents, which may explain the absence of antisocial behavior in this group (Patterson et al., 1992; Torry and Billick, 2011).

Regarding the 3rd year of evaluation, in teenagers' interaction with their parents, it is striking that CG teenagers express approval of their parents' behavior (*positive verbal*); this may indicate the good state of interactions between adolescents and parents in this group and therefore the absence of coercive interactions between them. Conversely, the constant change of topic of conversation (*termination*) by RG members may indicate escape strategies, as well as a low rate of exchange. These data, coupled with parents' requests to change the topic, contingent on the use of negative consequences (*hostile containment*) indicate the presence of a coercive process between the parents and adolescents in this group, which implies risk factors for adolescents to engage in antisocial behavior (Patterson et al., 1989, 1992; Santoyo and López, 1990). Parents in CG not only had a prosocial interaction with their teenage children but also modeled to them how act when a conflict issue is addressed.

By undertaking the analysis considering data from all the evaluations and only differentiating participants by experimental group, it was found that in GI parent-child exchanges, beginning to address a specific topic was associated with the utterance of *negative verbal* phrases by both by parent and teenager, which implies the constant disparagement of the person with whom one is interacting. In the case of adolescents, this disparagement is associated with the previous occurrence of verbalization in which the adolescent justifies his or her behavior (*debate*). This situation shows the use of aversive utterances in the exchanges, which may serve as a predictor of antisocial behavior since the display of physical and verbal aggression during childhood and adolescence has been associated with the development of this type of problems

in the literature (Gaeta and Galvanovskis, 2011; Snyder et al., 2012; Alink and Egeland, 2013; Rhee et al., 2013; Garaigordobil and Maganto, 2016; Jalling et al., 2016).

These data also show that the behavior modeled by RG parents is aggressive, since the display of verbal disapproval associated with the start of verbal behavior oriented toward the chosen conversation topic (*conversation topic*) by the father is also a risk factor for the development of antisocial behavior as is parents' authoritarianism (Loeber, 1990; Frías-Armenta et al., 2003; Quiroz et al., 2007; Antolín et al., 2009; López and Rodríguez-Arias, 2012).

The fact that RG teenagers' acts are not associated with the expression of disapproval by the parent could also denote the use of avoidance strategies by the teenager and thus the development of coercive interactions (Patterson et al., 1989, 1992; Santoyo and López, 1990).

The use of verbalizations in which the disapproval of adolescent behavior (*negative verbal*), prescribing changes in the adolescent's behavior (*containment*) and the anticipation of negative consequences if the prescribed changes are not made (*hostile containment*) create an inauspicious setting for negotiation, which explains the absence of engagement in "agreement" by parents and their offspring in the at-RG.

A different situation was identified in the case of CG. Parents' engagement in *negative verbal* behavior is associated with the justification of facts (*debate*) by adolescents. Moreover, the utterance of parent's disapproval was associated with previous use of *debating* behavior which may indicate the use of inductive behavior by parents, and the prescription of previously justified changes.

Another difference found in the parent-child interaction patterns between the at-risk and control groups is that in the latter, the questions used by parents are not only associated with obtaining elaborate answers from the teenager (*clarification*), but also result in the continuation of the conversation, allowing greater exchange between members of the dyad, which could be acting as protection factor (Quiroz et al., 2007; Bowker et al., 2016).

The more complexity patterns of teenager-parent interaction in CG were principally related to parents' behavior. Previous research had found that parental monitoring and responsiveness was associated with prosocial behaviors of the children (Obando et al., 2014; Slattery and Meyers, 2014; Bowker et al., 2016), in this investigation we had found that the principal aspect of parent-teenager exchanges was related to the presentation of a more varied behavioral repertoire by parents, corroborating our hypothesis about it.

It is important to note that this paper sheds light on the molecular process (Reid et al., 2002) which contributes to the emergence and maintenance of antisocial behavior in adolescents, through the identification of specific behavior patterns in adolescents at risk antisocial behavior in their interaction with their parents and peers, which differ from the rates and behavioral patterns engaged in by teenage-parent dyads which are not at risk of antisocial behavior. Since this is a cohort study, it is important to stress the importance of longitudinal studies covering this developmental stage (adolescence).

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of "Mexican Society of Psychology, and bioethics committe of Universidad Autónoma de Aguascalientes," with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by "the bioethics committe of Universidad Autónoma de Aguascalientes."

AUTHOR CONTRIBUTIONS

FC was the main researcher of this work; he provided de original idea and organize de research group. Later FC and AH work together on the design of the project, and with SR and KM they organized the work in field and in the training of the research assistants. FC, AH, SR, and KM had a role in the collecting data process and in the analysis and interpretation of the final data. AH and SR wrote the first draft, that late was developed by AH, FC, SR, and KM by revising and rewriting the contents. All of them worked on the final draft and agrees to be accountable for all aspects of the work in ensuring that questions related to the

accuracy or integrity of any part of the work are appropriately investigated and resolved.

REFERENCES

Aguilera, S. J., Pedroza, F. J., and Cervantes, A. R. (2013). "Roles desempeñados en el bullying: implicaciones prácticas," in *La Construcción de Identidades Agresoras: el Acoso Escolar en México*, eds F. J. Pedroza and S. J. Aguilera (México: CONACULTA), 31–48.

Alcázar, C. M. Á., and Bouso, S. J. C. (2008). La personalidad y la criminología. Un reto para la psicología. *Anu. Psicol.* 18, 99–111.

Alink, L., and Egeland, B. (2013). The roles of antisocial history and emerging adulthood developmental adaptation in predicting adult antisocial behavior. *Aggress. Behav.* 39, 131–140. doi: 10.1002/ab.21464

American Psychiatric Association [APA] (2002). *Manual Diagnóstico y Estadístico de los. (Trastornos) Mentales DSM-IV-TR*. Barcelona: Masson.

Anguera, M. T. (2003). "La observación," in *Evaluación Psicológica. Concepto, Proceso y Aplicación en las Áreas del Desarrollo y de la Inteligencia,* ed. C. Moreno Rosset (Madrid: Librería Sanz y Torres).

Anguera, M. T., Magnusson, M., and Jonsson, G. (2007). Instrumentos no estándar: planteamiento, desarrollo y posibilidades. *Av. Med.* 5, 63–82.

Antolín, L., Oliva, A., and Arranz, E. (2009). Contexto familiar y conducta antisocial infantil. *Anu. Psicol.* 40, 313–327.

Arce, R., Fariña, F., and Vázquez, M. J. (2011). Grado de competencia social y comportamientos antisociales, delictivos y no delictivos en adolescentes. *Rev. Latinoam. Psicol.* 43, 473–486.

Ayala, H. E., Pedroza, F. J., Morales, S., Chaparro, A. A., and Barragán, N. (2002). Factores de riesgo, factores protectores y generalización del comportamiento agresivo en una muestra de niños en edad escolar. *Salud Ment.* 25, 27–40.

Barkley, R. A., Fisher, M., Edelbrock, C. S., and Smallish, L. (1990). The adolescent outcome of hyperactive children diagnosed by research criteria: I. An 8-year prospective follow-up study. *J. Am. Acad. Child Adolesc. Psychiatry* 29, 546–557. doi: 10.1097/00004583-199007000-00007

Bowker, J. C., Rubin, K. H., and Coplan, R. J. (2016). "Social withdrawal," in *Encyclopedia of Adolescence*, ed. R. J. R. Levesque (New York, NY: Springer), 1–14. doi: 10.1007/978-3-319-32132-5_124-2

Braga, T., Gonçalves, L. C., Basto-Pereira, M., and Maia, Â. (2017). Unraveling the link between maltreatment and juvenile antisocial behavior: a meta-analysis of prospective longitudinal studies. *Aggress. Violent Behav.* 33, 37–50. doi: 10.1016/j.avb.2017.01.00

Campbell, S. B. (1991). "Longitudinal studies of active and aggressive preschoolers: individual differences in early behavior and outcome," in *Rochester Symposiumon Developmental Psychopatology*, eds D. Cicchetti and S. L. Toth (Hillsdale, NJ: Erlbaum), 57–90.

Carlo, G., Mestre, M. V., McGinley, M., Tur-Porcar, A., Samper, P., and Opal, D. (2014). The protective role of prosocial behavior on antisocial behaviors: the mediating effects of deviant peer affiliation. *J. Adolesc.* 37, 359–366. doi: 10.1016/j.adolescence.2014.02.009

Çelik, S., Diken, H. I., Çolak, A., Arkan, A., Aksoy, F., and Tomris, G. (2016). Effectiveness of the preschool version of the first step to success early intervention program for preventing antisocial behavior. *Educ. Sci. Theory Pract.* 16, 511–535. doi: 10.12738/estp.2016.2.0317

Dishion, T. J., Ha, T., and Véronneau, M. (2012). An Ecological analysis of the effcts of deviant peer clustering on sexual promiscuity, problem behavior, and childbearing from early adolescence to adulthood: an enhancement of the life history framework. *Dev. Psychol.* 48, 703–717. doi: 10.1037/a0027304

Ettekal, I., and Ladd, G. W. (2015). Developmental pathways from childhood aggression-disruptiveness chronic peer rejection, and deviant friendships to early-adolescents rule breaking. *Child Dev.* 86, 614–631. doi: 10.1111/cdev.12321

Frías-Armenta, M., López, E. A. E., and Díaz-Méndez, S. G. (2003). Predictores de la conducta antisocial juvenil: un modelo ecológico. *Estud. Psicol.* 8, 15–24. doi: 10.1590/S1413-294X2003000100003

Gaeta, M. L., and Galvanovskis, A. (2011). Propensión a Conductas antisociales y delictivas en adolescentes mexicanos. *Psicol. Iberoam.* 19, 47–54.

Garaigordobil, M. (2017). Conducta antisocial: conexión con bullying/ciberbullying y estrategias de resolución de conflictos. *Psychosoc. Interv.* 26, 47–54. doi: 10.1016/j.psi.2015.12.002

Garaigordobil, M., and Maganto, C. (2016). Conducta antisocial en adolescents y jóvenes: prevalenca en el país Vasco y diferencias en función de vaiables socio-demográficas. *Acción Psicol.* 13, 57–68. doi: 10.5944/ap.13.2.17826

Jalling, C., Bordin, M., Romelsjö, A., Källmén, H., Durbeej, N., and Tengström, A. (2016). Parent programs for reducing adolescent's antisocial behavior and substance use: a randomized controlled trial. *J. Child Fam. Stud.* 25, 811–826. doi: 10.1007/s10826-015-0263-y

Kazdin, A. E. (1987). Treatment of antisocial behavior in children: future directions. *Psychol. Bull.* 102, 187–203. doi: 10.1037/0033-2909.102.2.187

Li, M., Chen, J., Li, X., and Deater-Deckard, K. (2015). Moderation of harsh parenting on genetic and environmental contributions to child and adolescent deviant peer affiliation: a longitudinal twin study. *J. Youth Adolesc.* 44, 1396–1412. doi: 10.1007/s10964-015-0288-y

Loeber, R. (1990). Development and risk factors of juvenile antisocial behavior and delinquency. *Clin. Psychol. Rev.* 10, 1–41. doi: 10.1016/0272-7358(90)90105-J

López, L. S., and Rodríguez-Arias, J. L. (2012). Factores de riesgo y de protección en el consumo de drogas y la conducta antisocial en adolescentes y jóvenes españoles. *Int. J. Psychol. Res.* 5, 25–33.

Obando, D., Trujillo, A., and Trujillo, C. A. (2014). Substance use and antisocial behavior in adolescents: the role of family and peer-individual risk and protective factors. *Subst. Use Misuse* 49, 1934–1944. doi: 10.3109/10826084.2014.956365

Patterson, G. R. (1982). *Coercive Family Process*. Eugene, OR: Castalia.

Patterson, G. R., DeBaryshe, B. D., and Ramsey, E. (1989). A developmental perspective on antisocial behavior. *Am. Psychol.* 44, 329–335. doi: 10.1037/0003-066X.44.2.329

Patterson, G. R., Reid, J. B., and Dishion, J. T. (1992). *Antisocial Boys*. Eugene, OR: Castalia.

Pears, K. C., Kim, H. K., and Fisher, P. A. (2016). Decreasing risk factors for later alcohol use and antisocial behavior in children in foster care by increasing early promotive factors. *Child. Youth Serv. Rev.* 65, 156–165. doi: 10.1016/j.childyouth.2016.04.005

Pedroza, F. J., Cervantes, A. R., Aguilera, S. J., and Martínez, K. I. (2012). Interacciones sociales de adolescentes consumidores y no consumidores de sustancias adictivas. *Rev. Mex. Anál. Conducta* 38, 126–150.

Pellón, S. R. (2013). Watson, skynner y algunas disputas dentrol del conductismo. *Rev. Colomb. Psicol.* 22, 389–399.

Quiroz, V. N., Villatoro, J. A., Juárez, G. F., Gutiérrez, L. M. L., Amador, B. N. G., and Medina-Mora, I. M. E. (2007). La familia y el maltrato como factores de riesgo de conducta antisocial. *Salud Ment.* 30, 47–54.

Reid, J. B., Patterson, G. R., and Snyder, J. E. (2002). *Antisocial Behavior in Children and Adolescents: A Developmental Analysis and Model for Intervention*. Washington, DC: American Psychological Association. doi: 10.1037/10468-000

Rhee, S. H., Friedman, N., Boeldt, D., Corley, R., Hewitt, J., Knafo, I., et al. (2013). Early concern and disregard for others as predictors of antisocial behavior. *J. Child Psychol. Psychiatry* 54, 157–166. doi: 10.1111/j.1469-7610.2012.02574

Robins, L. N. (1986). "Changes in conduct disorders over time," in *Risk in Intellectual and Psychosocial Development*, eds D. C. Farran and J. D. McKinney (New York, NY: Academic Press), 227–259.

Rovis, D., Bezinovic, P., and Basic, J. (2015). Interactions of school bonding, disturbed family relationships, and risk behaviors among adolescents. *J. School Health* 85, 671–679. doi: 10.1111/josh.12296

Santoyo, C., and López, R. F. (1990). *Análisis Experimental del Intercambio Social*. Mexico: Trillas.

Slattery, T., and Meyers, S. (2014). Contextual predictors of adolescent antisocial behavior: the developmental influence of family, peer, and neighborhood factors. *Child Adolesc. Soc. Work J.* 31, 39–59. doi: 10.1007/s10560-013-0309-1

Smith, J. D., Dishion, T. J., Shaw, D. S., Wilson, M. N., and Patterson, G. R. (2014). Coercive processes and early-onset conduct problems from age 2 to school entry. *Dev. Psychopathol.* 26, 917–932. doi: 10.1017/S0954579414000169

Snyder, J., Schrepferman, L., Bullard, L., McEachern, A., and Patterson, G. (2012). Covert antisocial behavior, peer deviancy training, parenting processes, and sex differences in the development of antisocial behavior during childhood. *Dev. Psychopathol.* 24, 1117–1138. doi: 10.1017/S0954579412000570

Snyder, J., and Stoolmiller, M. (2002). "Reinforcement and coercion mechanism in the development of antisocial behavior: the family," in *Antisocial Behavior in Children and Adolescents: A Developmental Analysis and Model for Intervention*, eds J. B. Reid, G. R. Patterson, and J. Snyder (Washington, DC: American Psychological Association), 65–100.

Torry, Z., and Billick, S. (2011). Implications of antisocial parents. *Psychiatr. Q.* 82, 275–285. doi: 10.1007/s11126-011-9169-z

Van Ryzin, J. M., and Dishion, T. J. (2014). Adolescent deviant peer clustering as an amplifying mechanism underlying the progression form early substance use to late adolescent dependence. *J. Child Psychol. Psychiatry* 55, 1153–1161. doi: 10.1111/jcpp.12211

Using Systematic Observation and Polar Coordinates Analysis to Assess Gender-Based Differences in Park Use in Barcelona

Félix Pérez-Tejera [1,2]*, Sergi Valera [1,2] and M. Teresa Anguera [1,3]

[1] Department of Social Psychology and Quantitative Psychology, Faculty of Psychology, University of Barcelona, Barcelona, Spain, [2] Social Environmental and Organizational Psychology Research Group (PsicoSAO), Barcelona, Spain, [3] Institute of Neurosciences, University of Barcelona, Barcelona, Spain

This paper aims to assess gender differences in the usage of public open spaces (POS), as an everyday context. Forty POS in the city of Barcelona were studied over 3 months using systematic observation. To objectively measure park use, an observational instrument (EXOdES) was purposely designed combining a field format and several category systems. The instrument facilitated the record of configurations or co-occurrences of codes from different dimensions (i.e., time of day, age, race/ethnicity, activity setting, activity, and presence of vehicles), providing contextually rich data of more than 35,000 individuals and groups and the setting in which the activity occurs. Although a similar overall proportion of males and females were found using POS (55 vs. 45%), important differences by gender were found between people being alone (66 vs. 34%), and groups (53 vs. 47%). To identify regular patterns in the way that men and women use public parks, information on more than 18,000 groups of people was analyzed as a global data set. A multievent sequential analysis was performed considering gender composition as the given behaviors (i.e., groups of males, females, and gendered mixed). Thus, polar coordinates analysis was also performed, because it is a suitable reduction data technique in studies with a broad observational instrument and a large database. Results show important gendered and cultural differences in POS use. Women tend to reproduce traditional gender role, being often more engaged in care functions with children and elders rather than in any other activity or with people of their same age group. Of particular concern is the gap on park use observed in women of ethnic minority groups. Assessing specific group needs on park use is particularly relevant attending to their multiple health and social benefits.

Keywords: park use, public space, gender perspective, systematic observation, lag sequential analysis, polar coordinate analysis

INTRODUCTION

Public open spaces (POS) such as urban parks, open green spaces and squares, contribute to life quality in urban areas in many ways (Chiesura, 2004). Green spaces have stress-reduction and mental health benefits, as contact with nature has a number of restorative effects (Ulrich, 1984; Hartig et al., 1991; Ulrich et al., 1991; Hull and Michael, 1995; Kaplan, 1995; Hansmann et al., 2007; Collado and Staats, 2016). The largely free and accesible character of POS provides a setting for leisure activities and free opportunities for physical activity, which have been linked with multiple benefits to psychological and physical well-being, including weight management, controlling blood pressure, decreasing the risk of heart disease, strokes, breast cancer, and Type 2 diabetes (Godbey, 2009). Spending more time outdoors has also been linked with better health indicators because several indoor air pollutants and vitamin D deficiency, as a consequence of low sun exposure, are also associated with the pathogenesis of frequent chronic diseases (Viegi et al., 2004; Peterlik and Cross, 2005).

Additionally, from a psychological perspective, experiences in specific local places (e.g., public parks, squares, and markets) provide contexts for developing place-identity and might contribute to taste flow and well-being (Bonaiuto et al., 2016). POS are also essential for establishing social recognition and interaction, promoting friendship between neighbors, social cohesion, and a sense of community (Coley et al., 1997; Kuo et al., 1998; Cattell et al., 2008; Vargas and Merino, 2012). Cattell et al. (2008, p. 556) describe the beneficial properties of public spaces in community life: "Social interaction in public spaces, for example, can provide relief from daily routines, sustenance for people's sense of community, opportunities for sustaining bonding ties or making bridges, and can have a direct influence on wellbeing by raising people's spirits." In a similar way, after pointing out the negative correlation between social cohesion and neighborhood insecurity, Vargas and Merino (2012, p. 172) claimed that "it is likely that perceptions of insecurity might decrease if children, youth, families and elder populations are integrated in the space with social activities creating social networks and a sense of community." That is the crucial role of public spaces on social life.

Attending to the multiple benefits of POS on physical and psychosocial well-being, research has recently put more attention on questions related to environmental justice (Wolch et al., 2014). Using GIS-based measures, several studies have reported income and racial/ethnic disparities in access to recreation facilities, especially in the U.S. (Dahmann et al., 2010; Sister et al., 2010). On a recent review of the equity mapping literature on urban parks, Rigolon (2016) recently concluded that low socioeconomic status and ethnic minority communities have access to fewer parks, fewer park acres, and parks that are potentially more congested. Addressing social disparities in park provision not always require the creation of new public spaces, but also improving those that are underutilized.

Safety has been cited by both adolescents and adults, and in particular women, as one of the most important reasons for not using POS (Burgess et al., 1988; Valentine and Mckendrick,

1997; Molnar et al., 2004; Weir et al., 2006; Casper et al., 2013; Babey et al., 2015). In a review of qualitative research about park use, McCormack et al. (2010) found concerns as to the presence of "undesirable users" (e.g., drug users/dealers, homeless, and loiterers) also some park attributes related to injury safety (e.g., presence of glass, syringes, rocks, debris, heavy traffic) are also often mentioned as discouraging reasons for using public parks. This effect can be related to a disorder model about unsafety (Franklin et al., 2008), according to which both "social and physical incivilities are signs of lack of adherence to norms of public behavior" (Taylor and Hale, 1986, p. 154). Other studies have suggested that modifying park facilities could have a greater impact to increase park use than improving perceptions of park safety (Cohen et al., 2009; Lapham et al., 2016). Urban planners can play a key role in helping communities to have the same opportunities to access public parks. Assessing the type of users and activities that POS attract, can provide valuable information to identify existing disparities of access by certain specific groups. In contrast, as Sister et al. (2010) have stated, "the theoretical perspectives on social justice have seldom translated into practical methods and techniques applicable in the field, failing to provide specific tools for planners to assess, and address social disparities."

Systematic observation has been proven effective in the analysis of natural contexts, respecting the maximum display of naturality (Anguera, 2003). Contrary to self-reports, systematic observation is a direct method that can provide objective information with strong internal validity and allows for the simultaneous generation of information about the physical and social environment where the activity is taking place (McKenzie and van der Mars, 2015). Recently, the analysis of park use with systematic observation has received considerable more attention, but because most of the studies have been conducted in the United States and focused on physical activity levels, important areas of interest remain still unclear.

Previous research consistently has shown a gender gap on park use, suggesting the existence of structural and cultural factors that influence women's leisure opportunities in an urban context (Scraton and Watson, 1998). More males than females tend to use public parks, being males more physically active than women (Evenson et al., 2016; Derose et al., 2018). According to Krenichyn (2004, p. 118), "women are underrepresented in urban parks and plazas and their absence is attributable to actual or perceived vulnerability to crime and threatening or sexually aggressive behavior, or that they use parks most often in the context of family and child-care activities." Jackson and Henderson (1995, p. 48) also described women constrained in their leisure time "because of the social expectations (women are still primarily responsible for childcare in our society) and social controls (women make less money than men) associated with gender." As a consequence, opportunities for leisure in public settings may be especially limited for women (Skogan and Maxfield, 1981; Hutchison, 1994; Perkins and Taylor, 1996).

The main goal of the present study is to offer a tool that can assist planners in addressing specific questions regarding park use. From a methodological point of view, our objectives are: (1) to present an observational instrument designed to record

park use as naturally occurs in daily life and (2) to show an example of the possibilities that polar coordinates offer to analyse observational data. In this paper, we use this methodology from a gender perspective to explore gender disparities on park use in Barcelona, Catalonia (Spain). Barcelona is a city with a low and stable victimization index around 15%, basically referred to minor crimes and well-recognized urban public spaces (Valera and Guàrdia, 2014). Nevertheless, insecurity is usually defined by its citizens as one of the most important problems of the city, together with other topics also linked with fear of crime, such as cleanliness, immigration, vandalism, and poverty[1]. According to Subirats (2006), governance of the public space in Barcelona is today getting more complex as a result of economic, political, and social dynamics that implies greater job insecurity, more unemployed people on the streets, poverty, and ethnic diversification. A better understanding of how men and women use POS may ultimately lead to interventions to promote park use for all kinds of users and improve perceived safety on urban areas.

METHODS

Design

We employed an N/F/M observational design (Blanco-Villaseñor et al., 2003; Anguera and Hernández-Mendo, 2013), where N refers to nomothetic (observing numerous POS and groups of people), F refers to intersessional follow-up (recording of numerous sessions) and M refers to multidimensional (analysis of multiple criteria included in the observational instrument).

Participants

Forty POS distributed among all 10 districts of the city of Barcelona were analyzed (**Figure 1**). In order to have different levels of analysis, the sample included 20 POS in Sants-Montjuïc and 20 POS in nine different city districts. The election of the sample was oriented balancing the presence and absence of physical and social disorder signs. To organize observational data collection, 10 circular routes were defined, each one including 4 POS within <15 min walk or by public transportation. Exclusion criteria were (1) very small POS where the presence of an observer could easily produce reactivity and (2) an excessive distance between public spaces included on the same observational route. Final selection included open spaces ($n = 2$), open green spaces ($n = 18$), small town squares ($n = 13$), and large district parks ($n = 7$) across the city. When necessary, POS were divided into smaller targeted areas to facilitate systematic observation.

Materials

Observational Instrument

An observational tool (EXOdES) was specifically created to assess park use and the environmental features of space where activity took place. EXOdES is an *ad hoc* instrument (Sánchez-Algarra and Anguera, 2013) based on the combination of field format and category systems, which permits recording co-occurrent

behaviors regarding multiple criteria. This work was developed as part of a broader project, including the development of new observational instrument to assess park use and the consequences of fear of crime on the activity patterns of public space. We conceptualized four different set of factors or macro-criteria: (1) contextual information (observer, date, observational period, public space, location/activity setting), (2) individual criteria (age, gender, and ethnicity of both people being alone and groups, size of groups, ethnic diversity of groups, poverty signs/homelessness), (3) activity criteria (main activity, dogs, vehicles, problematic uses, substances use signs, violence), and (4) environmental criteria (brightness, cleanness, visual control, green space maintenance, litter, graffiti). Category systems were defined for those criteria with limited options (e.g., gender, age, and race/ethnicity) and catalogs were created for those criteria with unlimited possibilities (i.e., type of vehicles and the main activity sports) that could be extended in case of observing new responses not previously considered. An earlier pilot and more details about the observational instrument and procedure can be found in previous works (Pérez-Tejera et al., 2011; Pérez Tejera, 2012; Valera et al., 2018). Six criteria of the observational system were selected for the present study to describe park use: time of day, age group, race/ethnicity, location, activity, and vehicles (**Table 1**). Environmental factors were excluded for a question of space and other relevant criteria regarding park use -problematic uses, substances use signs, violence, poverty signs- were also excluded for being infrequent, although their park use implications can be explored in the future.

Observers

Eight observers and two digital recorders were contracted half-time by the City Council of Barcelona and coordinated by the researchers. Training consisted of in-class and field-based training and occurred over the course of 1 month. In-class training provided an overview of the study purpose, data collection materials, park observation protocols, and EXOdES training with photographs. Field-based training consisted of on-site visits to each park to review its location and to practice the data collection with EXOdES under investigator supervision. Observers participated in the elaboration of detailed maps of each park identifying all targeted areas within each (e.g., football field, play-ground equipment, and open space). The control of quality of data has been done through kappa Cohen's coefficient, that has been satisfactory, exceeding 80%. Also, correlation coefficient is higher than 0.80.

Procedure

Systematic records were performed between September 2010 and December 2010. All POS were visited 8 times per day (observation period): 10:00–11:00, 11:00–12:00, 12:00–13:00, 13:00–14:00, 16:00–17:00, 17:00–18:00, 18:00–19:00, and 19:00–20:00. After assuring high levels of inter-rater reliability during training, observations were conducted by 1 observer. Every weekday during the study period every observer was assigned to one of 10 routes including 4 POS, in a morning (from 10:00 to 14:00) or afternoon turn (from 16:00 to 20:00). Each observational session was defined as a 45-min observational

---

[1]Source: Municipal Services Survey. Barcelona City Council.

FIGURE 1 | Location map for observed POS.

period. After the first observational session, the observer moved to the next POS of the route and started a new 45-min observational session until complete the assigned route. With this procedure, short observation sessions were ensured reducing the risk of observer fatigue and reactance (Hoeben et al., 2018). At the end of the study, every POS was observed at 8 different observational periods, a median of 5 different days, by at least 3 different observers and 3 different weekdays, to diminish some bias. Observations were conducted only during good weather. When special events took place in the POS, observational sessions were rescheduled on the same weekday in the following weeks.

During each observational session, observational scans of target areas were performed periodically to obtain information about park use. A scan is a single observation or visual sweep from left to right across the target area. All individuals or groups observed in each location during a 45-min observational period were recorded naturally. In the case of individuals, age group (i.e., children, teens, adults, and elders), and race/ethnicity (i.e., White, Latin, Arab, Asian, and African) were recorded. The make-up of the groups were recorded accordingly: size of the group (i.e., 2, 3–5, 6–10, and 10–20), gender composition (i.e., men, women, mostly men, mostly women, and equally mixed), age (i.e., children, teens, adults, elders, children with teens, adults with elders, and children/teens with adults/elders), and ethnicity using the same taxonomy for individuals. Aditionally, groups were also classified regarding their ethnic homogenity (i.e., Whites, mostly whites, equally mixed, mostly non-whites, and non-whites). The activity setting or target area where

people were observed (e.g., sport court, playground, and open space), the activity (e.g., play, sports, and walk) and the presence or absence of vehicles (e.g., no vehicles, skate, and stroller) were also recorded. Thus, each individual or group using the space during an observational session were recorded as a configuration, providing information regarding the co-occurrent multidimensional criteria of the observational instrument.

This research was carried out in accordance with the Declaration of Helsinki. A review by an ethics committee and written informed consents were not required in this study as: (a) it involved the observation of people in public places where individuals or groups targeted for observation had no reasonable expectation of privacy; (b) it did not include any intervention staged by the researcher or direct interaction with the individuals or groups; and (c) it did not comprise collecting personal information disseminated through photographic, film or video footage in the research results.

Data Analysis
Configurations recorded in all 40 POS were compiled as a global data set. We estimated the number of people observed counting for the number of individuals and groups of two people. When the size of groups was 3–5, 6–10, or 10–20, the number was estimated based on modal values. Regarding gender in groups, we considered that 0.75, 0.5 and 0.25% were women when the gender composition was coded as mostly women, equally mixed and mostly men, respectively.

TABLE 1 | EXOdES park use observation instrument.

Criterion	Code	Description
Observation period (TIME)	1011	10:00–11:00
	1112	11:00–12:00
	1213	12:00–13:00
	1314	13:00–14:00
	1617	16:00–17:00
	1718	17:00–18:00
	1819	18:00–19:00
	1920	19:00–20:00
Gender (GEND)	GFEM	Female group
	GMAS	Male group
	GMIX	Mixed gendered group
Age (AGE)	GCHI	Children
	GYOU	Youths
	GADU	Adults
	GELD	Elders
	ADEL	Adults with elders
	CHYO	Children and youths
	CYAE	Children and/or youths with adults and/or elders
Race/ethnicity (ETHN)	WHIT	White
	LATI	Latin/Caribbean
	ARAB	Arab
	ASIA	Asian
	AFRI	African
Activity setting (SETT)	BENC	Benches or similar
	PLAG	Playgrounds
	OPEN	Open spaces
	COUR	Sport courts
	GREE	Green areas
	SOTH	Other settings
Activity (USE)	SITT	Enjoying the scenery, chatting or relaxing
	PLAY	Playing
	WALK	Walking
	FOOT	Playing football
	PETA	Playing boules
	OSPO	Playing other sports
	PICK	Picnicking
Vehicles (VEHI)	NOVE	No vehicles
	BICY	Bicycle
	SKAT	Skate or roller skater
	BABY	Stroller
	WHEE	Wheelchair
	DRIV	Motorized vehicles

Thus, information on the behavior of more than 18,000 groups were analyzed to search for regular structures hidden in data set according to gender. Prospective and retrospective multievent sequential analysis, from lag −5 to lag +5, were performed using GSEQ 5.1 (Bakeman and Quera, 1995, 2011). We used a simplified gender composition category -groups of males only (GMAS), females only (GFEM), and gendered mixed (GMIX)- as target behaviors, considering the rest of categories in the observational instrument as given criteria. Thus, several polar coordinate analysis were performed with HOISAN (Hernández-Mendo et al., 2012) to create maps with all possible interrelations between gender composition of observed groups and all categories of the field format.

Polar coordinate is a data reduction technique based on the Z_{sum} statistic, which was introduced by Cochran (1954), developed by Sackett (1980), and optimized by Gorospe and Anguera (2000). Standardized Z statistics derived from adjusted residuals (Bakeman, 1978) were used to compute prospective and retrospective Z_{sum} statistics. These values are then used to build maps showing the relationships between a focal behavior and one or more conditional behaviors. These relationships are considered significant ($p < 0.05$) when the vector length is >1.96 (excitatory) or <−1.96 (inhibitory). Each quadrant shows the type of relationship between the focal behavior and the corresponding conditional behavior as follows (**Figure 2**): Quadrant I: prospective and retrospective activation; Quadrant II: prospective inhibition and retrospective activation; Quadrant III: prospective and retrospective inhibition; and Quadrant IV: prospective activation and retrospective inhibition. Although this technique was specifically developed for use in sport research (Gorospe and Anguera, 2000; Perea et al., 2012; Aragón et al., 2016; Castañer et al., 2016; Tarragó et al., 2016), it has been also useful in other fields (Anguera et al., 2003; Herrero Nivela and Pleguezuelos Saavedra, 2008; Santoyo et al., 2017). To our best knowledge, it is the first time that it was applied to analyse daily life interactions in public spaces.

RESULTS

Research staff completed a total of 1,505 observational sessions made on 67 different days. During the study period, we estimate that 75,853 people (55 males vs. 45% females) were using POS during observational periods. Specifically, 16,209 people were observed as being alone (66 males vs. 34% females) and we estimate that 59,644 people were recorded in groups (53 males vs. 47% females). The complete observed categories among males only (GMAS), females only (GFEM), and gendered mixed groups (GMIX) are shown in **Table 2**. For all studied criteria, the chi-square for differences among gender groups was significant at the 0.001 level.

Table 3 shows the polar coordinates analysis numerical result, considering as focal behavior GMAS, GFEM, and GMIX. It includes the following information: name of the conditional behavior, quadrant, prospective and retrospective Z_{sum}, radius, and angle. The polar coordinate maps offer a visual representation of the statistically significant associations (activation or inhibition) between focal and conditional behaviors. In the present study, only significant relations between focal and conditional behaviors are presented. The association is shown both quantitatively (length of vector) and qualitatively (quadrant I, II, III, or IV). We have structured results into sections organized by the different 6 target criteria in EXOdES that have been analyzed.
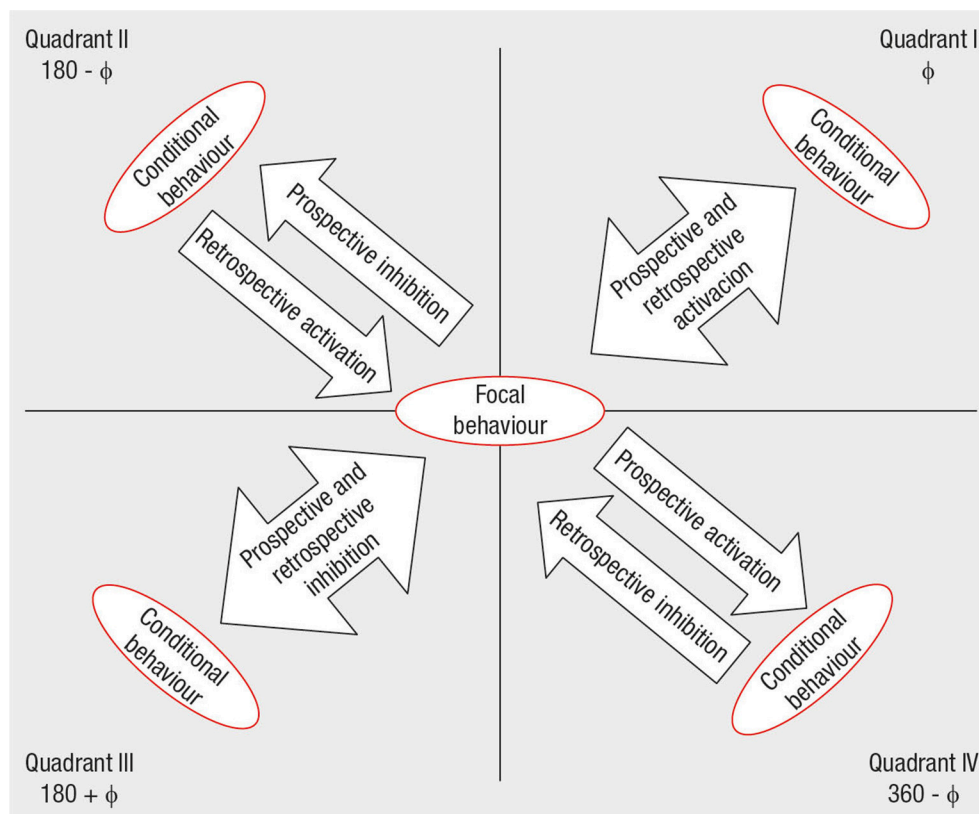
FIGURE 2 | Relationships between focal and target behaviors in a polar coordinate map (Extracted from Aragón et al., 2016, p. 5).

Time of Day

Studied POS have more capacity to attract groups of people during the afternoon, especially from 17:00 to 20:00. As shown in **Table 2**, during this observational period, 53.6% of groups were observed. In **Figure 3**, relationships between gender composition and observational periods are shown. Male groups have mutually inhibitory relationships with 17:00–18:00 (1718) and 18:00–19:00 (1819), also mutually excitatory relationships with the rest of observational periods. Contrary to men groups, female groups present mutually excitatory relationships with 12:00–13:00 (1213), 18:00–19:00, and particularly significant with 17:00–18:00, coinciding with the moment when children finish school in Spain. Regarding mixed groups, mutually excitatory relationships are found with 10:00–11:00 (1011), 18:00–19:00, and particularly stronger with 19:00–20:00 (1920).

Age

The most frequent composition group observed (36.0%) was that formed by children, youth or both, accompanied by adults, older adults or both (CYAE). This category comprises of different forms of child and youth care. After that, the most common groups were adults (23.2%), youths (18.05%), elders (10.3%), children (6.3%), adults with elders (5.0%), and children with youths (1.2%). In the second polar coordinate map, the relationship between gender and age groups are

shown. As we can see in **Figure 4**, male groups have mutually excitatory relations with all composition groups, except with groups of adults and older adults (ADEL), and particularly groups of children and/or youths with adults and/or older adults (CYAE), both with mutually inhibitory relations. Groups of adults with older adults, but particularly groups of children and/or youths supervised by adults and/or older adults are the only composition groups that are found to be mutually activated with groups of females. Regarding mixed groups, mutually excitatory associations are found with groups of youths (GYOU), adults (GADU), elders (GELD), and groups of adults with elders. Mixed gendered groups also have mutually inhibitory relationships with groups of children only (GCHI) and groups of children and/or youths supervised by adults and/or older adults.

Race/Ethnicity

Regarding race/ethnicity, most of the observed groups are Whites (77.9%), followed by Latins (12.3%), Asians (4.9%), Arabs (3.5%), and Africans (1.4%). These results are coherent with the heterogeneity of residents in the city of Barcelona, as according to the census 16.6% of its population is foreign, that being Europeans, Latins, and Asians the more common origins. The groups ethnically heterogeneous, those where whites and other minority groups are mixed, represent the 7.5% of

TABLE 2 | Observed frequency of analyzed criteria by gender composition.

	Male groups		Female groups		Mixed groups		Total	
	N	%	*N*	%	*N*	%	*N*	%
OBSERVATION PERIOD								
1011	373	6.0%	317	5.1%	349	6.0%	1,039	5.7%
1112	670	10.7%	631	10.1%	557	9.6%	1,858	10.2%
1213	751	12.0%	755	12.1%	629	10.9%	2,135	11.7%
1314	709	11.3%	621	10.0%	554	9.6%	1,884	10.3%
1617	632	10.1%	436	7.0%	495	8.6%	1,563	8.5%
1718	1,094	17.5%	1,507	24.1%	980	17.0%	3,581	19.6%
1819	1,078	17.2%	1,224	19.6%	1,137	19.7%	3,439	18.8%
1920	949	15.2%	749	12.0%	1,071	18.6%	2,769	15.2%
XSQ 257.62	df = 14	*p* = < 0.01						
AGE								
GCHI	541	8.7%	222	3.6%	394	6.8%	1,157	6.3%
GYOU	1,433	22.9%	695	11.1%	1,168	20.2%	3,296	18.0%
GADU	1,689	27.0%	925	14.8%	1,623	28.1%	4,237	23.2%
GELD	670	10.7%	428	6.9%	782	13.5%	1,880	10.3%
ADEL	170	2.7%	413	6.6%	325	5.6%	908	5.0%
CHYO	80	1.3%	67	1.1%	65	1.2%	212	1.2%
CYAE	1,666	26.7%	3,490	55.9%	1,415	24.6%	6,571	36.0%
XSQ 1,970.5	df=12	*p* = < 0.01						
RACE/ETHNICITY								
WHIT	4,368	71.5%	4,851	79.0%	4,734	83.7%	13,953	77.9%
LATI	821	13.4%	780	12.7%	591	10.5%	2,192	12.3%
ARAB	386	6.3%	155	2.5%	91	1.6%	632	3.5%
ASIA	382	6.2%	299	4.9%	198	3.5%	879	4.9%
AFRI	156	2.6%	56	0.9%	42	0.7%	254	1.4%
XSQ 421.95	df=8	*p* = < 0.01						
ACTIVITY SETTING								
BENC	2,620	41.9%	2,595	41.6%	2,865	49.6%	8,080	44.2%
PLAG	900	14.4%	1,927	30.9%	782	13.6%	3,609	19.8%
OPEN	1,733	27.7%	1,340	21.5%	1,547	26.8%	4,620	25.3%
COUR	578	9.2%	85	1.4%	129	2.2%	792	4.3%
GREE	197	3.2%	146	2.3%	237	4.1%	580	3.2%
SOTH	227	3.6%	142	2.3%	212	3.7%	581	3.2%
XSQ 1,293.17	df=10	*p* = < 0.01						
ACTIVITY								
SITT	3,453	58.9%	3,497	56.3%	3,889	69.3%	10,839	61.3%
PLAY	950	16.2%	1,826	29.5%	794	14.1%	3,570	20.2%
WALK	331	5.6%	623	10.0%	646	11.5%	1,600	9.0%
FOOT	427	7.3%	39	0.6%	66	1.2%	532	3.0%
PETA	163	2.8%	13	0.2%	32	0.6%	208	1.2%
OSPO	403	6.8%	72	1.2%	89	1.6%	564	3.2%
PICK	140	2.4%	139	2.2%	98	1.7%	377	2.1%
XSQ 1,729.23	df=12	*p* = < 0.01						
VEHICLES								
NOVE	5,125	82.0%	4,345	69.6%	4,696	81.4%	14,166	77.5%
BICY	275	4.4%	116	1.9%	150	2.6%	541	3.0%
SKAT	186	3.0%	67	1.1%	55	1.0%	308	1.7%
BABY	376	6.0%	1,506	24.1%	574	10.0%	2,456	13.5%
WHEE	96	1.5%	188	3.0%	203	3.5%	487	2.7%
DRIV	195	3.1%	17	0.3%	88	1.5%	300	1.7%
XSQ 1,285.15	df=10	*p* = < 0.01						

Only those categories accounting for more than 1.0% have been included; those with less than 1.0% were aggregated in other recoded or eliminated of the analysis.

TABLE 3 | Polar coordinate analysis of studied criteria considering gender composition the focal behavior.

Code	Male groups (GMAS)					Female groups (GFEM)					Mixed groups (GMIX)				
	Quadr.	Prosp.	Retrosp.	Radius	Angle	Quadr.	Prosp.	Retrosp.	Radius	Angle	Quadr.	Prosp.	Retrosp.	Radius	Angle
TIME_1011	I	4.75	2.58	5.41 (*)	28.45	III	−9.07	−4.71	10.23 (*)	207.45	I	4.4	2.17	4.91 (*)	26.28
TIME_1112	I	3.51	4.31	5.56 (*)	50.84	IV	1.85	−0.49	1.91	345.09	III	−5.47	−3.9	6.72 (*)	215.52
TIME_1213	I	1.37	1.44	1.99 (*)	46.46	I	2.34	2.52	3.44 (*)	47.16	III	−3.78	−4.05	5.54 (*)	226.93
TIME_1314	I	7.1	7.38	10.24 (*)	46.1	III	−3.16	−3.13	4.45 (*)	224.76	III	−4.03	−4.34	5.93 (*)	227.11
TIME_1617	I	11	12.08	16.34 (*)	47.68	III	−11.98	−10.63	16.02 (*)	221.59	IV	0.99	−1.48	1.78	303.85
TIME_1718	III	−10.95	−13.63	17.48 (*)	231.21	I	24.84	24.54	34.92 (*)	44.66	III	−14.16	−11.13	18.01 (*)	218.17
TIME_1819	III	−10.04	−9.21	13.63 (*)	222.51	I	5.03	3.51	6.13 (*)	34.91	I	5.13	5.82	7.76 (*)	48.6
TIME_1920	I	0.98	2.7	2.87 (*)	69.96	III	−18.26	−18.92	26.29 (*)	226.02	I	17.61	16.54	24.16 (*)	43.21
AGE_GCHI	I	10.88	10.74	15.29 (*)	44.62	III	−10.08	−8.52	13.2 (*)	220.2	III	−0.75	−2.12	2.25 (*)	250.52
AGE_GYOU	I	15.43	14.55	21.21 (*)	43.31	III	−22.69	−22.85	32.2 (*)	225.2	I	7.7	8.8	11.69 (*)	48.83
AGE_GADU	I	20.57	22.97	30.83 (*)	48.15	III	−32.83	−33.54	46.93 (*)	225.61	I	12.96	11.28	17.18 (*)	41.05
AGE_GELD	I	3.4	5.62	6.57 (*)	58.84	III	−11.82	−13.99	18.32 (*)	229.81	I	8.81	8.73	12.41 (*)	44.75
AGE_ADEL	III	−8.74	−9.33	12.78 (*)	226.87	I	5.23	4.5	6.9 (*)	40.74	I	3.59	4.83	6.02 (*)	53.39
AGE_CHYO	I	1.8	1.74	2.5 (*)	44.06	III	−1.39	−4.06	4.29 (*)	251.09	II	−0.41	2.42	2.46 (*)	99.63
AGE_CYAE	III	−33.95	−36.27	49.68 (*)	226.89	I	56.2	58.38	81.04 (*)	46.09	III	−23.49	−23.42	33.17 (*)	224.92
ETHN_WHIT	III	−21.17	−17.91	27.73 (*)	220.23	I	4.8	2.83	5.58 (*)	30.51	I	16.75	15.38	22.74 (*)	42.55
ETHN_LATI	I	7.45	3.97	8.44 (*)	28.07	I	4.62	5.17	6.94 (*)	48.19	III	−12.35	−9.33	15.48 (*)	217.06
ETHN_ARAB	I	16.83	14.97	22.53 (*)	41.66	III	−9.14	−6.82	11.4 (*)	216.74	III	−7.88	−8.32	11.46 (*)	226.53
ETHN_ASIA	I	6.13	5.8	8.44 (*)	43.43	III	−3.19	−1.78	3.65 (*)	209.17	III	−3.01	−4.11	5.09 (*)	233.71
ETHN_AFRI	I	16.05	17.46	23.71 (*)	47.42	III	−8.99	−9.73	13.25 (*)	227.26	III	−7.22	−7.88	10.69 (*)	227.49
SETT_PLAG	III	−20.76	−25.7	33.03 (*)	231.07	I	43.7	45.38	63 (*)	46.08	III	−23.45	−20.08	30.87 (*)	220.58
SETT_GREE	III	−3.6	−3.72	5.17 (*)	225.95	III	−6.03	−5.59	8.22 (*)	222.86	I	9.84	9.49	13.67 (*)	43.98
SETT_BENC	I	1.67	3.12	3.54 (*)	61.78	III	−11.48	−13.11	17.43 (*)	228.78	I	10.04	10.2	14.31 (*)	45.44
SETT_COUR	I	28.94	28.71	40.77 (*)	44.77	III	−18.95	−15.14	24.26 (*)	218.63	III	−10.22	−13.84	17.21 (*)	233.55
SETT_OPEN	I	4.44	5.89	7.38 (*)	52.98	III	−11.35	−11.85	16.4 (*)	226.24	I	7.07	6.08	9.32 (*)	40.7
SETT_SOTH	I	1.52	4.9	5.13 (*)	72.8	III	−5.79	−8.27	10.1 (*)	235.03	I	4.37	3.44	5.57 (*)	38.21
USE_SITT	I	3.56	5.06	6.19 (*)	54.83	III	−20.81	−21.43	29.88 (*)	225.84	I	17.73	16.8	24.42 (*)	43.45
USE_PLAY	III	−15.51	−21.08	26.17 (*)	233.66	I	39.44	40.33	56.41 (*)	45.63	III	−24.7	−19.87	31.7 (*)	218.82
USE_WALK	III	−12.72	−8.29	15.19 (*)	213.09	IV	0.48	−2.16	2.21 (*)	282.52	I	12.47	10.62	16.38 (*)	40.41
USE_OSPO	I	17.9	16.97	24.66 (*)	43.48	III	−12.92	−9.73	16.17 (*)	216.98	III	−4.96	−7.27	8.8 (*)	235.66
USE_FOOT	I	18.97	17.24	25.63 (*)	42.26	III	−10.03	−7.84	12.73 (*)	218.01	III	−9.02	−9.46	13.07 (*)	226.37
USE_PICK	I	2.32	3.05	3.84 (*)	52.77	III	−0.81	−1.2	1.45	235.97	III	−1.54	−1.87	2.42 (*)	230.55
USE_PETA	I	17.59	20.59	27.08 (*)	49.49	III	−10.8	−11.61	15.86 (*)	227.07	III	−6.82	−9	11.29 (*)	232.86
VEHI_NOVE	I	18.18	20.28	27.23 (*)	48.12	III	−31.11	−30.98	43.91 (*)	224.88	I	13.23	10.94	17.17 (*)	39.59
VEHI_BICY	I	5.3	6.18	8.14 (*)	49.41	III	−7.4	−6.51	9.85 (*)	221.32	I	2.16	0.33	2.18 (*)	8.73
VEHI_BABY	III	−30.86	−32.67	44.94 (*)	226.63	I	47.58	46.48	66.52 (*)	44.33	III	−17.11	−14.12	22.18 (*)	219.54
VEHI_WHEE	III	−7.08	−8.62	11.16 (*)	230.58	I	3.93	3.08	4.99 (*)	38.09	I	3.22	5.65	6.5 (*)	60.28
VEHI_DRIV	I	14.11	12.79	19.05 (*)	42.2	III	−12.33	−10.92	16.46 (*)	221.53	III	−1.83	−1.91	2.64 (*)	226.1
VEHI_SKAT	I	7.77	7.24	10.62 (*)	42.99	III	−4.37	−3.23	5.44 (*)	216.47	III	−3.47	−4.09	5.36 (*)	229.7

*Significant relationships (p < 0.05) between the focal behavior and conditional behaviors.

observed groups. In **Figure 5**, we show the relationships that have been found between gender and race/ethnicity. Male groups have mutually excitatory connections with all minority groups, particularly stronger with Africans (AFRI) and Arabs (ARAB). Whites (WHIT) is the only category with which male groups establish a strong mutually inhibitory relationship. Female groups, on the other hand, have mutually excitatory relations with Latins (LATI) and Whites, also mutually inhibitory relationships with Asians (ASIA) and particularly Arabs and

Africans. Mixed gendered groups have a mutually excitatory relationship only with Whites and mutually inhibitory relations with the rest of minority groups.

Activity Setting
The most heavily used activity settings were those where people can sit such as benches, little walls or stairs (44.2%), followed by open spaces (25.3%), playgrounds (19.8%), sport courts (4.3%), and green areas (3.2%). **Figure 6** examines the kind of relations
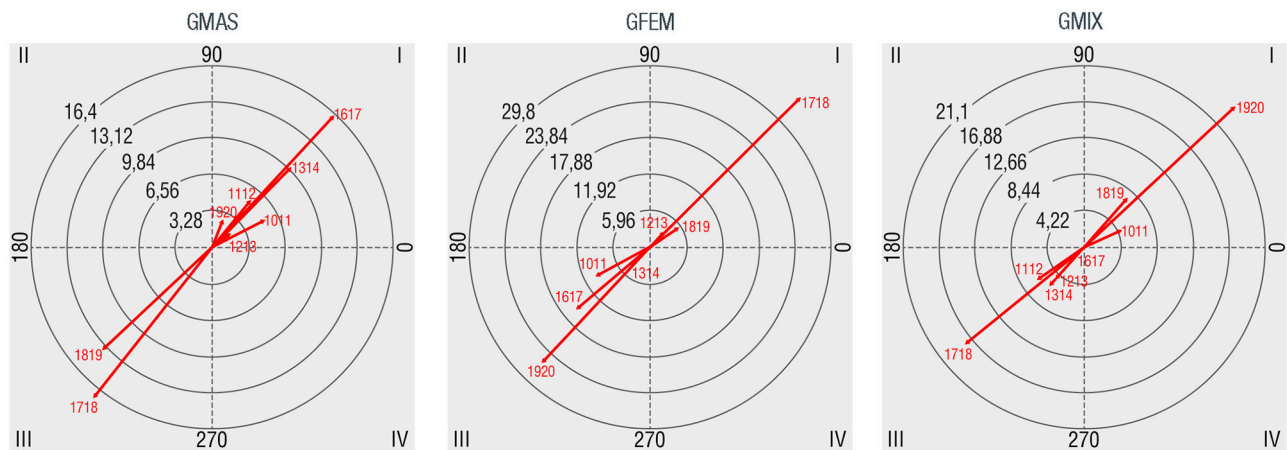
FIGURE 3 | Polar coordinate maps considering observational periods as target behavior (1011: 10:00–11:00, 1112: 11:00–12:00, 1213: 12:00–13:00, 1314: 13:00–14:00, 1617: 16:00–17:00, 1718: 17:00–18:00, 1819: 18:00–19:00, 1920: 19:00–20:00).
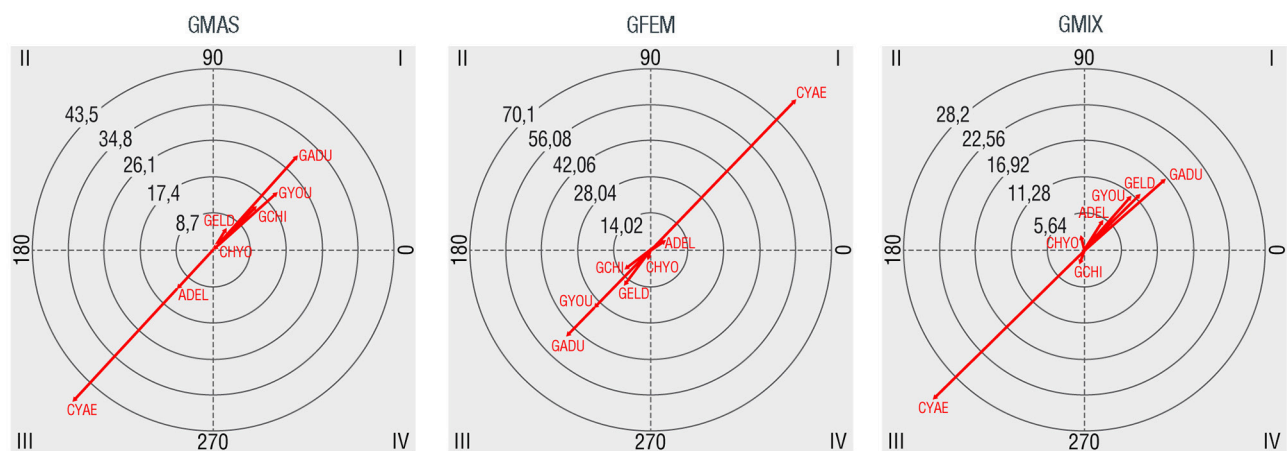
FIGURE 4 | Polar coordinate maps considering age composition of groups as target behavior (GCHI: children, GYOU: youths, GADU: adults, GELD: elders, ADEL: adults with elders, CHYO: children with youths, CYAE: children and/or youths with adults and/or elders).

established between activity setting and gender group. Male groups have mutually excitatory connections with benches, little walls or stairs (BENC), open spaces (OPEN), and particularly with sport courts (COUR). Green spaces (GREE) and specially playgrounds (PLAG) have mutually inhibitory connections with male groups. On the contrary, playground is the only activity setting that has a mutually excitatory relationship with female groups. Regarding mixed groups, they have mutually excitatory connections with all activity settings except with sport courts and playgrounds.

Activity
The most common activities observed were sitting or chatting (61.3%), followed by playing (20.2%), walking (9.0%), playing sports (7.4%), and picnicking (2.1%). Most frequently observed sports were football (3.0%) and boules (1.2%). Relationships that have been detected between the activity and gender are shown

in **Figure 7**. Male groups have mutually excitatory relationships with sitting/chatting (SITT) and picnicking (PICK), but the strongest relationships are established with playing sports as football (FOOT), boules (PETA) or others (OSPO). Regarding female groups, the only activity that is mutually activated is that related to game (PLAY) activities. In mixed gendered groups, activities that are found mutually activated are sitting/chatting and walking (WALK).

Vehicles
The analysis of vehicles is a complementary way of describing park use. From observed groups, 22.4% were carrying some type of vehicle, stroller being the most frequent (13.5%) followed by bicycles (3.0%), wheel chairs (2.7%), skates or roller skaters (1.7%), and other motorized vehicles as cars or motorcycle (1.7%). **Figure 8** shows the type of relationships found between gender groups and vehicles. Male groups

FIGURE 5 | Polar coordinate maps considering race/ethnicity composition group as target behavior (WHIT: whites, LATI: latins, ARAB: arabs, ASIA: asians, AFRI: africans).

FIGURE 6 | Polar coordinate maps considering location where the activity occurs as target behavior (BENC: benches or similar, PLAG: playground, OPEN: open space, COUR: sport courts, GREE: green areas, SOTH: other settings such as fountains or parking lots).

have mutually excitatory relations with no vehicles (NOVE), motorized vehicles (DRIV), skates (SKAT) and bicycles (BICY), and mutually inhibitory relationships with wheelchairs (WHEE) and particularly with baby carriages (BABY). Female groups have mutually excitatory relations with strollers and wheelchairs, also mutually inhibitory relationships with the rest of vehicles. Finally, mixed gendered groups have mutually excitatory relationships with bicycles, wheelchairs and particularly with no vehicles.

DISCUSSION

Public open spaces (POS) play a critical role in urban areas offering free opportunities for leisure and physical activity. They also help to increase social recognition and interaction with neighbors, which is the basis to improve social cohesion, trust, and perception of safety. Nevertheless, few studies have used systematic observation to analyse activity patterns on POS except

for some recent studies on outdoor physical activity mainly in the United States.

We have conducted a systematic observation study over 3 months, observing 40 POS distributed across all 10 districts of Barcelona to assess gendered differences on park use. An *ad hoc* observational instrument (EXOdES) was used to record sociodemographic characteristics of park users and their activities. In this paper, we estimated numbers of people using POS and analyzed gendered differences on several criteria: time of day, age group, race/ethnicity, activity setting, activity, and vehicles.

According to the census, less males than females (47 vs. 53%) live in Barcelona. Nevertheless, more males than females (55 vs. 45%) regardless of age group were seen using POS. The difference was particularly higher in individuals (66 vs. 34%) rather than in groups, where numbers were more similar (53 vs. 47%). This result is consistent with several previous studies
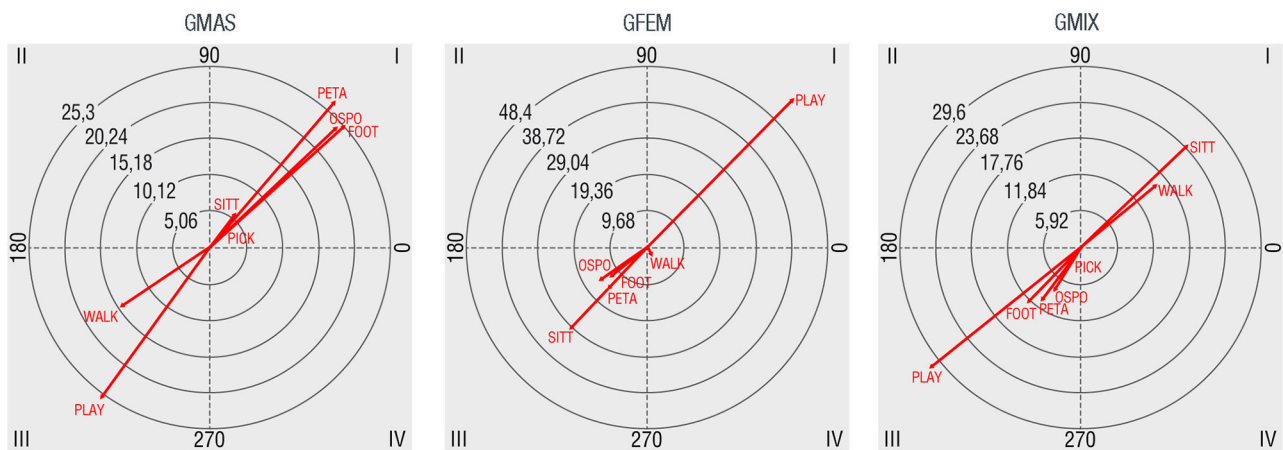
FIGURE 7 | Polar coordinate maps considering activity as target behavior (SITT: just enjoying the scenery, chatting or relaxing, PLAY playing, WALK: walking, FOOT: playing football, PETA: playing boules, OSPO: playing other sports such as volleyball, PICK: picnicking).
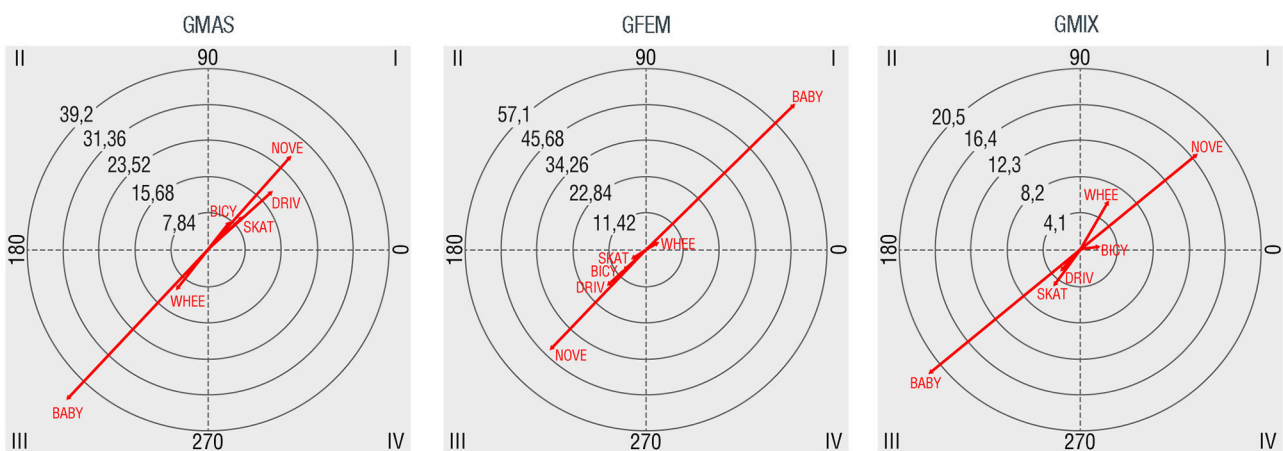
FIGURE 8 | Polar coordinate maps considering vehicles use as target behavior (NOVE: vehicle absence, BICY: bicycle, SKAT: skate or roller skates, BABY: stroller, WHEE: wheelchair, DRIV: motorized vehicles such as cars or motorcycles).

in other geographic areas. In a recent review of observational studies measuring physical activity levels on park users, Evenson et al. (2016) found that in 20 studies more males than females were observed using public parks, ranging from 51 to 67%, while just three of them reported fewer males or no gender differences. Thus, a considerable higher proportion of females were seen in the present study when compared to several of those included in Evenson's review. Regarding groups of the same age, more adults, and adolescents were seen than older adults and children, which is also consistent with literature. However, the most frequent mixed age group composition in our study was adults and/or older adults with children and/or youngs (36%), thus a considerable proportion of elders and children were observed using POS.

To particularly assess differences in the way that men and women use public space, data regarding to more than 18,000 groups of people have been analyzed considering their gender composition (males only, females only and mixed gendered groups). Using multievent sequential and polar coordinate analysis, several hidden patterns in dataset have been identified. Groups of women were more likely to use POS between 17:00 and 18:00 (after children finish school) rather than at other times of the day. They were particularly infrequent between 19:00 and 20:00 unless they were with other men.

Contrary to men, women were more frequently involved in groups with other children, adolescents, and elders, rather than with other women of their same age group. Women were also more likely to be close to playgrounds areas where they could supervise children, to be engaged in play activities with them, and to be seen with strollers, rather than any other amenities, activities or vehicles. All these results show that outdoor leisure of women is largely centered around traditional family roles as they spend more time with children, elderly or disabled relatives (Hutchison, 1994; Kavanagh et al., 2006). These care functions

were rarely seen in groups of males and mixed groups, according to codes that were mutually activated and inhibited in polar coordinate maps.

On the other hand, male groups were more likely to be observed at any time of day except from 17:00 to 19:00. Men were more often found amongst people of the same age group, using considerably more activity settings than women (e.g., sport courts, benches, and open spaces) and performing activities such as enjoying the scenery, chatting, relaxing, and picnicking. Consistently with literature, a higher proportion of men were engaged in moderate and vigorous activities, as playing football, basketball, boules, skating, or bicycling (McKenzie et al., 2006; Reed et al., 2008; Parra et al., 2010). From a health perspective, women's constraints on outdoor physical activity is of particular concern due to the important benefits on health indicators. A policy challenge is how to engage more women in sports while simultaneously supplying other sources of care for their young children (Cohen et al., 2007).

Two important questions arise when considering the role of race/ethinicity on park use. Firstly, a considerable fewer proportion of Asian, Arab, and African women were seen compared with groups of men of the same ethnic group. When female groups were observed, they were more likely to be Whites or Latinas rather than any other origin. However, groups of men have mutually excitatory relationships with all minority groups and an inhibitory relation just with Whites. These results reflect the exclusion of public space that many women from minority ethnic groups experience, which is coherent with previous ethnographies conducted in Barcelona (Garcia-Ramon et al., 2004; Ortiz et al., 2004). Research has shown that women from minority ethnic groups may have some specific constraints for park use, including a higher fear of sexual and racial attack, differences in roles and rights by gender as a result of more patriarchal structures (Ho et al., 2005), restrictions related to matters of honor especially on Muslim women (Peters, 2011), a socio-economic situation that decreases the importance of leisure pursuits and a "fear of dogs" mainly associated with religious reasons (Rishbeth, 2001). Secondly, while POS was frequented by a range of different ethnical groups, just 7.5% of observed groups were seen having contact between them. Ethic segregation may be highly functional for some groups when segregation is voluntary. Ethnic minorities "frequently want to be together in order to enjoy mutual support, rebuild family and neighborhood networks, and maintain their languages and cultures" (Castles, 1993). However, more efforts are needed to encourage informal social contact in POS between different ethnic groups. Promoting heterogeneity, tolerance and inter-ethnic understanding have also been linked with social cohesion and perception of safety (Vargas and Merino, 2012).

In the light of above exposed, a final reflection about the social quality of POS could be made. One of the most important consequences of fear of crime is the withdrawal of people from public spaces, especially vulnerable social groups (Jackson, 2011; Rader et al., 2012; Shippee, 2012). Fear of crime can make people prisoners at their own home (Hale, 1996). People who are afraid of being criminally victimized tend to stay more at home, limiting their social and cultural activities, reducing the quality of life,

and eroding social life. Additionally, limiting one's movement to safe places at safe times may have a feedback loop: limiting social interaction also increases fear in its turn (Liska et al., 1988), whereas, experiencing ethnic and social variety regularly may help to develop a sense of familiarity with strangers, reduce intolerance and increase social cohesion, perceptions of safety and well-being (Kazmierczak, 2013). From an urban ecological perspective (Saunders, 2001) social diversity has a great relevance on urban social management. For instance, Hristova et al. (2016) considers "brokerage" (or social connectivity), "serendipity,", "entropy," and "homogeneity" as mesures of social diversity. Indeed, POS should provoque spontaneous and unexpected social encounters, as well as those planned and trusted. Because POS are the main scenarios for urban social life, contact with strangers–viewed as an opportunity, not as a risk–should be psychosocially enriching, and a tool for pomoting social cohesion. In many cities, as we have also seen in Barcelona, too many POS are places appropiated by specific social groups in specific periods of time. This is particularly dramatical when we have analised gender patterns of occupation specially related to female traditional roles as well as cultural ethnic differences. Thus, considered, it is only a matter of time that POS will end up loosing their social relevance in favor of other more controlled and safer places. Conversely, a higher interest in promoting social diversity in a perceived safety environments could break off this tendency, now broadly extended in many urban environments (Low, 2003).

Some limitations of this study have been identified. Probably, the most important one was that observations were conducted only on weekdays from September to December. Thus, any conclusion about the activity patterns in studied public spaces should be restricted to this observational period. It would be essential to examine POS during weekends, as gendered patterns of public space use may be different, also during other times of the year to identify seasonal changes on park use. A second limitation included the sample selection bias. The sampling consisted of 40 POS in the city of Barcelona. At least 2 public spaces of all 10 districts of the city were represented to try to avoid an important bias. However, as 20 of them were concentrated in Sants-Montjuïc, results needed to be interpreted carefully. Additionally, as with most studies using systematic observation there was the possibility of generating reactance on park users. To minimize this bias, observers had instructions of being in locations where low visibility to park users were guaranteed. Although very few people respond with curiosity, there was an episode where the observer was asked to stop recording and leave, reflecting appropriation processes of public space by certain communities that characterize some places.

This study is an example of the possibilities that systematic observation offers for the study of naturally occurred interactions in everyday life. We have also shown the informative potential of polar coordinate technique when analyzing big observational data with results in form of easy-to-understand maps. Our results have documented men and women preferences on park use, in unisex, and mixed groups. Together, these findings can help urban planners and policy-makers to assess and address specific gender needs associated with environmental justice. The

approach can also provide relevant data to decide which parks need interventions or to examine the impact of park renovations on park use. Further research could also consider assessing social and environmental characteristics of POS and their implications on activity patterns and perceived insecurity.

ETHICS STATEMENT

This study was carried out in accordance with the Declaration of Helsinki. Systematic observation was performed anonymously.

AUTHOR CONTRIBUTIONS

SV and FP-T developed the project and the design of the study. MTA supervised the methods. FP-T coordinated data collection and did the data analysis, closely supervised by MTA. FP-T did the writing of the article that was critically review by MTA and SV. All authors approved the final, submitted version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Anguera, M. T. (2003). "Observational methods (general)," in *Encyclopedia of Psychological Assessment,* Vol. 2, ed R. Fernández-Ballesteros (London: Sage), 632–637.

Anguera, M. T., and Hernández-Mendo, A. (2013). La metodología observacional en el ámbito del deporte [Observational methodology in sport sciences]. *E Balonmano.Com J. Sport. Sci.* 9, 135–160.

Anguera, M. T., Santoyo, C., and Espinosa, M. C. (2003). "Evaluating links intensity in social networks in a school context through observational designs," in *Culture, Environmental Action, and Sustainability*, eds R. Garcia Mira, J. M. Camaselle, and J. R. Martinez (Göttingen: Hogrefe Publishing), 286–298.

Aragón, S., Lapresa, D., Arana, J., Anguera, M. T., and Garzón, B. (2016). An example of the informative potential of polar coordinate analysis: sprint tactics in elite 1,500-m track events. *Meas. Phys. Educ. Exerc. Sci.* 21, 26–33. doi: 10.1080/1091367X.2016.1245192

Babey, S. H., Tan, D., Wolstein, J., and Diamant, A. L. (2015). Neighborhood, family and individual characteristics related to adolescent park-based physical activity. *Prev. Med.* 76, 31–36. doi: 10.1016/j.ypmed.2015.04.001

Bakeman, R. (1978). "Untangling streams of behavior: sequential analysis of observation data," in *Observing Behavior,* Vol. 2: *Data Collection And Analysis Methods*, ed G. P. Sackett (Baltimore, MD: University of Park Press), 63–78.

Bakeman, R., and Quera, V. (1995). *Analyzing Interaction: Sequential Analysis with SDIS and GSEQ*. Cambridge: Cambridge University Press.

Bakeman, R., and Quera, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences*. Cambridge: Cambridge University Press.

Blanco-Villaseñor, Á., Losada, J. L., and Anguera, M. T. (2003). Data analysis techniques in observational designs applied to the environment-behaviour relation. *Medio Ambient. y Comport. Hum.* 4, 111–126.

Bonaiuto, M., Mao, Y., Roberts, S., Psalti, A., Ariccio, S., Ganucci Cancellieri, U., et al. (2016). Optimal experience and personal growth: flow and the consolidation of place identity. *Front. Psychol.* 7:1654. doi: 10.3389/fpsyg.2016.01654

Burgess, J., Harrison, C. M., and Limb, M. (1988). People, parks and the urban green: a study of popular meanings and values for open spaces in the city. *Urban Stud.* 25, 455–473. doi: 10.1080/00420988820080631

Casper, J. M., Harrolle, M. G., and Kelley, K. (2013). Gender differences in self-report physical activity and park and recreation facility use among Latinos in Wake County, North Carolina. *Ann. Behav. Med.* 45, S49–S54. doi: 10.1007/s12160-012-9435-9

Castañer, M., Barreira, D., Camerino, O., and Anguera, M. T. (2016). Goal scoring in soccer : a polar coordinate analysis of motor skills used by Lionel Messi goal. *Front. Psychol.* 7:806. doi: 10.3389/fpsyg.2016.00806

Castles, S. (1993). *Migrant incorporation in highly developed countries: an international comparison*. Available online at: http://ro.uow.edu.au/cmsocpapers/23

Cattell, V., Dines, N., Gesler, W., and Curtis, S. (2008). Mingling, observing, and lingering: everyday public spaces and their implications for well-being and social relations. *Health Place* 14, 544–561. doi: 10.1016/j.healthplace.2007.10.007

Chiesura, A. (2004). The role of urban parks for the sustainable city. *Landsc. Urban Plan.* 68, 129–138. doi: 10.1016/j.landurbplan.2003.08.003

Cochran, W. G. (1954). Some methods for strengthening the common $\chi2$ tests. *Biometrics* 10, 417–451. doi: 10.2307/3001616

Cohen, D. A., Golinelli, D., Williamson, S., Sehgal, A., Marsh, T., and McKenzie, T. L. (2009). Effects of park improvements on park use and physical activity. *Am. J. Prev. Med.* 37, 475–480. doi: 10.1016/j.amepre.2009.07.017

Cohen, D. A., McKenzie, T. L., Sehgal, A., Williamson, S., Golinelli, D., and Lurie, N. (2007). Contribution of public parks to physical activity. *Am. J. Public Health* 97, 509–514. doi: 10.2105/AJPH.2005.072447

Coley, R. L., Kuo, F. E., and Sullivan, W. C. (1997). Where does community grow? the social context created by nature in urban public housing. *Environ. Behav.* 29, 468–494. doi: 10.1177/001391659702900402

Collado, S., and Staats, H. (2016). Contact with nature and children's restorative experiences: an eye to the future. *Front. Psychol.* 7:1885. doi: 10.3389/fpsyg.2016.01885

Dahmann, N., Wolch, J., Joassart-Marcelli, P., Reynolds, K., and Jerrett, M. (2010). The active city? disparities in provision of urban public recreation resources. *Health Place* 16, 431–445. doi: 10.1016/j.healthplace.2009.11.005

Derose, K. P., Han, B., Williamson, S., and Cohen, D. A. (2018). Gender disparities in park use and physical activity among residents of high-poverty neighborhoods in Los Angeles. *Womens Health. Issues* 28, 6–13. doi: 10.1016/j.whi.2017.11.003

Evenson, K. R., Jones, S. A., Holliday, K. M., Cohen, D. A., and McKenzie, T. L. (2016). Park characteristics, use, and physical activity: a review of studies using SOPARC (System for Observing Play and Recreation in Communities). *Prev. Med.* 86, 153–166. doi: 10.1016/j.ypmed.2016.02.029

Franklin, T. W., Franklin, C. A., and Fearn, N. E. (2008). A multilevel analysis of the vulnerability, disorder, and social integration models of fear of crime. *Soc. Justice Res.* 21, 204–227. doi: 10.1007/s11211-008-0069-9

Garcia-Ramon, M. D., Ortiz, A., and Prats, M. (2004). Urban planning, gender and the use of public space in a peripherial neighbourhood of Barcelona. *Cities* 21, 215–223. doi: 10.1016/j.cities.2004.03.006

Godbey, G. (2009). *Outdoor Recreation, Health, and Wellness: Understanding and Enhancing the Relation.* Washington, DC: Resources for the Future.

Gorospe, G., and Anguera, M. T. (2000). Modificación de la técnica clásica de coordenadas polares mediante un desarrollo distinto de la retrospectividad: aplicación al tenis [Retrospectivity in polar coordinates analysis: application to tennis]. *Psicothema* 12, 279–282.

Hale, C. (1996). Fear of crime: a review of the literature. *Int. Rev. Vict.* 4, 79–150. doi: 10.1177/026975809600400201

Hansmann, R., Hug, S. M., and Seeland, K. (2007). Restoration and stress relief through physical activities in forests and parks. *Urban Forest. Urban Green.* 6, 213–225. doi: 10.1016/j.ufug.2007.08.004

Hartig, T., Mang, M., and Evans, G. W. (1991). Restorative effects of natural environment experiences. *Environ. Behav.* 23, 3–26. doi: 10.1177/0013916591231001

Hernández-Mendo, A., López-López, J. A., Castellano Paulis, J., Morales-Sánchez, V., and Pastrana, J. L. (2012). HOISAN 1.2: programa informático para uso en metodología observacional [I.T. *Program for Use in Methodology Observacional*]. *Cuad. Psicol. Del Deporte* 12, 55–78. doi: 10.4321/S1578-84232012000100006

Herrero Nivela, M. L., and Pleguezuelos Saavedra, C. S. (2008). Patrones de conducta interactiva en contexto escolar multicultural [Patterns of interactive behavior in a multicultural school context]. *Psicothema* 20, 945–950.

Ho, C., Sasidharan, V., Elmendorf, W., Willits, F. K., Graefe, A., and Godbey, G. (2005). Gender and ethnic variations in urban park preferences, visitation, and perceived benefits. *J. Leis. Res.* 37, 281–306. doi: 10.1080/00222216.2005.11950054

Hoeben, E. M., Steenbeek, W., and Pauwels, L. J. R. (2018). Measuring disorder: observer bias in systematic social observations at streets and neighborhoods. *J. Quant. Criminol.* 34, 221–249. doi: 10.1007/s10940-016-9333-6

Hristova, D., Williams, M. J., Musolesi, M., Panzarasa, P., and Mascolo, C. (2016). "Measuring urban social diversity using interconnected geo-social networks," in *Proceedings of the 25th International Conference on World Wide Web (WWW)*, 21–30.

Hull, R. B., and Michael, S. E. (1995). Nature-based recreation, mood change, and stress restoration. *Leis. Sci.* 17, 1–14. doi: 10.1080/01490409509513239

Hutchison, R. (1994). Women and the elderly in Chicago public parks. *Leis. Sci.* 16, 229–247. doi: 10.1080/01490409409513234

Jackson, E. L., and Henderson, K. A. (1995). Gender-based analysis of leisure constraints. *Leis. Sci.* 17, 31–51. doi: 10.1080/01490409509513241

Jackson, J. (2011). Revisiting risk sensitivity in the fear of crime. *J. Res. Crime Delinq.* 48, 513–537. doi: 10.1177/0022427810395146

Kaplan, S. (1995). The restorative benefits of nature: toward an integrative framework. *J. Environ. Psychol.* 15, 169–182. doi: 10.1016/0272-4944(95)90001-2

Kavanagh, A. M., Bentley, R., Turrell, G., Broom, D. H., and Subramanian, S. V. (2006). Does gender modify associations between self rated health and the social and economic characteristics of local environments? *J. Epidemiol. Community Health* 60, 490–495. doi: 10.1136/jech.2005.043562

Kazmierczak, A. (2013). The contribution of local parks to neighbourhood social ties. *Landsc. Urban Plan.* 109, 31–44. doi: 10.1016/j.landurbplan.2012.05.007

Krenichyn, K. (2004). Women and physical activity in an urban park: enrichment and support through an ethic of care. *J. Environ. Psychol.* 24, 117–130. doi: 10.1016/S0272-4944(03)00053-7

Kuo, F. E., Bacaicoa, M., and Sullivan, W. C. (1998). Transforming inner-city landscapes: trees, sense of safety, and preference. *Environ. Behav.* 30, 28–59. doi: 10.1177/0013916598301002

Lapham, S. C., Cohen, D. A., Han, B., Williamson, S., Evenson, K. R., McKenzie, T. L., et al. (2016). How important is perception of safety to park use? a four-city survey. *Urban Stud.* 53, 2624–2636. doi: 10.1177/0042098015592822

Liska, A. E., Sanchirico, A., and Reed, M. D. (1988). Fear of crime and constrained behavior specifying and estimating a reciprocal effects. *Soc. Forces* 66, 827–837. doi: 10.1093/sf/66.3.827

Low, S. (2003). *Behind the Gates: Life, Security and the Pursuit of Happiness in Fortress America.* New York, NY; London: Routledge.

McCormack, G. R., Rock, M., Toohey, A. M., and Hignell, D. (2010). Characteristics of urban parks associated with park use and physical activity: a review of qualitative research. *Health Place* 16, 712–726. doi: 10.1016/j.healthplace.2010.03.003

McKenzie, T. L., Cohen, D. A., Sehgal, A., Williamson, S., and Golinelli, D. (2006). System for observing play and recreation in communities (SOPARC): reliability and feasibility measures. *J. Phys. Act. Heal.* 3, S208–S222. doi: 10.1123/jpah.3.s1.s208

McKenzie, T. L., and van der Mars, H. (2015). Top 10 research questions related to assessing physical activity and its contexts using systematic observation. *Res. Q. Exerc. Sport* 86, 13–29. doi: 10.1080/02701367.2015.991264

Molnar, B. E., Gortmaker, S. L., Bull, F. C., and Buka, S. L. (2004). Unsafe to play? neighborhood disorder and lack of safety predict reduced physical activity among urban children and adolescents. *Am. J. Heal. Promot.* 18, 378–386. doi: 10.4278/0890-1171-18.5.378

Ortiz, A., Garcia-Ramon, M. D., and Prats, M. (2004). Women's use of public space and sense of place in the Raval (Barcelona). *GeoJournal* 61, 219–227. doi: 10.1007/s10708-004-3539-0

Parra, D. C., McKenzie, T. L., Ribeiro, I. C., Hino, A. A. F., Dreisinger, M., Coniglio, K., et al. (2010). Assessing physical activity in public parks in Brazil using systematic observation. *Am. J. Public Health* 100, 1420–1426. doi: 10.2105/AJPH.2009.181230

Perea, A., Castellano, J., Alday, L., and Hernández-Mendo, A. (2012). Analysis of behaviour in sports through polar coordinate analysis with MATLAB®. *Qual. Quant.* 46, 1249–1260. doi: 10.1007/s11135-011-9435-z

Pérez Tejera, F. (2012). Differences between users of six public parks in Barcelona depending on the level of perceived safety in the neighborhood. *Athenea Digital* 12, 55–66. doi: 10.5565/rev/athenead/v12n1.930

Pérez-Tejera, F., Valera, S., and Anguera, M. T. (2011). Un nuevo instrumento para la identificación de patrones de ocupación espacial [A new instrument to identify spatial occupancy patterns]. *Psicothema* 23, 858–863.

Perkins, D. D., and Taylor, R. B. (1996). Ecological assessments of community disorder: their relationship to fear of crime and theoretical implications. *Am. J. Community Psychol.* 24, 63–107. doi: 10.1007/BF02511883

Peterlik, M., and Cross, H. S. (2005). Vitamin, D, and calcium deficits predispose for multiple chronic diseases. *Eur. J. Clin. Invest.* 35, 290–304. doi: 10.1111/j.1365-2362.2005.01487.x

Peters, K. (2011). *Living Together in Multi-Ethnic Neighbourhoods: the Meaning of Public Spaces for Issues of Social Integration.* Wageningen: Wageningen Academic Pub.

Rader, N. E., Cossman, J. S., and Porter, J. R. (2012). Fear of crime and vulnerability: using a national sample of Americans to examine two competing paradigms. *J. Crim. Justice* 40, 134–141. doi: 10.1016/j.jcrimjus.2012.02.003

Reed, J. A., Arant, C.-A., Wells, P., Stevens, K., Hagen, S., and Harring, H. (2008). A descriptive examination of the most frequently used activity settings in 25 community parks using direct observation. *J. Phys. Act. Health* 5, 183–195. doi: 10.1123/jpah.5.s1.s183

Rigolon, A. (2016). A complex landscape of inequity in access to urban parks: a literature review. *Landsc. Urban Plan.* 153, 160–169. doi: 10.1016/j.landurbplan.2016.05.017

Rishbeth, C. (2001). Ethnic minority groups and the design of public open space : an inclusive landscape? *Landsc. Res.* 26, 351–366. doi: 10.1080/01426390120090148

Sackett, G. P. (1980). "Lag sequential analysis as a data reduction technique in social interaction research," in *Exceptional Infant. Psychosocial Risks in Infant-Environment Transactions*, eds D. B. Sawin, R. C. Hawkins, L. O. Walker, and J. H. Penticuff (New York, NY: Brunner/Mazel), 300–340.

Sánchez-Algarra, P., and Anguera, M. T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: impact of recording and coding among predominating perspectives. *Qual. Quant.* 47, 1237–1257. doi: 10.1007/s11135-012-9764-6

Santoyo, C., Jonsson, G. K., Anguera, M. T., and López-López, J. A. (2017). Observational analysis of the organization of on-task behavior in the classroom using complementary data analysis. *An. Psicol.* 33, 497–514. doi: 10.6018/analesps.33.3.271061

Saunders, P. (2001). "Urban ecology," in *Handbook of Urban Studies*, ed R. Paddison (London: Sage), 36–51.

Scraton, S., and Watson, B. (1998). Gendered cities: women and public leisure space in the 'postmodern city'. *Leis. Stud.* 17, 123–137.

Shippee, N. D. (2012). Victimization, fear of crime, and perceived risk: testing a vulnerability model of personal control. *Sociol. Perspect.* 55, 117–140. doi: 10.1525/sop.2012.55.1.117

Sister, C., Wolch, J., and Wilson, J. (2010). Got green? addressing environmental justice in park provision. *GeoJournal* 75, 229–248. doi: 10.1007/s10708-009-9303-8

Skogan, W. G., and Maxfield, M. G. (1981). *Coping with Crime*. Beverly Hills, CA: Sage Publications.

Subirats, J. (2006). Apunts sobre espai públic, convivència i ciutadania [Notes on public space, coexistence and citizenship]. *Nous Horitzons* 183, 8–14.

Tarragó, R., Iglesias, X., Lapresa, D., and Anguera, M. T. (2016). Complementariedad entre las relaciones diacrónicas de los T-Patterns y los patrones de conducta en acciones de esgrima de espada masculina de élite [Complementarity between diachronic T-pattern relationships and behavioral patterns in elite male fencing]. *Cuad. Psicol. del Deport.* 16, 113–128.

Taylor, R. B., and Hale, M. (1986). Testing alternative models of fear of crime. *J. Crim. Law Criminol.* 77, 151–189. doi: 10.2307/11 43593

Ulrich, R. S. (1984). View through a window may influence recovery from surgery. *Science* 224, 420–421. doi: 10.1126/science.6143402

Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A., and Zelson, M. (1991). Stress recovery during exposure to natural environments. *Environ. Psychol.* 11, 201–230. doi: 10.1016/S0272-4944(05) 80184-7

Valentine, G., and Mckendrick, J. (1997). Children's outdoor play: exploring parental concerns about children's safety and the changing nature of childhood. *Geoforum* 28, 219–235.

Valera, S., and Guàrdia, J. (2014). Perceived insecurity and fear of crime in a city with low-crime rates. *J. Environ. Psychol.* 38, 195–205. doi: 10.1016/j.jenvp.2014.02.002

Valera, S., Pérez-Tejera, F., Anguera, M. T., and Sicilia, L. (2018). Evaluating the uses and environmental characteristics of 40 public parks and squares in Barcelona by means of systematic observation/Evaluación de los usos y características ambientales de 40 parques y plazas en Barcelona mediante observación sistemátic. *Psyecology* 9, 118–151. doi: 10.1080/21711976.2018.1432525

Vargas, D., and Merino, M. (2012). Public spaces in Mexico as social cohesion promoters: an structural modeling perspective. *Well Being Soc. Policy* 9, 157–177.

Viegi, G., Simoni, M., Scognamiglio, A., Baldacci, S., Pistelli, F., Carrozzi, L., et al. (2004). Indoor air pollution and airway disease. *Int. J. Tuberc. Lung Dis.* 8, 1401–1415.

Weir, L. A., Etelson, D., and Brand, D. A. (2006). Parents' perceptions of neighborhood safety and children's physical activity. *Prev. Med.* 43, 212–217. doi: 10.1016/j.ypmed.2006.03.024

Wolch, J. R., Byrne, J., and Newell, J. P. (2014). Urban green space, public health, and environmental justice: the challenge of making cities 'just green enough'. *Landsc. Urban Plan.* 125, 234–244. doi: 10.1016/j.landurbplan.2014.01.017

Detection of Ludic Patterns in Two Triadic Motor Games and Differences in Decision Complexity

Miguel Pic Aguilar[1]*, Vicente Navarro-Adelantado[1] and Gudberg K. Jonsson[2]

[1] Consejeria Educacion Canarias, University of La Laguna, San Cristóbal de La Laguna, Spain, [2] Human Behavior Laboratory, University of Iceland, Reykjavik, Iceland

The triad is a particular structure in which an ambivalent social relationship takes place. This work is focused on the search of behavioral regularities in the practice of motor games in triad, which is a little known field. For the detection of behavioral patterns not visible to the naked eye, we use Theme. A chasing games model was followed, with rules, and in two different structures (A↔B↔C↔A and A → B → C → A) on four class groups (two for each structure), for a total of 84, 12, and 13 year old secondary school students, 37 girls (44%) and 47 boys (56%). The aim was to examine if the players' behavior, in relation to the triad structure, matches with any ludic behavior patterns. An observational methodology was applied, with a nomothetic, punctual and multidimensional design. The intra and inter-evaluative correlation coefficients and the generalizability theory ensured the quality of the data. A mixed behavioral role system was used (four criteria and 15 categories), and the pattern detection software Theme was applied to detect temporal regularities in the order of event occurrences. The results show that time location of motor responses in triad games was not random. In the "maze" game we detected more complex ludic patterns than the "three fields" game, which might be explained by means of structural determinants such as circulation. This research points out the decisional complexity in motor games, and it confirms the differences among triads from the point of view of motor communication.

Keywords: Theme, triad, motor game, structure, T-patterns

INTRODUCTION

Motor games with rules enclose players behavior regularity due to the expectations of the roles. The regularities of these behaviors also depend on the communication structure to which the players are subjected. Play behaviors are an orderly way of communicating in games with rules. The relationships between the roles of the games and its time are two key aspects of the analysis of the relationships showed by the players, in this way triadic relationships are defined. Triad motor games are composed of three decisional units (three individuals or three groups of individuals), in a context of motor communication (Parlebas, 1981), with players' autonomy to act according to strategic interests (Pic and Navarro, 2017).

To speak about regularity of behavior in games is to talk about the outcome of a logic of each motor situation (Parlebas, 1981, 1988, 2005a,b,c). This logic has been developed by Parlebas' motor praxiology, and studied in sports games. However, it is necessary to reach a deeper level of analysis of issues that are hidden from the chaining of roles in different games. Knowing the time regularity

of the motor sequences used by the players provides more information about the background of the transcendent playing relations. Linking the role to the time requirement implies ensuring a more precise, consistent, and revealing pattern.

Motor games show varying levels of complexity; one of these complexities is provided by the triad game (Parlebas, 2011). The inclusion of a third triadic element incorporates coalliance (Gamson, 1961) into the playful scenario (Navarro and Pic, 2016) increasing the habitual complexity of the dual game. In general, games have a consubstantial random component that comes from their degree of uncertainty but, in the face of this, players struggle to control that uncertainty and to approach their strategy. Triad motor games have in complexity a research challenge, because they delve into cooperation (Pic and Navarro, 2017) even though players may be adversaries.

Complexity in triadic games remains a research challenge. Complexity studies have indicated the importance of including the time scale in research (Balagué et al., 2013). In triad, property activation is relevant and susceptible to becoming constraints (Davids et al., 2008). To analyze complexity, three deepening levels are available: (a) the structure of the game, (b) the role and (c) the observed behavior. The complexity of the triad structures (Simmel, 1950; Caplow, 1956, 1959; Wasserman, 1975; Wasserman and Faust, 2013) refers to the amount of relationships established between the teams in the game, and the decisional range of the roles of a game brings with it a series of possible relationships with other roles. Finally, the "strategic whole" is therefore unsurpassable and, through the role, the behaviors put into practice by the players must be made operative in order to unravel the true complexity.

In dual games, the identity of each party is reflected in the antagonism of their relationships. Nevertheless, in the triadic census (Moody, 1998), based on the relation typology, a great variety of connections is verified, which could activate certain properties. Two properties such as reciprocity and circulation affect structures 1 (A↔B↔C↔A) and structure 2 (A → B → C → A) in a different way.

Reciprocity is understood as the two-way communication connection between two teams (A↔B), while circulation is the sequence of connections from one source to the same starting point, completing one cycle. Structure 2 does not have reciprocity, but with one-way circulation, however, structure 1 has 3 reciprocities and two-way circulation, that is, it has two origins. Their reciprocity is affected in the confrontation between the three teams, because each duel would annul itself reciprocally. In relation to the circulation property in structure 2, in simultaneous capture games, if the players of team A take prisoners of the players of team B, it means that most of the players of team C will remain as free players because there are scarce possibilities for team B players to get some catch from team C. That is, team A, following the rules of the game and fulfilling their pretensions, hinder their chances of winning the game. These situations, scarcely studied in the motor game, were already analyzed in depth with a social approach by Caplow (1959). It should be noted that we understand complexity as the set of elements, relationships and emergent properties based on a strategic sense. Therefore, in structure 1, there would be

greater complexity and average probabilities for the paradox to appear, while in structure 2 the complexity would be less, but the structural paradox would appear due to the relational disposition of its elements.

The internal logic of the motor game leads to focus attention on the role as it is a route of play when players act. Role is a structural indicator that helps to operationalize the motor game, as well as to order the strategic procedures that the players put into practice to reach their objectives. In an apparently uncontrolled context in a game, each driving situation is unique and unrepeatable. Role represents the path, the structure of communication support, and T-patterns (detected with Theme) emerge from the observed behaviors. In this sense, the T-patterns found are not rules of decision for the game, but seemingly random events which are timely and behaviorally organized, though. For this reason, an ecological methodology (Anguera and Hernández-Mendo, 2014) is required to advance in the study of these structures based on the motor interaction of three teams, as well as the inclusion of the time parameter (Magnusson, 2000). To find out T-patterns, the observation of games offers methodological advantages with which to reduce the complexity of the triadic game.

In the observational methodology (Lapresa et al., 2013a; Anguera and Hernández-Mendo, 2014, 2016) the use of analysis techniques that help to evidence the construction of behavioral structures with time regularity has been increasing. As it is well known, THEME is a software (Magnusson, 1996, 2000) that detects T-patterns (Borrie et al., 2002; Jonsson et al., 2006; Casarrubea et al., 2015) by combining ordered events which occur at relatively invariable time distances.

Motor games have not been studied using the T-pattern algorithm before, despite their communicational richness, but sports have. There are differences of a time sense between the first and second part in soccer matches in high level competitions (Cavalera et al., 2015) detected by using an observational methodology and THEME (Magnusson, 1996, 2000). T-patterns have been identified in the motor interactions of F.C. Barcelona (Camerino et al., 2012) and finalized strikes in goal in futsal (Lapresa et al., 2015). In basketball, the effectiveness of offensive play (Fernandez et al., 2009; Lapresa et al., 2013a) and foot position were studied, among other criteria, to try to optimize pitch (Garzón et al., 2011), and it was also studied in combat sports (Camerino et al., 2014).

Justification of the hypothesis: the decisions of a triadic motor game players show the complexity of game structures. Strategic regularity represents a degree of strategic organization which players put into practice when playing. The detection of T-patterns is evidence of the logic of strategic sense, ordered in roles. The decisional chance is greater the less able the players are to carry out actions that allow obtaining advantages. When confronting two triadic motor games with different distribution of directed graphs (communication flows), two communication conditions are tested when developing the game strategy; the game "the maze" is more complex due to the number of reciprocities, while "the three fields" does not have any reciprocity. In addition, the one-way communication of the game "the three fields," directly activates a structural

paradox with consequences on the strategy of teammates and/or rivals.

In this line of research, aimed at the motor game, the objective was to look for temporary regularities in two triad games, under two different communication structures, through game roles and their observable behaviors in practice. For this study, the following generic and specific objectives were considered, respectively.

- Identifying systemic properties in two triadic games and analyzing their influence in practice.
- Discriminating the scope of triadic behavior according to the roles of the game.

Hypothesis:

- The increase in decisional complexity and strategic regularity (T-patterns) is greater in favor of the triadic game with a greater number of reciprocities in its communication.

METHODS

Design

An observational methodology design was selected for the study. It is a relatively recent methodological approach (Anguera and Hernández-Mendo, 2014, 2016) with application to sports and physical education (Fernández et al., 2012; Hernández-Mendo and Planchuelo, 2013). Specifically, a N/P/M design (Blanco-Villaseñor et al., 2003; Anguera et al., 2011) is applied: (a) Nomothetic (N) because the motor behaviors of different players were recorded; (b) Punctual (P), because the registered games were raised in a precise moment; and finally, (c) Multidimensional (M), since different dimensions (observational criteria) constituting the observation tool were taken into account.

Participants

The number of players was 84, consisting of 37 (44%) girls and 47 (56%) boys between 12 and 13 years old (M = 12.5; DT = 1) from two secondary schools in the Canary Islands (Spain). The two institutes were located in different cities and islands. The cities were middle-class urban places. Both the groups and the institutes were selected according to accessibility and intentionality (Anguera et al., 1995). The students played two motor games, distributed in 4 class groups; in each center there were two groups (group A, group B) for each game (game 1: Maze and game 2: Three fields). The first game (maze, modified), in both secondary schools, was played by 21 Players and distributed into 7 participants per team (group 1). The second game (Three fields) was practiced with an identical distribution of players in both secondary schools. This study was carried out in accordance with the recommendations of *ethics committee for research and animal welfare* of the University of La Laguna (Spain) with written informed consent from all parents or legals tutors of all participants (Declaration of Helsinki).

Materials

The two triadic games analyzed were "the maze" (modified) and "the three fields" (modification), which are both chase games.

In "the maze" (A↔B↔C↔A), three teams are formed, with the same strengths, and all players try to capture each other simultaneously, under action conditions that regulate only one part of the body for contact. The captured player assumes the role "prisoner," remaining crouched in the place where he/she was captured; "prisoners" can be released if they are saved by a free player, under the role of "savior" (Navarro, 1995). The team that first turns all adversaries into "prisoners" wins. It is an ambivalent and stable motor communication network (Parlebas, 1988). In the game "the three fields" (modified) (A → B → C → A), the chase cycle between the teams is regulated.

To analyze these games, the motor interactions of the roles were taken as a reference (**Table 1**). In addition, four indicators were used: roles, group interaction (intragroup, intergroup), communication (emission, reception), and valence (positive and negative; Heider, 1946). All interactions are computed taking into account the vertex or node representing each team (A, B, C) and their corresponding emissions (positive or negative) and receptions (positive or negative), giving rise to three values. In **Table 1** and for game 2, the rating (3,3,3) in the "catcher" role as negative emission means that 3 is the value of the negative emission flux of A on the roles of team's B and C, specifically with 1 reception for each catcher, dodger, and savior. And all this happens as intergroup interaction.

"The maze" (modified) has a total of 42 motor interactions (6 positive and 36 negative: 1 to 6, in disequilibrium in favor of rivalry). That is, for each cooperative interaction, six antagonists were found. In contrast, when quantifying the motor interaction in "the three fields" game (modified) relational values decrease to 24 motor interactions (6 positive and 18 negative: 1 to 3, in imbalance in favor of rivalry over solidarity). Therefore, a structural difference between these two games comes from doubling antagonism over cooperation. Consequently, these are two games structures with marked differences.

Observational Record System

For the construction of the registration tool, a previous exploratory study (Pic and Navarro, 2014) was used, detecting which strategic options were most demanded by players. From these options, nesting's on the praxiological core "role" were identified. The valence variable was not included in the registration tool because the behavior implies a positive valence (cooperative behavior), a negative valence (rivalry behavior) or an ambivalent behavior (valence positive and negative behavior).

An "*ad-hoc*" tool was used (**Table 2**) to record the behaviors carried out by boys and girls in triad motor games. The categories are exhaustive and mutually exclusive, merging the multidimensionality of the field format and the referent of the category system.

The catcher role (C) is identified because it bears the initiative of the game, and it is observable through effective captures (CA) or pursuit actions (PA), which brings together at least two players involved with negative emissions. Sometimes, (C) tries to defend players who are already prisoners and it is then when (DEF) takes place, while passive was the player who was not involved in the game (P). When it was observed that a player

TABLE 1 | Motor interactions of "the maze" game (modified) and "the three fields" (modified), following the indicators "role," "game group," "valence," and "communication."

		Intragroup				Intergroup			
		Emissions		Receptions		Emissions		Receptions	
		+	−	+	−	+	−	+	−
Game	Roles								
The Maze	Capture	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	6,6,6	0,0,0	2,2,2
	Dodger	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	2,2,2
	Prisioner	0,0,0	0,0,0	1,1,1	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0
	Savior	1,1,1	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	2,2,2
	Total	1,1,1	0,0,0	1,1,1	0,0,0	0,0,0	6,6,6	0,0,0	6,6,6

6 intragroup interactions (6 positive, 0 negative) In total (A, B, C = 3 positive emissions and 3 positive receptions)

36 intergroup interactions (0 positive and 36 negative) In total (A, B, C = 18 negative emissions and 18 negative receptions)

Total: 42 motor interactions (6 positive and 36 negative: 1 to 6 in favor of rivalry over solidarity)

		Intragroup				Intergroup			
		Receptions		Emissions		Receptions		Emissions	
		+	−	+	−	+	−	+	−
Game	Roles								
The three fields	Capture	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	3,3,3	0,0,0	1,1,1
	Dodger	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	1,1,1
	Prisioner	0,0,0	0,0,0	1,1,1	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0
	Savior	1,1,1	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	0,0,0	1,1,1
	Total	1,1,1	0,0,0	1,1,1	0,0,0	0,0,0	3,3,3	0,0,0	3,3,3

6 intragroup interactions (6 positive, 0 negative) In total (A, B, C = 3 positive emissions and 3 positive receptions)

18 Intergroup Interactions (0 positive and 18 negative) In total (A, B, C = 9 negative emissions and 9 negative receptions; 0 positive emissions and 0 positive receptions)

Total: 24 motor interactions (6 positive and 18 negative: 1 to 3, in disequilibrium in favor of rivalry over solidarity)

belonging to different teams pursued an opponent player, it was considered (ALZAAC), being an ambivalent behavior. Although the previous role had more initiative, the "dodger" role (E) offered a reciprocal response to the catch action (C). In this sense, the dodger (E) will act by dodging on the opponent (EA) but he/she can also flee (HA) or move to free places (DLL). When it is observed that a fellow player intends to interfere in a chase between a catcher and a fleeing player, it was considered (AC). When the action is developed cooperatively between players (ambivalence behaviors) of different teams, it was considered (ALZAE). If a player was caught and did not admit it, this conduct was considered (NR). The prisoner role (P) occurs when a player has been previously captured. If the player is simply standing without facilitating his/her release we say that he/she is not in attention (A), but other times the player facilitates the saving action (CE). The fourth role, liberator (L), tries to save fellow prisoner players (TUFC) but sometimes the release was verified between players of different teams (TUFA).

Procedure

The images, video recordings, were taken in the educational centers of each group. In order to have a clear observability of each record (Anguera, 2003) and to have at least two recordings of each sequence, inclusion criteria were used. The consent of the study participants and parents/guardians was obtained. The collection of images was made through long distance recordings, making impossible the recognition of the faces of the players in those images. Although these games are usually longer, the first 3 min of each game carried out by each class group were selected. During each practical experience, the spatial conditions

Criteria	Categorie	Description
Catcher (C)	CA	Catches an opponent
Catcher (C)	PA	Chases an opponent
Catcher (C)	DEF	Defends a prisoner
Catcher (C)	P	Passivity
Catcher (C)	ALZAAC	Alliance with adversary
Dodger (E)	EA	Dodges an opponent
Dodger (E)	HA	Runs away from an opponent
Dodger (E)	AC	Helps a fellow escape
Dodger (E)	DLL	Moves to free places
Dodger (E)	NR	Does not recognize being caught
Dodger (E)	ALZAE	Alliance between dodging adversaries
Prisioner (P)	A	In atention
Prisioner (P)	CE	Changes space to make it easier to release
Liberator (L)	TUFC	Touches a prisoner (fellow prisoners)
Liberator (L)	TUFA	Touches a prisoner (adversary prisoners)

were identical, with a play space of 20 × 20 m and a similar surface.

The measurements related to the itineraries of the roles and properties that attend the two play structures served as an a priori platform, thus systematizing the recording of motor behaviors through a system composed of four criteria and 15 categories. The follow-up of an observational methodology (Anguera and Hernández-Mendo, 2014, 2016) allowed us to delve into the scarce knowledge available regarding the behavior of players in the motor triad as a play scenario.

The final records were made by two experts in motor games. Previously, and to construct the recording tool, the observers analyzed images taken from different motor games, but developed under the same structures of the present study, with the purpose of agreeing on the degrees of freedom of the categories and suitable nuclear criteria. Once the tool was available, and using the software LINCE (Gabín et al., 2012), the images of the explained games were recorded. When the acceptable levels of reliability in the social sciences had been calculated, and checked, the definitive images were recorded using the same software.

The data quality control was carried out, in order to calculate the inter-observer and intra-observer reliability and validity. Pearson and Spearman correlation coefficients were used, as well as the theory of generalizability (Cronbach et al., 1972). Generalizability analysis was used to estimate accuracy, validity, reliability (Blanco-Villaseñor et al., 2014). In addition, the role sistem acts as a theoretical construct for the description of observed behaviors (Parlebas, 1981).The lowest values reached were 0.954 inter-operator per 0.964 intra-observer, Spearman coefficients in both cases. The JGRC/M model was used to calculate the variance (0%) attributable to the measurements made by each observer at two different times. Also, the JGRC/O model was used to calculate the inter-observer variance, repeating the procedure in the comparison made by both evaluators, obtaining values of 1%. The tools used to complete the

data quality were the Generalizability Study GT program, v.2.0E (Ysewijn, 1996), the SAS statistical package (v.9.1.3) and the SPSS program (v.20).

Data Analysis

After transcribing video recording data was analyzed using Theme 6.0. Searching for temporal patterns default search parameters were used, except level of significance was set at 0.005 and minimum occurrences at 2.

RESULTS

The two most complex T-patterns detected for groups 1 and 2 are displayed in **Figure 1**. The event time plot (top part of figure) illustrates the temporal distribution of transcribed events (the horizontal line representing the observation period and the vertical line occurrence time event types registered). For group 1, 117 different event types were registered, mounting to 278 data points (data rate 0.08). For group 2, 105 different event types were registered, mounting to 305 data points (data rate 0.10). The pattern statistics differed between the two groups. For group 1, 402 different patterns were detected, occurring 833 times. For group 2, 1,466 different patterns were detected, occurring 3,048 times.

The group 1 most complex pattern detected, displayed in **Figure 1** (left side), consists of 15 different events, occurring with significantly similar time interval between themselves, on two occasions during the observation period. The pattern duration covers 61% of the observation period. The group 2 most complex pattern detected, displayed in **Figure 1** (right side), consists of 18 different events, occurring with significantly similar time interval between themselves, on two occasions during the observation period. The pattern duration covers 89% of the observation period.

In the first game, specifically with group 1 composed of three teams (TR, TG, TB), a pattern tree graph with a greater verticality than in group 2 could be observed. The most dependent relations of group 1 were intragroup and not intergroup. The behavior of the TR team player (b5tr, ca) links its appearance in the game to the captive player (g1tr, ca). The TG player (g1tg, a) is imprisoned a prisoner and timely attached to the dodge action (b5tg, ea). The same player (b5tg, ea) is captured and related to the player's catch action (b7tg, ca).

The player of the TG team (g1tg, tufc) performs release actions related to fellows that are captured (b7tg, a). Finally, from the pattern tree graph, we find a group of three behaviors that link the release of partners (b6tr, tufc) to the existence of fellow prisoners (b5tr, a) and (g1tr, a), similar to the previous case.

In the second group of the same game, the teams (Y, NC, and B) showed through the pattern tree graph more intergroup behaviors with recurrent time dependence than intragroup ones. A prisoner of team Y (b6y, a) with a catcher of team B (b5b, ca). We observed a group of three behaviors of the NC team of the prisoner role in attention (b5nc, a) with displacements to free places (b7nc, dll) and fleeing (b5nc, ha). Also, (b5y, ea) and (b5b, a), like (b6y, tufa) with (b7y, ea) and (b5b, ca).
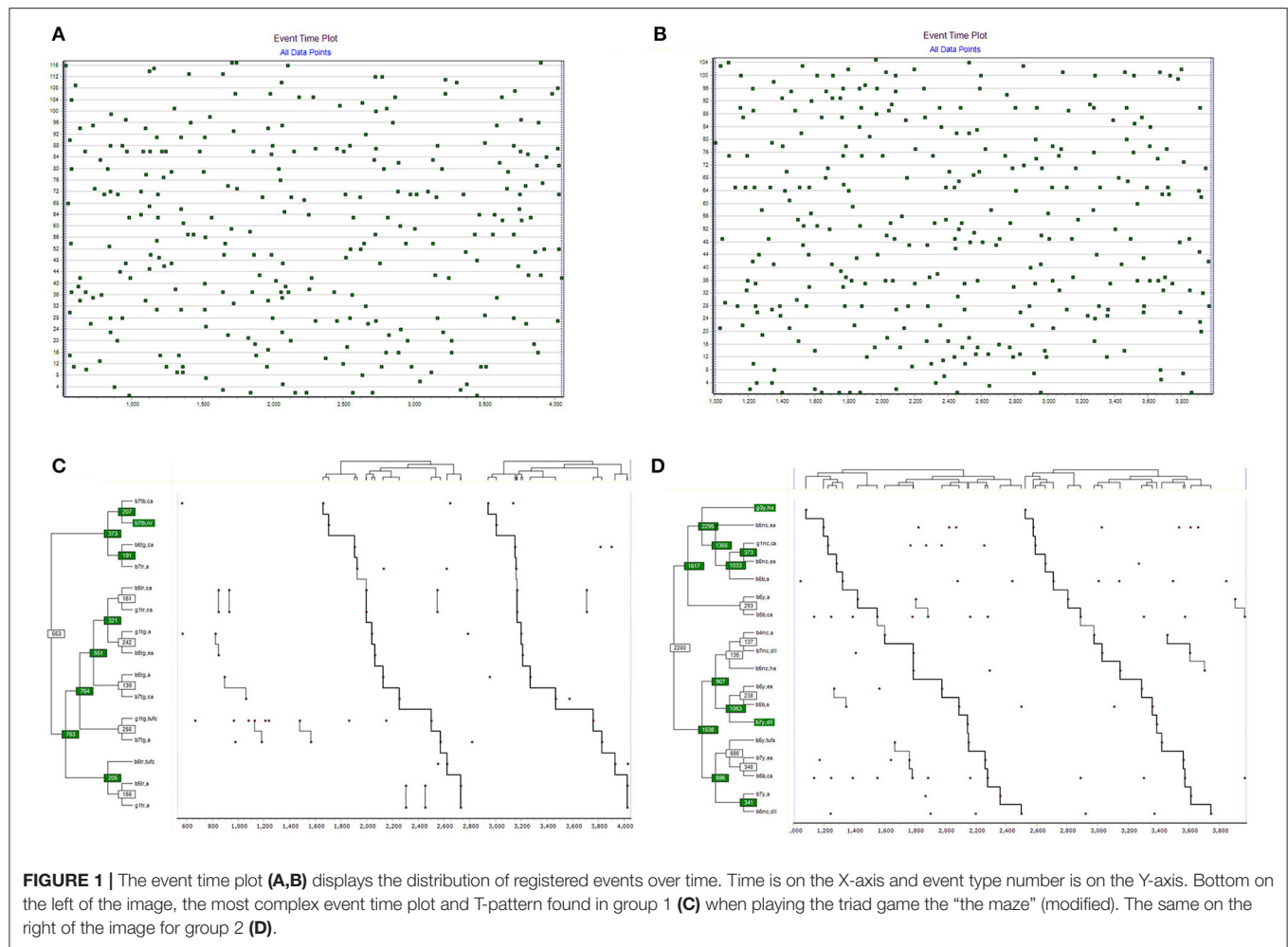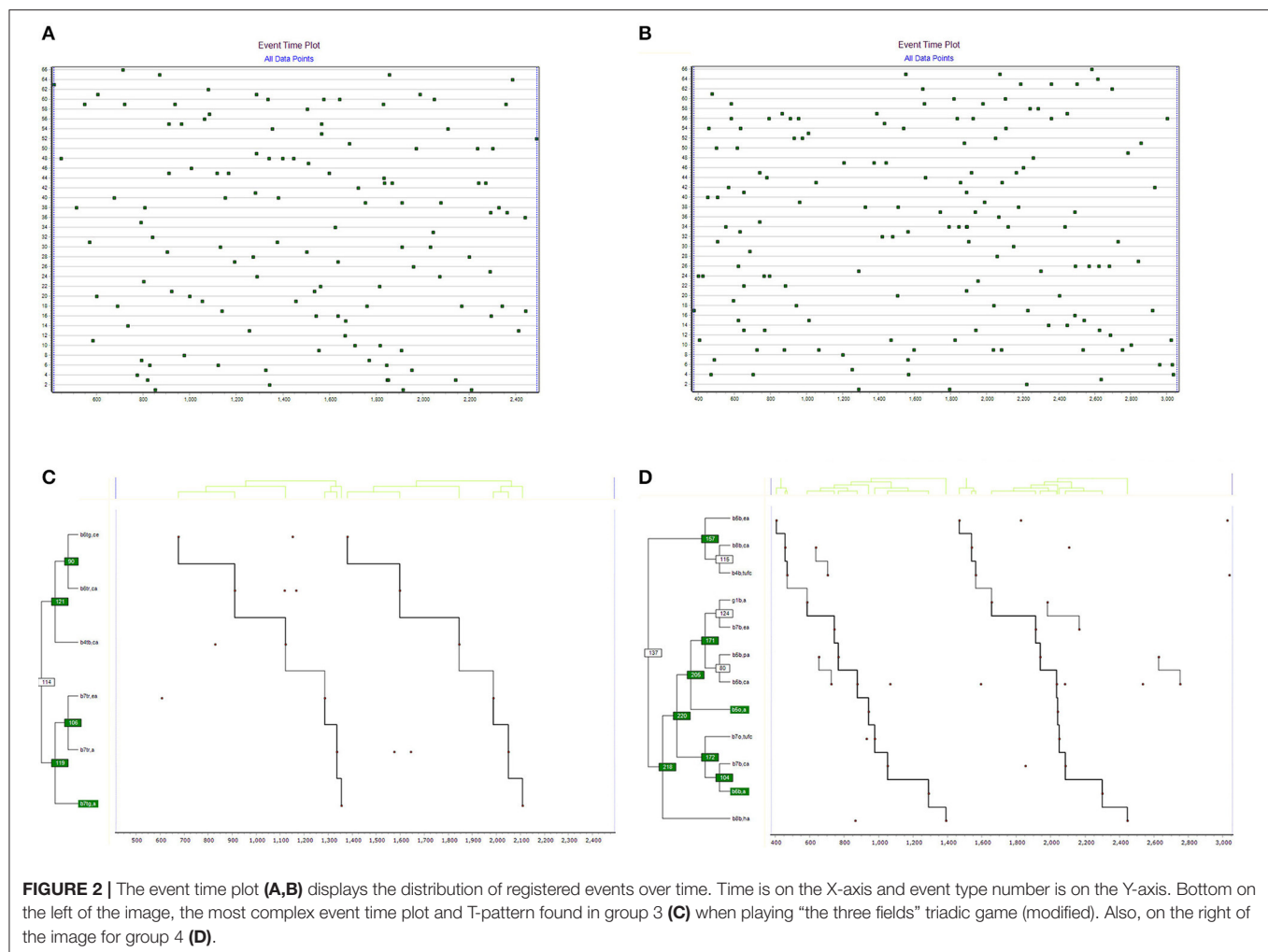
FIGURE 1 | The event time plot **(A,B)** displays the distribution of registered events over time. Time is on the X-axis and event type number is on the Y-axis. Bottom on the left of the image, the most complex event time plot and T-pattern found in group 1 **(C)** when playing the triad game the "the maze" (modified). The same on the right of the image for group 2 **(D)**.

The two most complex T-patterns detected for the groups 3 and 4 are displayed in **Figure 2**. The event time plot (top part of figure) illustrates the temporal distribution of transcribed events (the horizontal line representing the observation period and the vertical line occurrence time event types registered). For group 3, 25 different event types were registered, mounting to 66 data points (data rate 0.06). For group 4, different event types were registered, mounting to 66 data points (data rate 0.06). The pattern statistics differed between the two groups. For Group 3, 36 different patterns were detected, occurring 73 times. For group 4, 97 different patterns were detected, occurring 97 times.

The group 3 most complex pattern detected, displayed in **Figure 2** (left side), consists of 6 different events, occurring with significantly similar time interval between themselves, on two occasions during the observation period. The pattern duration covers 68% of the observation period. The group 4 most complex pattern detected, displayed in **Figure 2** (right side), consists of 11 different events, occurring with significantly similar time interval between themselves, on two occasions during the observation period. The pattern duration covers 73% of the observation period.

Group 4 shows a greater time dependence on game behaviors than group 3. In this group, the captive player of the TG team (b6tg, ce) caused captures (b6tr, ca) in the player belonging to the TR team. The two previous behaviors relate to the captures of a player of the TB team (b4tb, ca). Also, the dodging of the player belonging to the team TR (b7tr, ea) caused that same player to fall prisoner (b7tr, a). The two responses above were linked to the TG team reaction (b7tg, a).

The most obvious time recurrences were established with team B players. First, catches of team B (b8b, ca) with releases that another player from the same team performed (b4b, tufc). Also between the player (g1b, a) of the same team and (b7b, ea). Another linkage found corresponded to behaviors within team B and player (b5b, pa) and (b5b, ca). Other relationships that were part of the behavioral cluster, although less remarkable, would be those shown among the last four behaviors mentioned, to cause the identification of the behavior of team O (b5o, a) as a captured player. When team O freed teammates (b7o, tufc), actions were activated on team B to capture (b7b, ca), while prisoners on hold (b6b, a). Finally, the flight behavior of team B (b8b, ha) was activated.

FIGURE 2 | The event time plot **(A,B)** displays the distribution of registered events over time. Time is on the X-axis and event type number is on the Y-axis. Bottom on the left of the image, the most complex event time plot and T-pattern found in group 3 **(C)** when playing "the three fields" triadic game (modified). Also, on the right of the image for group 4 **(D)**.

DISCUSSION

A powerful data pattern detection technique was used, Theme 6.0 (Magnusson, 1996, 2000) to address the comparison of two triad structures of motor games to confirm or reject the existence of T-patterns in practice. The detection of T-patterns provides an accurate fusion of motor responses during the game to generate a time cluster with which to solve the triadic complexity; this fusion of motor responses emerged under the same system of roles for both games and it only limited the actions of the action structure of the triad.

The analysis of play structures (Pic and Navarro, 2017) in motor games, taking roles into account (**Table 1**), anticipated some key elements: (a) smaller relational limitations of structure 1 on structure 2; (b) increase of the antagonistic density of game 1 on game 2; (c) influence of the activation of structural properties in the two triad games and their consequences. This anticipated forecast theoretically requires to be contrasted with the results obtained. Therefore, we next assess the complexity of the T-patterns found in relation to the properties noticed in the structures, and their transformation from game roles to observed behaviors. In this way, advancing in the conciliation

between play practice and a previous communication analysis tries to overcome the conjecture. In general, the analysis is based on ecological conditions (Araújo et al., 2006), facilitated by a methodology that allows the registration in natural environments (Anguera and Hernández-Mendo, 2016), and by the inclusion of a time dimension (Magnusson, 1996, 2000), in full harmony with the reality of the triadic events of motor practice (Pic and Navarro, 2017).

Complexity of T-Patterns and Relational Constraints on Triad Structures

One aspect of T-patterns is the complexity of patterns detected. The more structured behavioral phrases of the cluster detected by Theme (Jonsson et al., 2006) confirm the high relational complexity of the players when playing. Specifically, between the two groups practicing the same game the maze (modified) made a total of 1,868 T-patterns composed of 3,881 different behaviors, whereas in "the three fields" (modified) game they only reached 133 T-patterns composed of 170 ludic behaviors. Although it is true that the structural formula of "all against all" facilitated the relational exchange between participants, the motor response is

put open to debate before the context influence (Araújo et al., 2014), subjecting players to comply with what is allowed by the rule within a given structure and under the organization of the role. In view of this structural condition, the results confirmed the existence of T-patterns in both triadic games. However, the time recurrence shown by the pattern tree graphs in the maze game (modified) reveals differences with respect to "the three fields" game (modified), joining the patterns found in other specific contexts (Fernandez et al., 2009; Lapresa et al., 2013a,b; Cavalera et al., 2015).

The origin of these differences refers to the nature of motor interaction (Parlebas, 1981) and to the intensity of antagonism (Heider, 1946; **Table 1**). In the maze (modified) we found that the space for collaboration was scarce (one collaborative behavior by six antagonists) compared to the three fields (modified), with one collaborative behavior by three antagonists. This extremely antagonistic scenario in of the maze game, requires many prisoner releases for the ludic system to be able to adapt (Passos et al., 2016). The facilitation of the capture as a priority objective to win needs more releases than in "the three fields" game (modification). In this sense, it is a less rigid formula and more adaptable to errors and successes of the players. The number of T-patterns reflects this high combinatorial complexity.

The search for T-patterns shared by both games or groups was unsuccessful and calls for an interpretation. Triadic structures are complex relational archetypes, subject to the emergence of written properties in the structure, but knowing these play systems does not guarantee their prediction. The differences between the four groups when playing two game structures showed different T-patterns. Perhaps because of the decision-making ability of the players, or maybe because of the restrictions of Theme to identify T-patterns with the subjects of the study labeled with event occurrences, or due to the vertiginously driven demands for situational adaptations.

Players may in many cases select automatic responses when facing the difficulty of having short time frames to react. On the other hand, it should not be forgotten that players do not have a decisive recipe (Araújo et al., 2014). Therefore, and in this sense, the decision is subject to great variation and, consequently, the groups and structures of the game do not explain by themselves the similarities in the obtained in the pattern tree graphs.

Triadic T-Patterns and Emerging Properties (Circulation and Reciprocity)

What keys have been activated in the games studied following the detection of T-Patterns? Reciprocity, when it is antagonistic (mirror-like, e.g., catcher-dodger), is determinant in the triad analysis because it hinders circulation fulfillment. That is to say, it is inversely proportional: to more reciprocity between the sum of existing duels, less circulation in one direction. This property justifies the structural paradox when the relation is fulfilled in a single sense, as happens in "the three fields" (modified). In this sense, it is a property of the circulation network, transformed into a triad constrains (Davids et al., 2008). It is thus that

the lack of strategic organization detected by Theme in "the three fields" (modified) is not accidental but causal, and it was more disorganized. In "the maze" there was a greater decisional alternative (Araújo et al., 2006) to address the problems that arose from the game.

If we focus on the salvator role, on it rests essentially the continuity of the game, in a systemic sense, but it affected each group in a different way. In group 1, the TUFC record was identified by THEME as T-pattern. However, group 2 practicing the same game (the maze) was found to be TUFA (action to free opponents of the rival team, developed by rescuers). According to this, the need for releases in the maze mentioned in the previous section seems to be reinforced due to the communication structure (**Table 1**), but also to the specificity of each group, since it was only in the second of them that Theme identified it.

Specifically, in the last relational framework identified by Theme, in the second group of "the maze" game, which indicates that the team would have been in a difficult situation, also supported because the rival team could make effective captures (b5b, ca), is the release of opponents made by the player (b6y, tufa) linked to the player of the same team when performing dodges (b7y, ea). It may be this criticality what leads the teams to partnership with rivals. Based on the above, its positive or negative good value (Heider, 1946) is vital to understand the strategic specificity of each group and the systemic need for liberation for both groups during play.

We have already alluded to a smaller elaboration of the T-patterns in the three fields in front of the game the maze, backed by the number of T-patterns but also by a greater strategic and temporary structuring of the motor action. The collaborations or release actions between players in the first game showed different behaviors regarding the previous game. Release actions among fellow players (TUFC) were only identified in group 4. That is, the lower number of T-patterns identified in the three fields affected the savior role. In group 3, there was no collaboration to make releases between fellow players or opponents, which could indicate a state of normative incomprehension, culturally understandable in the players for a lack of triadic experiences that group 4 put into practice. The property of circulation and its paradoxical effects on the motor decision, could explain the high strategic disorganization identified by THEME in the third triad group. Again, structure and play roles described the pattern of the motor decision, showing the situational demands that were brought by the groups and found by the relevant detection of temporary recurrences in view of previous studies in sports (Jonsson et al., 2006).

In conclusion, T-patterns have solved the underlying complexity of the two different game structures and their groups, beyond the systemic solution, showing how players are temporarily confronted in natural contexts of practice.

CONCLUSIONS

This study has demonstrated the advantages of using a technique pattern detection to address the internal complexity of the motor triad. The inclusion of a time dimension has meant an advance

for the interpretation and analysis of data from ecological contexts, confirming the properties of the ludic structure.

The decisional complexity of the T-Patterns was different in "the maze" and "the three fields" triadic games. It was also reinforced by the number and composition of T-patterns. The relational strategies identified in the four groups were different, since no similarities were found; which confirms the high complexity of each game developed by each group. This high complexity shows a specific variability for each triadic motor game.

LIMITATIONS

Among the limitations that accompanied the study, the inclusion of the spatial criterion as a facet of it would be worth mentioning. Similar studies could replicate this research, with different populations and age groups, to have a thorough knowledge about the triadic effects. Performing analysis aimed at detecting the inhibition and activation of particular play behaviors would add an important explanatory value. The theory of motor play needs to put on hold the circle of appreciation and move on to research,

in order to know what these and other motor games hide, as exceptional formulas of social interaction.

AUTHOR CONTRIBUTIONS

MA and VN-A have contributed to the theorical and methodological development of the article while GJ has contributed with the data analysis. Results and discussion have been prepared by all authors.

FUNDING

REFERENCES

Anguera, M. T. (2003). "Observational methods (general)," in *Encyclopedia of Psychological Assessment, Vol. 2,* ed R. Fernández-Ballesteros (London: Sage), 632–637.

Anguera, M. T., and Hernández-Mendo, A. (2014). Metodología observacional y psicología del deporte. Estado de la cuestión. *Rev. Psicol. Dep.* 23, 103–109.

Anguera, M. T., and Hernández-Mendo, A. (2016). Advances in mixed methods observational studies in sports sciences. *Cuad. Psic. Dep.* 16, 17–30.

Anguera, M. T., Arnau, J., Ato, M., Martínez, R., Pascual, J., and Vallejo, G. (1995). *Métodos de Investigación en Psicología*. Madrid: Síntesis.

Anguera, M. T., Blanco, A., Hernández-Mendo, A., and Losada, J. L. (2011). Observational designs: their suitability and application in sports psychology. *Cuad. Psic. Dep.* 11, 63–76.

Araújo, D., Davids, K., and Hristovski, R. (2006). The ecological dynamics of decision making in sport. *Psychol. Sport. Exerc.* 7, 653–676. doi: 10.1016/j.psychsport.2006.07.002

Araújo, D., Silva, P., and Ramos, J. P. (2014). Affordance-based decisions guide team synergies during match performance. *Res. Phys. Edu. Sport. Health* 3, 19–26.

Balagué N., Torrents, C., Hristovski, R., Davids, K., and Araújo, D. (2013). Overview of complex systems in sport. *J. Syst. Sci. Complex.* 26, 4–13. doi: 10.1007/s11424-013-2285-0

Blanco-Villaseñor, A., Castellano, J., Hernández-Mendo, A., Sánchez-López, C. R., and Usabiaga, O. (2014). Aplicación de la TG en el deporte para el estudio de la fiabilidad, validez y estimación de la muestra [Application of the generalizability theory in sport to study the validity, reliability and estimation of samples]. *Rev. Psicol. Dep.* 23, 131–137.

Blanco-Villaseñor, A., Losada, J. L., and Anguera, M. T. (2003). Data analysis techniques in observational designs applied to the environment-behaviour relation. *Medio Ambient. Comport. Hum.* 4, 111–126.

Borrie, A., Jonsson, G. K., and Magnusson, M. S. (2002). Temporal pattern analysis and its applicability in sport: an explanation and exemplar data. *J. Sport. Sci.* 20, 845–852. doi: 10.1080/026404102320675675

Camerino, O., Chaverri, J., Anguera, M. T., and Jonsson, G. K. (2012). Dynamics of the game in soccer: detection of T-patterns. *Eur. J. Sport. Sci.* 12, 216–224. doi: 10.1080/17461391.2011.566362

Camerino, O., Prieto, I., Lapresa, D., Gutiérrez-Santiago, A., and Hileno, R. (2014). Detección de T-patterns en la observación de deportes de combate. *Rev. Psicol. Dep.* 23, 147–155.

Caplow, T. (1956). A theory of coalitions in the triad. *Am. J. Sociol. Rev.* 21, 489–493. doi: 10.2307/2088718

Caplow, T. (1959). Further development of a theory of coalitions in the triad. *Am. J. Sociol.* 64, 488–493. doi: 10.1086/222544

Casarrubea, M., Jonsson, G. K., Faulisi, F., Sorbera, F., Di Giovanni, G., Benigno, A., et al. (2015). T-pattern analysis for the study of temporal structure of animal and human behavior: a comprehensive review. *J. Neurosci. Methods* 239, 34–46. doi: 10.1016/j.jneumeth.2014.09.024

Cavalera, C., Diana, B., Elia, M., Jonsson, G. K., Zurloni, V., and Anguera, M. T. (2015). T-patterns analysis in soccer games: relationship between time and attack actions. *Cuad. Psic. Dep.* 15, 41–50. doi: 10.4321/S1578-84232015000100004

Cronbach, L. J., Gleser, G. C., Nanda, H., and Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements*. New York, NY: Wiley.

Davids, K., Button, C., and Bennett, S. (2008). *Dynamics of Skill Acquisition: A Constraints-Led Approach*. Champaign, IL: Human Kinetics.

Fernandez, J., Camerino, O., Anguera, M. T., and Jonsson, G. K. (2009). Identifying and analyzing the construction and effectiveness of offensive plays in basketball by using systematic observation. *Behav. Res. Methods* 41, 719–730. doi: 10.3758/BRM.41.3.719

Fernández, M., Sánchez, C. R., Jiménez, F., Navarro, V., and Anguera, M. T. (2012). Sistema de codificación y análisis de la calidad del dato para una intervención inclusiva en Educación Física. *Rev. Psicol. Dep.* 21, 67–73.

Gabín, B., Camerino, O., Anguera, M. T., and Casta-er, M. (2012). Lince: multiplatform sport analysis software. *Proc. Soc. Behav. Sci.* 46, 4692–4694. doi: 10.1016/j.sbspro.2012.06.320

Gamson, W. (1961). An experimental test of theory of coalitions formation. *Am. Sociol. Rev.* 26, 565–573. doi: 10.2307/2090255

Garzón, B., Lapresa, D., Anguera, M. T., and Arana, J. (2011). Análisis observacional del lanzamiento de tiro libre en jugadores de baloncesto base. *Psicothema* 23, 851–857.

Heider, F. (1946). Attitudes and cognitive organization. *J. Psychol.* 21, 107–112. doi: 10.1080/00223980.1946.9917275

Hernández-Mendo, A., and Planchuelo, L. (2013). El incremento del desarrollo moral en las clases de Educación Física. *Rev. Iberoamericana Psicol. Ejercicio Dep.* 9, 393–421.

Jonsson, G. K., Anguera, M. T., Blanco-Villaseñor, A., Losada, J. L., Hernández-Mendo, A., Ardá, T., et al. (2006). Hidden patterns of play interaction in soccer using SOF-CODER. *Behav. Res. Methods.* 38, 372–381. doi: 10.3758/BF03192790

Lapresa, D., Anguera, M. T., Alsasua, R., Arana, J., and Garzón, B. (2013a). Comparative analysis of T-patterns using real time data and simulated data by assignment of conventional durations: the construction of efficacy in children's basketball. *Int. J. Perform. Anal. Sport.* 13, 321–339. doi: 10.1080/24748668.2013.11868651

Lapresa, D., Arana, J., Anguera, M. T., and Garzón, B. (2013b). Comparative analysis of sequentiality using SDIS-GSEQ and THEME: a concrete example in soccer. *J. Sport. Sci.* 31, 1687–1695. doi: 10.1080/02640414.2013.796061

Lapresa, D., Camerino, O., Cabedo, J., Anguera, M. T., Jonsson, G. K., and Arana, J. (2015). Degradation of T-patterns in observational studies: a study on the effectiveness in futsal. *Cuad. Psic. Dep.* 15, 71–82. doi: 10.4321/S1578-84232015000100007

Magnusson, M. S. (1996). Hidden real-time patterns in intra- and inter-individual behavior. *Eur. J. Psych. Assess.* 12, 112–123. doi: 10.1027/1015-5759.12.2.112

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behav. Res. Methods Instrum. Comput.* 32, 93–110. doi: 10.3758/BF03200792

Moody, J. (1998). Matrix methods for calculating the triad census. *Soc. Netw.* 20, 291–299. doi: 10.1016/S0378-8733(98)00006-9

Navarro, V., and Pic, M. (2016). Regulating Co-alliance in a triad motor game. *Ágora PE Sport* 18, 276–288.

Navarro, V. (1995). *Study of Children's Behavior in a Game with Rules. Analysis of Game Structure, Age and Gender.* Doctoral, thesis, University of Las Palmas de Gran Canaria. Available online at: http://hdl.handle.net/10553/2016

Parlebas, P. (1981). *Contribution à un Lexique Commenté de L'action Motrice.* Paris: INSEP.

Parlebas, P. (1988). *Elements of Sociology in Sport.* Málaga: Unisport.

Parlebas, P. (2005a). Mathematisation elementaire de l'action dans les jeux sportifs. *Math. Sci. Hum.* 170, 95–117.

Parlebas, P. (2005b). Modelling in games and sports. *Math. Sci. Hum.* 170, 11–45. doi: 10.4000/msh.2968

Parlebas, P. (2005c). Elementary mathematisation of the action in sport games. *Math. Sci. Hum.* 170, 5–9. doi: 10.4000/msh.2966

Parlebas, P. (2011). Cursed trio or fruitful triad? The case of the "Rock-Paper-Scissors" *Math. Sci. Hum.* 196, 5–25. doi: 10.4000/msh.12107

Passos, P., Araújo, D., and Davids, K. (2016). Competitiveness and the process of co-adaptation in team sport performance. *Front. Psychol.* 7:1562. doi: 10.3389/fpsyg.2016.01562

Pic, M., and Navarro, V. (2014). Triadic games through a system of categories formed by roles. *Rev. Inter. Dep. Col.* 18, 15–23.

Pic, M., and Navarro, V. (2017). Triad communication and specificity of motor games. *Rev. Int. Med. Cienc. Ac.* 17, 523–539. doi: 10.15366/rimcafd2017.67.009

Simmel, G. (1950). "The triad," in *The Sociology of George Simmel,* ed K. Wolff (New York, NY: The Free Press-Mcmillan Publisher Company), 145–162.

Wasserman, S. (1975). "Random directed graph distributions and the triad census in social networks," in *NBER Working Paper,* Vol. 113 (Cambridge, MA), 1–47. Available online at: www.nber.org/papers/w0113.pdf

Wasserman, S., and Faust, K. (2013). *Análisis de Redes Sociales. Métodos y Aplicaciones.* Madrid: CIS.

Ysewijn, P. (1996). *About Software for Generalizability Studies (GT).* Switzerland: Mimeograph.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Systematic Observation of an Expert Driver's Gaze Strategy—An On-Road Case Study

Otto Lappi[1]*, Paavo Rinkkala[2] and Jami Pekkanen[1,2]

[1] Cognitive Science, University of Helsinki, Helsinki, Finland, [2] Traffic Research Unit, University of Helsinki, Helsinki, Finland

In this paper we present and qualitatively analyze an expert driver's gaze behavior in natural driving on a real road, with no specific experimental task or instruction. Previous eye tracking research on naturalistic tasks has revealed recurring patterns of gaze behavior that are surprisingly regular and repeatable. Lappi (2016) identified in the literature seven "qualitative laws of gaze behavior in the wild": recurring patterns that tend to go together, the more so the more naturalistic the setting, all of them expected in extended sequences of fully naturalistic behavior. However, no study to date has observed all in a single experiment. Here, we wanted to do just that: present observations supporting all the "laws" in a single behavioral sequence by a single subject. We discuss the laws in terms of unresolved issues in driver modeling and open challenges for experimental and theoretical development.

Keywords: naturalistic tasks, observational instruments, eye tracking, fixation classification, gaze coding, eye movements, driving, expertise

INTRODUCTION

This paper takes a slightly unusual, even unorthodox, approach to studying car drivers' visual behavior. Observations of an extended behavioral sequence of just one highly experienced driver are used as a case study of the rich and varied strategies experienced subjects employ in natural tasks.

Even the apparently simple task of driving down a winding country road involves sophisticated gaze information processing and behavior (at this point wish to take a peek at the **Supplementary_Movie_1_full_video.mp4** which we will be analyzing in this observational study). It is only the ease with which humans are able to routinely perform such tasks belies their underlying complexity. This complexity is starkly revealed in artificial intelligence and robotics (e.g., the design of autonomous cars), where real-time interaction with complex 3D environments has turned out to be one of the most daunting tasks for a machine to perform. Much of the complexity is not well revealed, however, by highly simplified laboratory tasks, where a subject is asked to maintain fixation on a "fixation target" or pursue a "pursuit target" (typically small dots or geometrical shapes on a blank background) or search for a "feature conjunction target" among distractors. None of these tasks reveals the dynamic interplay of gaze, action and the spatial world. Yet the active strategies of natural behavior is something we need to understand if we are to understand vision—for if we do not understand how humans actively sample and organize the information available in rich naturalistic stimuli, we do not understand the visual input to the brain (Ahissar and Assa, 2016; Lappi, 2016).

Based on the data available in the literature—collected in different experiments with different tasks and presented from the point of view of different theoretical interests—it can be difficult to get a "feel" for the behavioral patterns. Especially so for researchers who are not themselves involved in eye tracking, and have not accumulated the type of tacit knowledge that comes from working with raw data. There are two main reasons why this is particularly true in eye tracking. One is that the *quantitative* patterns are still most often given using very coarse *aggregate measures* such as the horizontal dispersion in "visual search" (e.g., Underwood et al., 2002) or area of interest dwell times aggregated over a "trial" (e.g., Foulsham et al., 2011). This type of aggregate representation tremendously simplifies statistical analyses, but does not reveal the role individual fixations play in the underlying strategies—what information might be gleaned from the fixations, or how they support the on-going behavior. Yet work on natural task gaze strategies has consistently shown that individual fixations focus with high specificity on targets most relevant to the task (Land, 2006; Tatler et al., 2011). This adaptive character of individual fixations is rendered unobservable when the eye movements are not looked at the level of fixations, but instead aggregated into scalar variables such as horizontal gaze direction variability, AOI gaze catch %, fixation counts or saccade counts. The other reason is that *qualitative* descriptions of "what the participants were looking at" are usually given only *verbally*, or with the help of a single still image from a gaze video (e.g., Figure 13 in Land, 2006; Figures 4, 5 in Wann and Wilkie, 2004). Neither way of representing the observations is conducive to giving the reader an overall understanding or the dynamical aspects of the phenomenon: how gaze target selection, and more generally gaze-interaction with the complex natural settings, evolve over time.

One of the goals of this study is, therefore, to show the richness of natural gaze behavior in an active real-world task—driving. Driving is in many ways an ideal domain to study real-world locomotion. For one thing, it is easier to instrument a car than a pedestrian for reliable measurement. Also constraining limb actions to steering wheel and pedal movements in itself brings about a useful reduction in degrees of freedom, the road environment is typically more stereotypical and simple in layout than most of our locomotor surroundings. Finally, the driving task can be studied at all levels of expertise from driving school students to professional drivers, with most of the adult population in modern countries potential test subjects somewhere in between.

Driving, and the way we use eye movements to sample the spatial world from behind the wheel, are also often invoked as an important real-world application of models of attention, perception, or memory—at least in introductory vignettes (e.g., Regan and Gray, 2000; Wolfe and Horowitz, 2004). Kowler (2011) in reviewing 25 years of research into "how eye movements cope with real world visual and cognitive demands" identified driving (along with tasks such as reading, and sports) as a core task the understanding of which would reveal much of interest about how the visual system works.

Driver eye movements have been investigated for decades, in both real world driving (e.g., Shinar et al., 1977; Land,

1992; Land and Lee, 1994; Green, 2002; Lappi et al., 2013b; Lehtonen et al., 2014) and driving simulator experiments (e.g., Wilson et al., 2007; Mars, 2008; Mars and Navarro, 2012; Lemonnier et al., 2015). Yet, if we want to understand driver gaze behavior in terms of where people look (the identity of each fixation target) and when (how they sequentially allocate gaze time to multiple parallel targets), then the extensive literature actually presents a somewhat fragmented picture of specific tasks and experimental settings, but no "overall story." The main goal of this paper is to describe and portray in their natural ecological context strategies that have been revealed in quantitative experimental work. We use prior experimental work to systematize our qualitative analysis: eye movements are fleeting and difficult to pick up and codify through observation with the naked eye. We therefore organize our observations of the gaze behavior of our expert subject by organizing them in terms of seven typically recurring aspects of gaze control in naturalistic tasks—the "seven qualitative laws of gaze behavior in the wild" (Lappi, 2016)—which have been observed severally in a wide variety of tasks, but to date never yet all together.

"Seven Qualitative Laws" of Natural Gaze Behavior in the Wild (and Their Application to Driving)

Eye tracking in naturalistic tasks has begun to reveal recurring patterns of gaze behavior that turn out to be surprisingly regular and repeatable. Based on, and extending, previous reviews (Regan and Gray, 2000; Hayhoe and Ballard, 2005; Land, 2006; Kowler, 2011; Tatler and Land, 2011; Tatler et al., 2011) seven "qualitative laws" of gaze behavior in the wild were identified by Lappi (2016). (L1–L7, **Table 1**) These "laws" were defined as recurring patterns that tend to go together, the more so the more naturalistic the setting, all of them expected in most extended sequences of fully naturalistic behavior.

However, to date they have been observed singly or a few at a time in tasks as various as making tea (Land et al., 1999), making a sandwich (Hayhoe et al., 2003), drawing from a model (Land, 2006), steering a car (Land, 1992; Land and Lee, 1994; Lappi et al., 2013b) and a number of sports such as tennis (Ripoll and Fleurance, 1988), cricket (Land and McLeod, 2000; Mann et al., 2013), and squash (Hayhoe et al., 2012). That is, no empirical study (or review) to date has exhibited these common recurring patterns in a single task. Thus, to bolster the claim (Lappi, ibid.) that *all* these patterns would be expected to be present in extended sequences of fully naturalistic behavior we wanted in the present study to present observations of all the "laws" in a single behavioral sequence of a single subject. We use driving as our behavior of choice, following and building on the analysis in Lappi (2014).

We next proceed to describe the observational methodology used in this study. Then, in the Results and Discussion section we analyze the gaze patterns qualitatively, i.e., in terms of how and what they reveal about the seven qualitative laws of gaze behavior in the wild. We close with a discussion of open issues and challenges to existing perceptual-cognitive and

TABLE 1 | General gaze strategies in naturalistic tasks, and how they are manifested in driving.

Seven qualitative laws	Examples in driving	Notes
L1. Gaze patterns are highly repeatable and stereotypical	The way the driver scans the road surface (i) visual orientation to the curve apex during curve approach and turn-in (sometimes called "tangent point orientation," Land and Lee, 1994) (ii) optokinetic nystagmus (OKN) elicited during cornering (Authié and Mestre, 2011; Lappi and Lehtonen, 2013; Lappi et al., 2013b; Itkonen et al., 2015).	Stereotypy is found within and between individuals, given task constraints and physical context.
L2. Gaze is focused on task-relevant objects and locations	Scanning intersections, traffic, potential hazards etc. (e.g., Lemonnier et al., 2015). In curve driving, gaze concentration in a fairly small visual region in the view of the road ahead (some 10° horizontally, 5° vertically) (e.g., Land and Lee, 1994; Lappi et al., 2013b; Lehtonen et al., 2014).	Top-down control, rather than the visually most salient ones repeatedly "capturing" gaze.
L3. Individual fixations have interpretable functional roles	In the curve driving literature steering models attempt to account for how *Guiding Fixations* to various "steering points" may produce visual input for different control mechanisms (Land, 1998; Salvucci and Gray, 2004; Boer, 2016; for reviews see Wann and Land, 2000; Lappi, 2014). See also *look-ahead fixations* (L5) and Results and discussion for their possible role(s).	The roles are not always intuitive. The pattern of gaze can often be surprising to the subject as we are usually relatively unconscious of our eye movements, and cannot report them verbally.
L4. If possible, targets are fixated "just in time" (Ballard et al., 1995) 4.1. In complex tasks gaze is tightly coupled to the information requirement of the imminent subtask 4.2. Typically gaze leads action by about 1 s	In driving, the gaze-to-steering time delay is typically about 1–2 s (Land, 1992; Chattington et al., 2007; Land and Tatler, 2009), and the time headway typically about 2 s (Lappi and Lehtonen, 2013; Lehtonen et al., 2014). Lead time is the delay from gaze shift toward the bend to steering input. Time headway is the path distance from current location to the point of fixation divided by speed. These properties have been used to define operationally *guiding fixations* in driving (Lappi and Lehtonen, 2013), cf.L3 above.	"Just in time" means the moment they become relevant for guiding the next action, in contrast to than scanning the targets well ahead of time (requires cognitive resources for maintaining information in short-term memory, and carries risk the obsolescence of that memory). Unless a(sub)task requires continuous monitoring/tracking, gaze disengages—i.e., switches to a new target—before (sub)task completion.
L5. Visual sampling is intermittent[i] 5.1. In sequential tasks "just–in–time fixation" (or guiding fixations) are interleaved with look-ahead fixations 5.2. In dual-task performance guiding fixations (and look-ahead fixations) are interleaved with fixations to targets relevant to a parallel task	Look-ahead fixations have been identified in approach to a bend (Lappi and Lehtonen, 2013; Lehtonen et al., 2013, 2014, cf. gaze polling in Wilkie et al., 2008). Glancing at the instruments can be interpreted as another form of intermittency, i.e., gaze time shared between tasks (Johnson et al., 2014).	In most tasks, we do not "stare" at a single target for a prolonged period of time, but we instead sample the visual world with fixations lasting typically 100–500 ms interspersed by rapid, intermittent, saccades when visual input is degraded and actively suppressed by the visual system (see e.g., Land and Tatler, 2009). Visual information is not available to the brain continuously but as fairly discrete samples.
L6. Memory is used to (re)orient in 3D space	Long-term memory contribution to driving has been little studied, but Shinoda et al. (2001) show that traffic signs are more reliably noticed when they occur in locations that would be expected by prior knowledge of the traffic system than unexpected locations. Land and Tatler (2001) discuss the visual strategy of an expert racing driver in terms of a rich memory representation of the lap, and the possibility of orienting gaze, head and the vehicle in a way that takes into account road geometry beyond the range currently in view.	This can be done even to targets currently outside the field of view, implying trans-saccadic spatial memory (Tatler and Land, 2011). Also the fact that there are few fixations to irrelevant objects (i.e., visual search) implies stable contextual representation of where the task–relevant objects and locations are in 3D space (cf. e.g., Land et al., 1999).
L7. Gaze control is always part of "embodied" eye/head/body/locomotor control	On the one hand, gaze shifts are achieved by rotating not only the eyes but also rotating and translating the head and the body. On the other hand, head and body movements in space are compensated for by gaze–stabilizing eye and head rotations. For the brain, scanning and fixating targets and changing one's point of vantage are not necessarily separate "modular" tasks where the total output (gaze) would be a linear sum of separate locomotor, head and oculomotor systems (see discussion in Steinman, 1986; Collewijn et al., 1992; Lappi, 2016).	One shortcoming of visual steering models as interpretations of fixation behavior (L3) is that they only formalize the steering response as a function of visual information that is assumed to be available through appropriately coordinated gaze behavior (Donges, 1978; Salvucci and Gray, 2004). I.e., they do not discuss how fixation behavior itself is organized as part of a general gaze-steering strategy (but see Land, 1992; Land and Furneaux, 1997).

[i] This intermittency formulation is more general than the guiding fixations/look-ahead fixations formulation in *Lappi (2014)* and *Lappi (2016)*. *Guiding fixations and look-ahead fixations are defined in terms of visual requirements of different phases of a single "task". Sharing gaze time on the other hand happens "between tasks". But that definition presupposes an a priori delination of task strucutre, which in naturalistic tasks is not trivial. In driving, say, is operating the vehicle to be understood as one task, and monitoring traffic (via mirrors) another task (gaze time sharing between tasks)? Or should mirror-checking be interpreted as preparation for subtasks, such as a lane change (GF/LAF within a task)? Or is even interleaving guiding and look-ahead fixations in curve negotiation to be understood as sharing gaze time between control and anticipation tasks? The distinction seems more semantic than substantial, at least unless a highly rigorously specified task model is available.*

control theoretical driver models arising from observing real-world behavior in context.

METHODS

Participant

The subject was a 43 year old male licensed driving school instructor with 25 years of driving experience and 18 years of experience in professional driver education. He was recruited by personal contact. He had normal uncorrected vision and a valid driver's license. He reported no medical conditions that might affect eye movements.

Ethics Statement

This study was carried out in accordance with the recommendations of Finnish Advisory Board on Research Integrity. The protocol was approved by the ethics committee of the Faculty of Behavioral Sciences, University of Helsinki. Written informed consent in accordance with the Declaration of Helsinki was obtained from the participant. This was done in the form of a fixed-format consent form explaining the purpose of the study, the procedure, and intended use of the data (publication of anonymous data for scientific purposes). A paper copy of the consent form was archived.

Test Site, Equipment and Procedure

The test road (Velskolantie, Espoo: N 60.273951, E 24.654733) was a 5.13 km low-standard two-lane rural road (5.5 m pavement width, painted edge lines) with low traffic density. The test vehicle was a MY 2001 Porsche Typ986 3.2 (trade name "Boxster S") with a manual transmission (Dr. Ing. h.c. Ferdinand Porsche AG, Stuttgart, Germany). The car was not familiar to the driver, but as he was an expert with a wide experience of operating different vehicles he displayed no apparent difficulty with the controls, adapting immediately. The eye tracker was a Pupil Labs Binocular 120 (Pupil Labs UG haftungsbeschränkt, Berlin, Germany). The headset has a forward-looking world camera with an approximately 100° (horizontal) by 56 degrees (vertical) field of view, and two eye cameras. The sampling rate for the eye cameras was set to 30 Hz. The Pupil software with in-house custom code ran on an ASUS Zenbook UX303LB 2.4 GHz, with Linux Debian 4.2.6. and kernel 4.2.0. A custom built headband was used to secure the headset more firmly.

Upon arriving at the test site, the participant was briefed on the procedure, after which he filled the informed consent form. The driver was shown the test route on a map, and explained that the instruction was simply to drive the route "as they normally would." After adjusting the driving position, the eye-tracker was calibrated, and the calibration accuracy was immediately checked by the same 15-point procedure (see below). The researcher operating the eye tracker (PR) in the passenger seat gave instructions at crossroads leading to and from the test route proper. There were no intersections or crossroads on the test route. The road was run in both south-north and north-south directions. A post-calibration was then performed allowing us to determine calibration accuracy and also to improve it offline in post-processing.

Eye Tracker Calibration

The eye tracker was calibrated using 15 points in the visual field (**Figure 1**). Note that rather than presenting targets at 15 physical locations, a single target (about 5 m in front of the vehicle) was used, and the participant was asked to adopt different head poses, moving the target to different parts of the (head-referenced) visual field. Extensive pilot testing was done to arrive at a protocol whereby the instructions are clear and natural to the participants and they can follow them in an efficient and
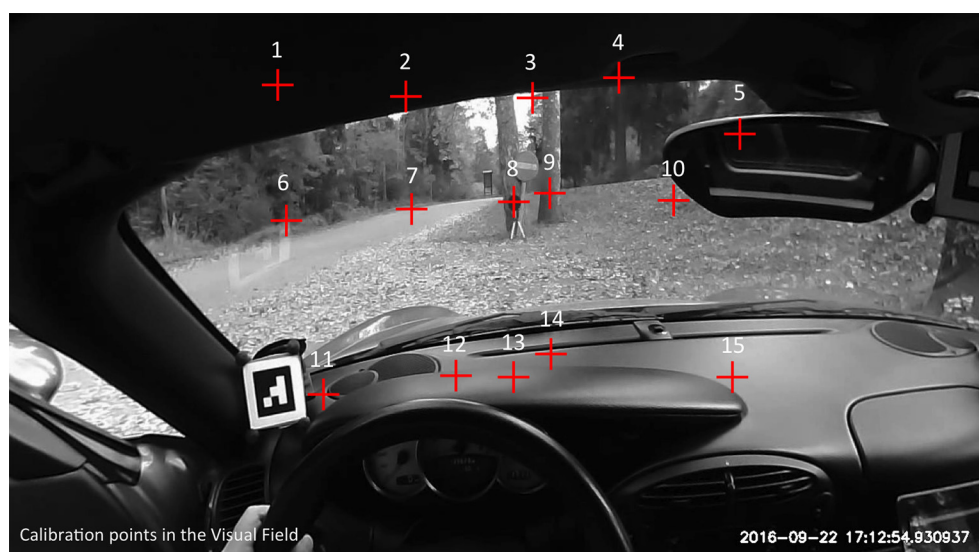
FIGURE 1 | Locations of the calibration points in the (head centered) visual field. Note that instead of multiple targets, we used a single target (on the tripod, at fixation point 8). The participant was instructed to adopt different head poses that moved the target to different locations in the visual field.

repeatable way. While this method does not give us *complete* control of the positioning of the target locations in the field of view, it nevertheless has a number of advantages. From a practical point of view, it does not require a large and cumbersome 15-point calibration frame to be transported—a single target on a tripod suffices. Second, the target can be placed at a large distance (rather than, say, on the vehicle bonnet or even at an arm's length inside the cockpit), thereby reducing parallax error.

Post-processing

In post-processing camera lens distortion was corrected using OpenCV v.2.4.9.1. CV2 undistort tool. Calibration stability was checked manually, and where it was deemed bumps or headset movement had shifted the calibration, it was adjusted manually. These adjustments were small, and mainly to the vertical coordinate. Time stamp (ts) of each frame based on Unix time stamp was burned into the video. This produced the final gaze-overlay video for analysis.

The video was inspected visually in slow motion for recurring typical fixation patterns, and an iterative method was used to arrive at a codification of gaze targets with the following desiderata: (i) It should be as free as possible from any specific theoretical or functional interpretation, that is, it should not be confined to any specific theoretical point of view in the literature on driver eye movements but accommodate all, (ii) the classification should allow for a reasonably unambiguous classification of all fixations in the video, (iii) the classification should give a good balance between categorizing *all* fixations in an *informative* way, but with as few categories as possible. (iv) the fixation classes should be mutually exclusive, that is, each fixation should be categorizable into one and only one class. This classification was the basis for our General Observations that are intended to characterize the overall pattern in the present gaze data (presented in Section General observations).

After this initial rough classification was in place, a more detailed analysis of episodes most relevant to the core task of steering the vehicle was done. Specific bends were selected on the basis that they should contain sufficient variability and detail to allow meaningful discussion of "the seven qualitative laws of gaze behavior in the wild" (see Section Introduction). The selected episodes were annotated, marking the beginning and end point and a putative classification of each fixation using a custom video annotation tool (https://github.com/jampekka/scvideonaxu). The rule used was that for a sequence of gaze positions to qualify as a fixation, gaze should remain stable at a fixed position or a fixed target object or location for a minimum of three frames (~90 ms). This putative classification was then discussed and refined in debriefing sessions within the research group to come up with a final classification. Episode videos were then prepared that display the annotations overlaid on the video (**Supplementary Movies 2–5**), serving as basis for more detailed illustrations of the more general "the seven qualitative laws" [presented in Section Detailed description of selected episodes (illustrations of the "qualitative laws")].

RESULTS AND DISCUSSION

We find it first useful to *show* the entire gaze video to give the reader a "feel" of the dynamical characteristics of gaze behavior on the road (rather than just individual frames and descriptions of the *spatial* fixations locations, or time *series* gaze data without the physical context). The full video is given as supplement (**Supplementary_Movie_1_full_video.mp4**). Observing this video will give the reader an idea of the richness and complexity of natural behavior even in a fairly controlled setting with little traffic, no intersections, no expansive vistas etc. The reader familiar with, say, the driver modeling literature and eye tracking experiments can from here get a feel for the gaps in present experimental work and models, which may hopefully inspire development of future experiments and models.

However, eye movements are fleeting and because we have poor conscious access to our own eye movement behavior it may be difficult to develop intuitions and identify the patterns through untutored observation with the naked eye. Here prior experimental work and models can, conversely, be useful in providing a framework for interpreting the observed behaviors, rendering "observable" behaviors that could otherwise be easily missed.

We will first present general observations about the overall pattern of how the driver scans the scene (in terms of a rough fixation classification), on the entire behavioral sequence. In the next section we will look in more detail at some "episodes," behavioral subsequences, which we consider to reveal interesting phenomena when looked at in light of the seven "qualitative laws." In both cases we present still images as figures, and verbal descriptions in the main text, to give the reader anchoring points and communicate our interpretations. But we urge the reader to consult the videos given as Supplementary Material.

General Observations

The first and perhaps most striking features are the high frequency of gaze shifts (gaze lability) and the amount of head movements. These are the features most people almost immediately and unprompted have remarked on when they have seen the video, as we are usually not *aware* of this lability of our eyes, and the frame in which they are supported (the head). This is because when the head and eye rotation are *actively controlled*, the perceptual system knows the motor command send to the eye/head system (efference copy), and can predict and thereby take into account sensor motion in creating a stable percept and maintaining orientation (for discussion see Angelaki and Hess, 2005; Burr and Morrone, 2012; Ahissar and Assa, 2016).

It is clear from the video that our driver never "stares" at any particular location or object for any extended period of time. Instead, the entire scene is scanned all the time with the rapid saccade–fixate–saccade pattern characteristic of visual (and) manual exploration. Remember that the participant is a driving school instructor: this behavior is consistent with the instruction given in Finnish driving schools to continually "rotate the gaze." **Figure 2** gives an overview of the "scanning pattern."

The fixations could be partitioned into different categories in a number of ways. One rather natural classification scheme is given
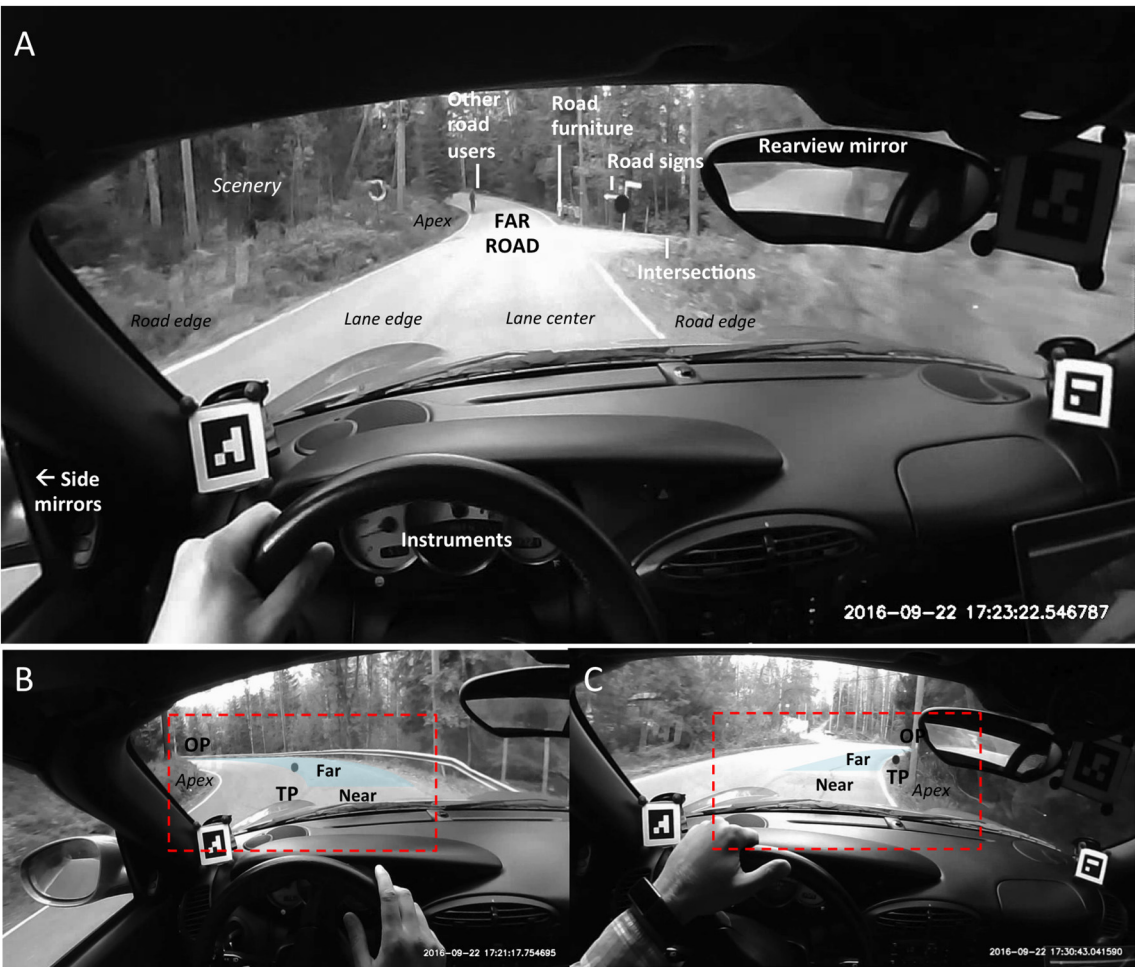
FIGURE 2 | (A) Descriptive classification of fixations used in this study, shown here in representative video frames on a straight. **(B,C)** Illustrate The Far Road "triangle" in a left turn **(B)** and a right turn **(C)**. See also **Figure 4**. OP, Occlusion Point; TP, Tangent Point. See main text for explanation.

in **Table 2** (cf. **Figure 2**). We use this classification to organize our general observations (G1–G7) of how each of the seven target classes figure in the overall scanning pattern.

G1 The Driver Tends to Keep His Eyes on the (Far) Road, Unless Other Relevant Targets Present Themselves, and Always Quickly Returns to It

After scanning for other targets, gaze always returns to the road ahead. From inspecting the video it is clear that in terms of dwell time the (far) road would be the predominant gaze category. Note that the gaze almost exclusively seeks out the "far" road region, as opposed to the road immediately in front of the car. This is consistent with the two-level/two point control models (Donges, 1978; Land, 1998; Salvucci and Gray, 2004; Boer, 2016) that are based on the assumption that experienced drivers use gaze to obtain visual *preview* of road geometry used for *anticipatory* control (matching the *predictable* road curvature), as opposed to near road information for *compensatory* control (maintaining lane position against *unpredictable* perturbations). That there are very few fixations to the near road does not

TABLE 2 | Overall fixation target classification and general observations (G1–G7) about the video.

Fixation target class	General observation
The (Far) Road	G1. The driver tends to keep his eyes on the (far) road, unless other relevant targets present themselves, and always quickly returns to it.
Instruments and Mirrors	G2. Instruments and mirrors are checked regularly.
Road users (traffic)	G3. Other road users in view are monitored, often with repeated fixations.
Intersections (side roads)	G4. Side road intersections are usually checked with a sideways glance.
Road Signs	G5. Most road signs are checked with a sideways glance.
(Other) Road Furniture	G6. Other road furniture is occasionally checked with a sideways glance.
Scenery	G7. "Scenery" not otherwise specified is rarely fixated.

mean by any means that the driver would not be using visual information for stabilizing control, though. Indeed, it has been shown an experienced driver can monitor near road information

peripherally (Summala et al., 1996), freeing the experienced driver to allocate more attention and overt gaze to the far road region than a novice (cf. Mourant and Rockwell, 1972; Lehtonen et al., 2014).

Of course, this behavior is also consistent with driving instruction frequently exhorting beginning drivers to try to look far enough ahead. But how far is far? There is no specific distance or time distance in the literature that would define "far" vs. "near" road. Typically, time headway to the "far" region in bends assumed to be about 1–2 s (cf. Lehtonen et al., 2014), which fits well with modeling literature as well (Boer, 2016).

In bends, it is conventional to use the *lane edge tangent point* (TP) as the point for segregating "near" and "far" road space (Land, 1998; Salvucci and Gray, 2004; Lappi, 2014). This is the point on the inside of the curve at which the visual orientation of the lane edge reverses its direction in the driver's visual field (**Figures 2B,C**). Note that the TP is a *travel point* (i.e., a point that moves with the observer in the 3D scene frame of reference, even though it may sometimes remain stationary in the observer's egocentric frame). Actual distance and time distance to the TP in the 3D world is therefore variable, and how far is "far" thus depends on curve geometry (time distance also depends on driving speed).

Observing the video, we can see that in simple bends the TP and Occlusion Point (OP) together with the lane edge opposite to TP create a Far Road "triangle" (**Figures 2B,C**). For a good portion of the time, this gives as good qualitative characterization of where we look when "we look where we are going" on the road[1]. The *Occlusion Point* (OP) is the point furthermost part of the road to which a continuous, unobstructed preview of a possible trajectory (future path) is visible (Lappi et al., 2013a), i.e., the point where "the road disappears from view." Like the TP, OP also is a *travel point*; it moves ahead as the observer travels along the road and does not follow the local optic flow (**Figure 3**; clear illustrations are e.g., 17:28:16.79–17:20:28.15 and 17:30:40.29–17:30:46.02).

G2 Instruments and Mirrors Are Checked Regularly

The driver's visual field does not only cover the external 3D roadspace, of course, but also relevant targets in the vehicle frame. These include the instrument panel (speedometer, tachometer) and mirrors (rear view and side mirrors). These are scanned predominantly (but not exclusively) on straights (e.g., 17:30:46.02–17:31:06.26; 17:27:57.13–17:28:02.61), presumably because there is less task load than in curves (Tsimhoni and

---

[1]Methodological note to researchers used to making Area of Interest (AOI) based analyses. As an AOI the Far Road Triangle would make more sense in characterizing the *general pattern* than for example the TP cantered circular AOI used to investigate tangent point orientation (Land and Lee, 1994; Lappi et al., 2013b). Note, however, that as with the distance and time distance of TP the *size* and the *shape* of this AOI will vary dramatically, making any statistical analysis of gaze catch percentages or dewll times problematic. (For the related AOI overlap problem of fixed-size AOI gaze catch percentage/dwell time analysis see Lappi et al., 2013a; Lappi, 2014). There appears to be no *straightforward* and *general* way to bring the traditional AOI method from controlled lab studies to dynamic tasks. Methodological innovation is called for—because of differences in the "design of the stimulus," not just noise and sampling rate issues (for discussion see Lappi, 2015).

Green, 2001), and less need for visually monitoring the road ahead. Stabilizing steering control requires little overt gaze, and the much longer time headways to visual occlusion means that there is much less time pressure for spotting hazards.

G3 Other Road Users in View Are Monitored, Often with Repeated Fixations

Oncoming vehicles, or pedestrians/bicyclists coming the other way and being overtaken are monitored by fixations. (e.g., 17:22.09–17:22:16.53; 17:29:06.83–17:29:10.96 [cars emerging from blind bends]; 17:32:57.75–17:33:05.49 [coming up on two bicyclists simultaneously]).

G4 Side Road Intersections Are Usually Checked with a Sideways Glance

Whenever there is a side road or a road from a yard that intersects the road, the driver tends to scan it with a fixation or often multiple fixations (e.g., ts 17:20:34.56–17:20:40.79 [side roads on both sides of a straight]; ts 17:22:08.55–17:22:09.02 [sideways glance to a side road on the left side of a right hand bend]).

G5 Most Road Signs Are Checked with a Sideways Glance

Road signs are fixated—even from quite impressive distances. These include road signs proper (*speed limit* signs *poor road surface* and *bends* caution signs, *stop* sign), as well as street name and navigational instruction signs (e.g., 17:20:17.47–17:20:21.52 [speed limits, bends]; 17:20:25-48–17:20:28.40 [bumpy road]; 17:34:08.52–17:34:25-72 [multiple]).

G6 Other Road Furniture Is Occasionally Checked with a Sideways Glance

Post boxes and other mid-sized objects near the road are occasionally "checked out" (e.g., ts 17:24:37.57–17:24:38.45). Here the driver is likely using high-spatial-resolution foveal vision for detailed analysis and object recognition of a target already *localized* and *individuated* as distinct from the ambient background using peripheral vision.
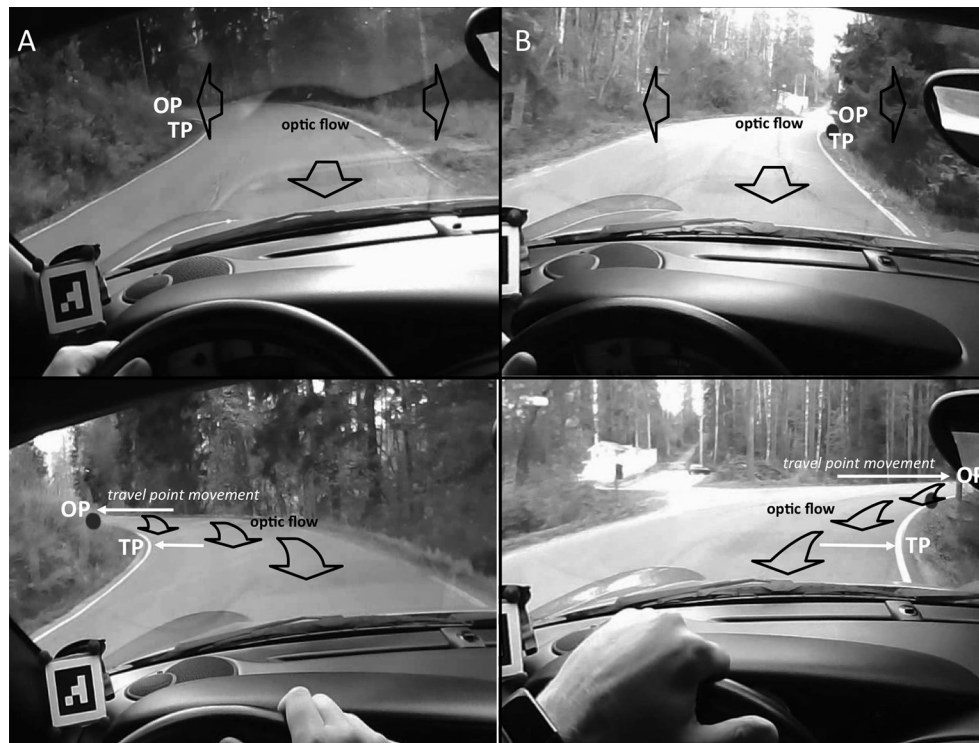
G7 "Scenery" Not Otherwise Specified Is Rarely Fixated

There are actually very few fixations at "scenery" not covered in the above categories. This is in itself important. The *absence* of any significant number of fixations on the general scenery, that is, the concentration of gaze on specific target and *the apparent absence of any visual search* in itself indicates that *peripheral visual information is used in a very efficient way to guide the gaze at the relevant locations with high accuracy and reliability*. Note that we therefore prefer to use the terms visual *exploration* or *scanning* rather than visual *search*, see further discussion below.

In sum, the general pattern of gaze coordination is the following: First, the default gaze mode is "eyes on the road," or "looking where you are going." Second, glances elsewhere are performed when *a specific relevant target* to look at has been identified (also, there must be "spare capacity" to allocate gaze time to non-steering related targets).

FIGURE 3 | Schematic illustration of how a bend "opens up" as the Occlusion Point travels up the road (and horizontally in the visual field), revealing more of the road. (A) Left hand bend. Top panel: approaching. Bottom panel: turning in. **(B)** Right hand bend. Top panel: approaching. Bottom panel: turning in. The Occlusion Point (like the tangent point) is a *travel point*, not a fixed 3D location in the scene. Travel point motion in the visual field (indicated by the white arrows) does not match the optic flow (indicated by the black block arrows). Fixating a travel point may be a tracking fixation, achieved with a pursuit movement. The same is clearly true also for tracking stationary fixed 3D scene objects or locations, but here the tracking will match optic flow. OP, Occlusion Point; TP, Tangent Point.

Detailed Description of Selected Episodes (Illustrations of the "Qualitative Laws")

We next illustrate "the seven qualitative laws" by selecting specific episodes from the extended behavioral sequence for more detailed fixation-by-fixation observation. To recap, the "laws" are:

L1. Gaze patterns are highly repeatable and stereotypical.
L2. Gaze is focused on task-relevant object and locations.
L3. Individual fixations have interpretable functional roles.
L4. If possible, targets are fixated "just in time."
L5. Visual sampling is intermittent.
L6. Memory is used to (re)orient in 3D.
L7. Gaze control is always part of "embodied" eye/head/body/locomotor control.

We will go through them in order, pointing out in each case relevant observations in the present data, open issues in the experimental modeling literature and deeper theoretical connections among the laws that may be non-obvious.

L1 Repeatable and Stereotypical Gaze Patterns

With a case study approach cannot tell from the data alone whether any stereotypy observed is idiosyncratic

to the participant. But when we find repeatable patterns that are reported in the literature we can consider them general.

Scanning the Far Road in bends. Perhaps the most robust coordination pattern is the fairly systematic scanning of the Far road (**Figures 2**, **4**) in bends. As the Occlusion Point travels up the road, revealing more of the scene behind (**Figure 3**), the gaze seeks out the road surface/inside road edge emerging into view.

This orientation toward the inside edge of an upcoming bend (the *apex* region) is sometimes called "tangent point orientation." But when one observes the actual scan pattern, it should clear there is much more complexity involved than the driver aiming to stabilize gaze on a single point. What does the full picture of scanning the far road in bends look like, then? *TP* orientation occurs, especially in the *approach* and *entering* phases of the bend (Land and Lee, 1994), but there are also fixations to lane edges further ahead, beyond the tangent point, and to the road surface in the Far Zone region. These *guiding fixations* (GF) are interspersed by saccades to make *look-ahead fixations* (LAFs) further into the bend and beyond. When the view up the road is occluded, as is the case on this road which runs through woods, some of these LAFs reach the OP (see **Figure 4**), but not all. As the driver enters the bend, fixations of the TP and the inside lane edge beyond the tangent point, and
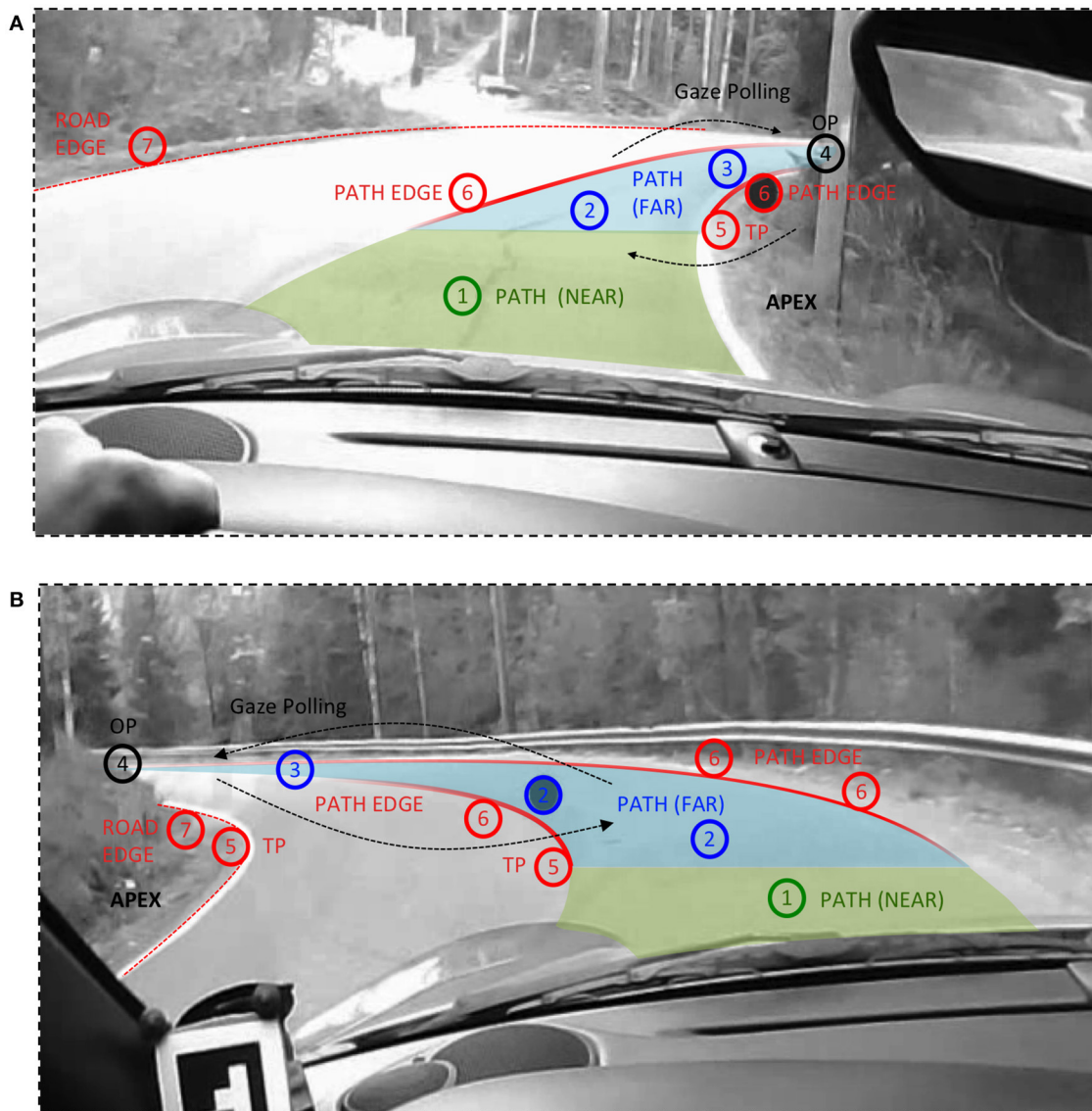
FIGURE 4 | Labeling of fixations at "the road ahead" in curve driving. NB. These underlying road images are the insets from **Figures 2B,C**. Far Road (blue) is defined as driver's own lane beyond the tangent point distance. Near Road (green) extends from in front of the vehicle to the tangent point level. (1) Path (near), (2) Path (far), (3) Path (look-ahead), (4) Occlusion Point (look-ahead), (5) Path Edge (tangent point), (6) Path Edge, and (7) Road Edge (tangent point).

fixations to the road surface in the Far Zone, mainly beyond the tangent point continue, interspersed by LAFs further up the road.

Because of this rich pattern and multiplicity of gaze targets, it is better to reserve the term *tangent point orientation* to fixations at or very near (within 3° of) the tangent point itself, performed at the very end of the approach and beginning of the entry (i.e., straddling the turn-in). This is according to the definition in the original Land and Lee (1994) study[2]. As a methodological side comment, note that because of the visual projection of

road geometry into the forward-looking visual field can bring these points very near to one another, making the definite determination of the actual gaze target of many individual fixations difficult, and traditional AOI methods unreliable (Lappi et al., 2013a; Lappi, 2014). But while in many individual cases the classification of a given single fixation could be ambiguous bases on instantaneous gaze position alone, the overall pattern of multiple gaze targets in the Far Zone is clear.

In the general case, due to the highly dynamic character of the way the road surface presents itself in the visual field (**Figure 4**), Far road fixations cannot be separated from LAFs

---

[2]Generalizing the term "tangent point orientation" to the entire scan pattern would be perhaps warranted if "steering by the tangent point" could be considered a *general* account of "where we look when we steer." But when there are

other theories, and the empirical picture is clearly more complex, this sort of terminological conflation is to be avoided.

by any hard-set distance or *gaze angle* criterion[3]. Instead, we propose that LAFs should be defined by the following *return saccade*. That is, in a LAF a fixation on the TP/far road/road edge is followed by a saccade further up the road (but not necessarily all the way to the OP), and a *return saccade* back to the road surface/lane edge closer to the vehicle (i.e., *gaze polling*, Wilkie et al., 2008). This "zig-zagging" pattern is very evident in most bends—for example the right hand bend in the beginning of **Supplementary_Movie_2_17_21_31_rl.mp4** and the left hand bend in **Supplementary_Movie_3_17_23_48_l.mp4**—and we have seen it to occur more or less frequently in the raw data of every driver we have ever tested in our previous studies.

L2 Gaze Focused on Task-Relevant Object and Locations

Although, the scanning patterns in driving are sometimes called "search" (primarily when complex situations call for identification and interpretation of multiple potential *hazards* e.g., Underwood et al., 2003; Wolfe and Horowitz, 2004), for the analysis of the core cognitive requirements of driving (high-speed vehicle control) we prefer the term *scanning* or visual *exploration*, as there is hardly any evidence of visual search proper—at least not in the way the term is used in experimental psychology. In visual search paradigms eye movements are used to look for a target among distractors, *where the* target is *masked* by the clutter and there is hence substantial *uncertainty over target location* (as in a typical Feature Integration Theory paradigm search matrix, or a Where's Waldo? image). The way the fixations appear to *immediately* find specific targets even at quite impressive distances (the road, other road users, traffic signs, see e.g., fixations to the three traffic signs in **Supplementary_Movie_2_17_21_31_rl.mp4**) suggests this is not the case in driving, as in fact we see very few fixations to "scenery" (general observation G7).

Intermittency combined with a high concentration of gaze on *relevant* targets (lack of search) implies efficient peripheral vision processes for target identification and saccade planning. Here we should bear in mind, though, that the participant is an expert driving school instructor, for whom traffic signs, for example, are highly relevant in terms of carrying out in-car instruction. Thus, he might have superior parallel covert "search" strategies compared to more typical drivers. Whether traffic signs are as "salient" to everyday drivers in (terms of being able spot them from distance with peripheral vision and "attracting" gaze) is not clear.

L3 Interpretable Functional Roles for Individual Fixations

Let us return to the far road fixation targets in curve driving. How can the individual fixations in the scanning pattern be interpreted? This is an important question, because the reason most eye movement research focuses on *fixation* behavior is that fixation is considered functionally as the "window" when

new visual information is available to the brain, punctuated by saccades during which relatively little information is received, and analysing where and when fixations are made is taken as a road to inferring underlying cognitive processes. We would like to point out here that for a moving observer, a fixation tracking a fixed target in the scene—and most travel points as well—is, from an *oculomotor* point of view, a pursuit movement (see discussion in Lappi, 2016).

In very broad terms, far road fixations can be considered simply "looking where you are going." But *interpreting* this strategy—*why* we should look where we are going in the first place—needs to take place in terms of *how* the brain processes the information gleaned from the fixation(s), and uses it real-time control of gaze and locomotion. Overt behavior (gaze position) does not uniquely specify the information that might be gleaned, and many interpretations for the (guiding) far road fixations have been put forward (for review see Wann and Land, 2000; Land and Tatler, 2009; Lappi, 2014). A detailed discussion of all the different interpretations, and their underlying theoretical motivations and commitments, are beyond scope of this paper. References to key papers are given in **Table 3**. The fixation classification scheme (**Table 3**, **Figure 4**) is intended to be compatible with any and all theoretical interpretations, i.e., not committed to any particular theoretical viewpoint or interpretation.

Far Path targets (**Table 3**, class 2) are postulated in several steering models to be steering points, as is the tangent point (class 5). (While the travel points and waypoints on the path would instantaneously occupy the same location, they will generally move in different directions in the visual field, cf. **Figure 3**, they will be useful for quite different steering strategies, see (Lappi, 2014) for detailed discussion). Look-ahead fixations (classes 3,4), LAFs on the other hand are considered to be different from such guiding fixations, because the tight gaze-steering coordination needs to be uncoupled during a LAF. They may nevertheless support higher-level trajectory planning (see L6, below). Note that we have used the term Path Edge (classes 5 and 6), which we define as those parts of the lane edges, in the Far region, which at any particular moment in time constrain available paths. (We reserve the term *lane edge* for the *entire* edges of the driver's own lane, extending beyond these "path edge" regions—even beyond the current field of view ahead and behind the vehicle). Road edge (classes 5 & 7) refer to the edge of the opposing lane.

L4 Targets Fixated "Just in Time"

Driving is a self-paced task in that the driver has a choice in the speed s/he wishes to travel at. However, once a speed is chosen the targets emerge at a given pace and obsolescence rate (cf. Senders et al., 1967; Kujala et al., 2015). This places a high importance on the accurate *timing* of fixations and saccades.

At an aggregate level this temporal coordination is reflected in a robust ca. 1 s gaze-action delay. Chattington et al. (2007) report gaze lead time (peak of gaze-steering cross correlation) of 0.98 s for 60 s epochs. We have observed similar cross correlations on the same road as used in this study for seven participants (*unpublished data* from Lappi et al., 2013a).

---

[3]When the curve geometry and speed are fixed, such criteria can be used to operationally define LAFs in *that particular physical context* (e.g., Lappi and Lehtonen, 2013; Lehtonen et al., 2014).

TABLE 3 | Labeling used, and interpretations found in the literature, for fixations in the Far road.

	Fixation class	Characterization	Interpretations available in the literature
1	Near path	The road surface immediately in front of the vehicle that the vehicle will imminently travel over (path).	*Compensatory control* in a two-level framework (Donges, 1978); *Near Point* in a two-point model (Salvucci and Gray, 2004).
2	Far path	The road surface ahead, in the bends at the tangent point level.	*Anticipatory control* in a two-level framework (Donges, 1978); *Far Point* in a two-point model (Salvucci and Gray, 2004). *Guiding fixations* on the *Future Path* (Boer, 1996, 2016; Wann and Land, 2000; Wann and Swapp, 2000; Wann and Wilkie, 2004; Wilkie et al., 2008).
3	Far path (look-ahead fixation)	The road surface further ahead, in the bends beyond the tangent point level.	*Guiding fixations/Look-ahead fixations* on the *Future Path* (Lehtonen et al., 2013, 2014). Gaze polling (Wilkie et al., 2008).
4	Occlusion point	The furthest point the road surface is continuously visible to.	*Look-ahead fixations*; trajectory planning and/or monitoring oncoming traffic (Lehtonen et al., 2013, 2014).
5	Tangent point (road/path edge)	Where the visual orientation of the lane edge reverses its direction.	Steering by the Tangent Point (Raviv and Herman, 1991; Land and Lee, 1994); *Far point* in a two-point model (Land, 1998).
6	Path edges	The edges of the driver's own lane in the far region, where they constrain the path the driver can choose.	Road geometry constraints on the *Field of Safe Travel* (Gibson, 1938); potential *Line Crossing* locations (Godthelp, 1986); *Safety Line* (Mars and Navarro, 2012).
7	Road edge	The edge of the opposing lane.	Potential *Line Crossing* locations?

i.e., class G1 in **Table 2**. *The numbers refer to* **Figure 4**.

At an individual fixation level, judgments of gaze-action delay depend on an interpretation of which action(s) each individual fixation actually supports. For steering related guiding and look-ahead fixations this question is still unresolved (cf. previous point). The most frequently referred to phenomenon remains the final fixation at or near the tangent point before turning into a bend (Land and Lee, 1994). This is clear for example in the blind right hand bend in **Supplementary_Movie_4_17_30_39_r.mp4**. However, given that the TP region is frequently fixated several times in anticipation (not just "just in time"), and that other locations in the Far Road Triangle are fixated in curve negotiation, the full picture of how *fixation timing* and *locomotor action timing* are related remains unresolved (cf. the next point).

The just-in-time strategy also implies the visual system needs to be able to recover either from peripheral visual information or from memory *where* the need-to-know information is at any moment in time. Cf. discussion of lack of search above (L2) and the role of memory below (L6).

L5 Intermittent Sampling

We remind the reader of the general observation that our driver never "stares" at any particular location or object for any extended period of time. Scanning the scene with the rapid saccade–fixate–saccade pattern happens all the time. That is, the overall pattern gives the intermittent "feel" that is characteristic of visual *exploration* (and exploratory behavior in other modalities, e.g., manual exploration).

Intermittency is clear for example in **Supplementary_Movie_5_17_30_55_lr.mp4** in the way looking "where you are going" is interspersed with fixations to a traffic sign and other road furniture, an intersection on

the right, and the side mirror. Also in the right hand bend in the beginning of **Supplementary_Movie_2_17_21_31_r.mp4** where the fixations in the Far Road region are interspersed by look-ahead fixations and sideways glances at a traffic sign. What are the implications of such a sampling pattern for control and cognitive processing?

Visual steering control models in psychology (for a review see Lappi, 2014) and driver models in vehicle dynamics engineering (for a review see Macadam, 2003) generally do *not* address this intermittency in visual input (but see Johns and Cole, 2015 for discussion and one of few experimental studies investigating the effects of intermittency of visual input to steering control; cf. "active gaze" approach in artificial intelligence and mobile robotics, Ballard, 1991; Fermüller and Aloimonos, 1995). Rather, the current state-of-art approaches in sensing and control are typically "reactive" systems, i.e., the system passively receives continuous input and produces output control signals in response. In contrast, in psychology and cognitive science it's well established—and apparent also in the present data - that in active dynamic tasks humans "proactively" sample visual information as needed, leading to input that is *intermittent*, and determined by the active observer (e.g., via eye movements) rather than imposed by the environment as a "forcing function." This allows humans to transcend their relatively slow information processing and limited sensory resolution to achieve impressively high aptitude in high-speed steering control (in driving and other domains).

On the other hand given that some relevant visual information may be available not *continuously* but as *discrete fixations*, critical action decisions may only be doable at certain points in time, or at least there are likely to be limited optimal "windows" for timing

locomotor action initiation relative to oculomotor actions. This issue has been studied in e.g., sports psychology in the literature on the quiet eye phenomenon (Vickers, 2016); but so far it has not been studied in the driving domain, nor modeled in driving models, beyond the above mentioned general 1 s lead time (cross correlation) between "apex orientation" and steering.

L6 Memory Used to (re)Orient in 3D

Memory processes cannot be readily "read off" from gaze behavior in our locomotor task—especially given that the route was only driven once (observing change in gaze behavior over multiple runs would be more informative of memory processes, as would analysis of landmark use in familiar surroundings, cf. Spiers and Maguire, 2007, 2008). Also note that the fact that in the present data even first fixations on a target are achieved without search, based on peripheral information (cf. discussion on L2, e.g., traffic signs) means that fixation-without-search in highly familiar surroundings (such as one's kitchen; Tatler and Land, 2011) can*not* be interpreted as proof of memory use. Also the just-in-time fixation strategy (L4) emphasizes the online nature of eye-hand-body coordination and "letting the world be its own model" (cf. Brooks, 1991)—as opposed to maintaining information in memory (which requires cognitive resources and faces the problem of that information becoming obsolete).

On the other hand memory and *intermittency* (L5) are deeply connected at a theoretical level, because it is to a large extent intermittency that makes memory (as opposed to pure online control) powerful. Anticipation allows humans to transcend their relatively slow information processing and limited sensory resolution. Modern cognitive theories of skilled action are predicated on the hypothesis that humans make predictions of the immediate future, choose actions on the basis of these predictions (for reviews of this *predictive* approach to anticipation and control, see e.g., Bubic et al., 2010; Henderson, 2017; for a critique of predictive control and defense of anticipation from merely *prospective* control see Zhao and Warren, 2015).

The relevant memory processes would be navigational long-term memory—an area of intense active research in the cognitive and computational neurosciences (Spiers and Barry, 2015). Integrating this literature to the theory of skilled driving would significantly advance out understanding of the driving task, and the role of these representations in real-world tasks generally.

Here we suggest one of the roles of look-ahead fixations—justifying the strategy of taking gaze time away from imminent needs of the primary control task—is to maintain and update this trans-saccadic memory of scene layout. That is, LAFS are relevant for steering (with a substantially higher delay than the 1 s for guiding fixations)—*both* for selection of, or parameter setting for, motor plans (updating "*inverse* models" in control theoretical terms), *and* creating a richer internal (*forward*) model of the state of the environment and the prediction of likely effects of action.

Precisely what *kind* of trans-saccadic memory underlies spatial orientation, and maintains our coherent experience of space (cf. Land and Furneaux, 1997; Tatler and Land, 2011; Burr and Morrone, 2012; Spiers and Barry, 2015) is an important but underappreciated issue in understanding (expert) driver behavior

L7 Gaze Control Part of "Embodied" Eye/Head/Body/Locomotor Control

One of the first remarks a number of people have made upon viewing the video (**Supplementary_Movie_1_full_video.mp4**) is expressing surprise at how much the participant's head moves. We experience the world as stable, even when the platform from which we observe it is not static. To achieve this *visual stability* in mobile contexts the brain must be able to take into account, in very sophisticated ways, both controlled (active, predictable) head movement and (passive, unpredictable) perturbations (Angelaki and Hess, 2005; Tatler and Land, 2011; Lappi, 2016).

Compensatory eye movements (vestibulo-ocular and optokinetic responses) and compensatory head movement stabilize gaze stable against unpredictable perturbances. Eye-head coordination is, on the other hand, *guided* by top-down attentional processes when target motion is *predictable*. A target moving in the visual field is tracked (pursuit) and large gaze shifts (saccade) achieved in part by synergistically turning the head, not just rotating the eyes in their sockets.

Synergistic eye/head gaze shifts are most apparent in fixations to the side mirrors. Pursuit of roadside objects, on the other hand, is usually done with eyes only, even for high eccentricities. In contrast, pursuit of OP is accompanied by head rotation anticipating vehicle rotation—even though the eccentricity is small, the gaze rotation anticipating or guiding locomotor rotation seems to recruit also head rotation, suggesting it is not only stimulus eccentricity but also stimulus relevance to ongoing motor action that is important in eye–head coordination. Even saccades and eccentric pursuit of *non-locomotor* targets can be done without a substantial head component (i.e., the head is kept aligned with locomotor path, side roads are checked with a sideways glance, for example), whereas orienting to the locomotor path (presumably for preview guidance information) involves a substantial head component—even when the eccentricity of the locomotor target is small. That is to say: the head is tightly coupled to steering-related "guidance" but less tightly coupled from non-steering related "scanning"—the head/gaze coupling is also *intermittent*.

GENERAL DISCUSSION

In this paper we present and analyse qualitatively an extended observational sequence where we have measured an expert driver's gaze behavior while driving on a real road with no experimental instruction. With this naturalistic task setting we hope to elicit typical behaviors this expert would use to cope with the complexity and ambiguity inherent in the real-world task of driving—within the limitations to the "naturalness" inherent in using an instrumented vehicle approach. This goal is heuristic[4]: to identify patterns one can "see" in naturalistic settings, but have not yet been codified in experimental procedures and quantitative contexts.

---

[4]In the traditional philosophical rather than the computer science sense of the word: organized, systematic activity aiming at empirical *discovery* rather than empirical *justification* (e.g., Lakatos, 1980).

Naturalistic (observational) and controlled (experimental) work should complement one another. On the one hand, controlled experiments run the risk of becoming too far abstracted away from the task constraints and behavioral strategies and patterns that actually make up the behavior of interest in the real world. Without real world observations complementing laboratory results, this may go unnoticed! Observing visual strategies in naturalistic real-world tasks can therefore provide important ecological validation to the design and results of lab experiments, or suggest ways to make laboratory designs more representative of real-world task settings. On the other hand, simply making notes of "where people look" does not produce good science. Using quantitative techniques to extend observational capabilities beyond those of the naked eye can allow one to record novel aspects of natural behavior. That is, quantitative measurement procedures that have been developed in experimental work can produce observational data for more qualitative analysis as well. At a more conceptual level, especially with eye movements which are so rapid and to which we have so little introspective access, prior laboratory/experimental work can be highly valuable, even essential, in coming up with the descriptive framework of concepts and procedures (the non-experimental observational *paradigm*, if you will).

We hope further systematic observational studies may be inspired by, and extend the results of, this one. This type of research is missing in the literature on naturalistic task gaze strategies but should be useful in moving between fully naturalistic settings and experimentally controlled tasks; in both directions and for mutual benefit. Of course, only controlled experiments will be able to reveal the internal workings of brain mechanisms—but at the cost of restricting the behavioral context to very restricted and often simplified tasks, and typically imposing artificial constraints whose effects on strategies may be unknown, yet substantial. We feel it is important to keep a balance between the goals of experimental rigor and faithfulness to the phenomena. Observational studies such as this one can also identify gaps in existing knowledge, e.g., by showing behavioral patterns in more detail, suggesting new data analysis procedures, or even completely new experiments.

AUTHOR CONTRIBUTIONS

OL: conceived the study, wrote the first draft and revised the manuscript; OL and PR: designed and piloted the data collection procedure; PR: collected the data; OL, PR, and JP: analyzed the results, prepared the videos/figures and wrote the final draft; JP: contributed the analysis/data annotation tools.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The supplementary movies are available via Figshare under a CC-BY license:

Supplementary Movie 1 Full Movie (1.37 GB) | doi: 10.6084/m9.figshare. 4498466

Supplementary Movie 2 (22.56 MB) | doi: 10.6084/m9.figshare.4498613

Supplementary Movie 3 (12.62 MB) | doi: 10.6084/m9.figshare.4498619

Supplementary Movie 4 (11.51 MB) | doi: 10.6084/m9.figshare.4498622

Supplementary Movie 5 (23.52 MB) | doi: 10.6084/m9.figshare.4498625

REFERENCES

Ahissar, A., and Assa, E. (2016). Perception as a closed-loop convergence process. *eLife* 5:e12830. doi: 10.7554/eLife.12830

Angelaki, D. E., and Hess, B. J. (2005). Self-motion-induced eye movements: effects on visual acuity and navigation. *Nat. Rev. Neurosci.* 6, 966–976. doi: 10.1038/nrn1804

Authié C. N., and Mestre, D. R. (2011). Optokinetic nystagmus is elicited by curvilinear optic flow during high speed curve driving. *Vis. Res.* 51, 1791–1800. doi: 10.1016/j.visres.2011.06.010

Ballard, D. H. (1991). Animate vision. *Artif. Intell.* 48, 57–86. doi: 10.1016/0004-3702(91)90080-4

Ballard, D. H., Hayhoe, M. M., and Pelz, J. B. (1995). Memory representations in natural tasks. *J. Cogn. Neurosci.* 7, 66–80.

Boer, E. R. (1996). "Tangent point oriented curve negotiation," in *Intelligent Vehicles Symposium, 1996, Proceedings of the 1996 IEEE*, 7–12. doi: 10.1109/IVS.1996.566341

Boer, E. J. (2016). What preview elements do drivers need? *IFAC-PapersOnLine* 4, 102–107. doi: 10.1016/j.ifacol.2016.10.469

Brooks, R. A. (1991). Intelligence without representation. *Artif. Intell.* 47, 139–159. doi: 10.1016/0004-3702(91)90053-M

Bubic, A., Von Cramon, D. Y., and Schubotz, R. I. (2010). Prediction, cognition and the brain. *Front. Hum. Neurosci.* 4:25. doi: 10.3389/fnhum.2010. 00025

Burr, D. C., and Morrone, M. C. (2012). Constructing stable spatial maps of the world. *Perception* 41, 1355–1372. doi: 10.1068/p7392

Chattington,. M., Wilson, M., Ashford, D., and Marple-Horvat, D. E. (2007). Eye–steering coordination in natural driving. *Exp. Brain Res.* 180, 1–14. doi: 10.1007/s00221-006-0839-2

Collewijn, H., Steinman, R. M., Erkelens, C. J., Pizlo, Z., and van der Steen, J. (1992). "Effect of freeing the head on eye movement characteristics during three-dimensional shifts of gaze and tracking," in *The Head-Neck Sensory Motor System*, eds A. Berthoz, P. P. Vidal and W. Graf (Oxford: Oxford University Press), 412–418.

Donges, E. (1978). A two-level model of driver steering behavior. *Hum. Factors* 20, 691–707.

Fermüller, C., and Aloimonos, Y. (1995). Vision and action. *Image Vis. Comput.* 13, 725–744. doi: 10.1016/0262-8856(95)98754-H

Foulsham, T., Walker, E., and Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vis. Res.* 51, 1920–1931. doi: 10.1016/j.visres.2011.07.002

Gibson, J. J., and Crooks, L. E. (1938). A theoretical field analysis of automobile–driving. *Am. J. Psychol.* 51, 453–471. doi: 10.2307/1416145

Godthelp, H. (1986). Vehicle control during curve driving. *Hum. Factors* 28, 211–221.

Green, P. (2002). "Where do drivers look while driving (and for how long)?" in *Human Factors in Traffic Safety*, eds R. E. Dewar and P. L. Olson (Lawyers & Judges Publishing Company), 77–110.

Hayhoe, M., and Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn. Sci.* 9, 188–194. doi: 10.1016/j.tics.2005.02.009

Hayhoe, M. M., McKinney, T., Chajka, K., and Pelz, J. B. (2012). Predictive eye movements in natural vision. *Exp. Brain Res.* 217, 125–136. doi: 10.1007/s00221-011-2979-2

Hayhoe, M., Shrivastrava, A., Mruczek, R., and Pelz, J. (2003). Visual memory and motor planning in a natural task. *J. Vis.* 3, 49–63. doi: 10.1167/3.1.6

Henderson, J. M. (2017). Gaze control as prediction. *Trends Cogn. Sci.* 21, 15–23. doi: 10.1016/j.tics.2016.11.003

Itkonen, T., Pekkanen, J., and Lappi, O. (2015). Driver gaze behavior is different in normal curve driving and when looking at the tangent point. *PLoS ONE* 10:e0135505. doi: 10.1371/journal.pone.0135505

Johns, T. A., and Cole, D. J. (2015). Measurement and mathematical model of a driver's intermittent compensatory steering control. *Veh. Sys. Dyn.* 53, 1811–1829. doi: 10.1080/00423114.2015.1100748

Johnson, L., Sullivan, B., Hayhoe, M., and Ballard, D. (2014). Predicting human visuomotor behaviour in a driving task. *Philos. Trans. R. Soc. B* 369:20130044. doi: 10.1098/rstb.2013.0044

Kowler, E. (2011). Eye movements: the past 25 years. *Vision Res.* 51, 1457–1483. doi: 10.1016/j.visres.2010.12.014

Kujala, T., Mäkelä, J., Kotilainen, I., and Tokkonen, T. (2015). The attentional demand of automobile driving revisited. *Hum. Factors* 58, 163–180. doi: 10.1177/0018720815595901

Lakatos, I. (1980). *The methodology of scientific research programmes: Philosophical papers Vol. 1*, eds J. Worrall and G. Currie (Cambridge: Cambridge University Press).

Land, M. F. (1992). Predictable eye-head coordination during driving. *Nature* 359, 318–320. doi: 10.1038/359318a0

Land, M. F. (1998). "The visual control of steering," in *Vision and Action*, eds L. R. Harris and M. Jenkin (Cambridge: Cambridge University Press), 163–180.

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Prog. Retin. Eye Res.* 25, 296–324. doi: 10.1016/j.preteyeres.2006.01.002

Land, M. F., and Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philos. Trans. R. Soc. B* 352, 1231–1239. doi: 10.1098/rstb.1997.0105

Land, M. F., and Lee, D. (1994). Where we look when we steer. *Nature* 369, 742–744. doi: 10.1038/369742a0

Land, M. F., and McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nat. Neurosci.* 3, 1340–1345. doi: 10.1038/81887

Land, M. F., and Tatler, B. W. (2001). Steering with the head: the visual strategy of a racing driver. *Curr. Biol.* 11, 1215–1220. doi: 10.1016/S0960-9822(01)00351-7

Land, M. F., and Tatler, B. W. (2009). *Looking and Acting: Vision and Eye Movements in Natural Behaviour*. Oxford: Oxford University Press.

Land, M., Mennie, N., and Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Percept. Lond.* 28, 1311–1328. doi: 10.1068/p2935

Lappi, O. (2014). Future path and tangent point models in the visual control of locomotion in curve driving. *J. Vis.* 14, 1–22. doi: 10.1167/14.12.21

Lappi, O. (2015). Eye-tracking in the wild: the good, the bad and the ugly. *J. Eye Mov. Res.* 8, 1–21. doi: 10.16910/jemr.8.5.1

Lappi, O. (2016). Eye movements in the wild: oculomotor events, gaze behavior and frames of reference. *Neurosci. Biobeh. Rev.* 69, 49–68. doi: 10.1016/j.neubiorev.2016.06.006

Lappi, O., and Lehtonen, E. (2013). Eye-movements in real curve driving: pursuit-like optokinesis in vehicle frame of reference, stability in an allocentric reference coordinate system. *J. Eye Mov. Res.* 6, 1–13. doi: 10.16910/jemr.6.1.4

Lappi, O., Lehtonen, E., Pekkanen, J., and Itkonen, T. (2013a). Beyond the tangent point: gaze targets in naturalistic driving. *J. Vis.* 13, 1–18. doi: 10.1167/13.13.11

Lappi, O., Pekkanen, J., and Itkonen, T. H. (2013b). Pursuit eye-movements in curve driving differentiate between future path and tangent point models. *PLoS ONE* 8:e68326. doi: 10.1371/journal.pone.0068326

Lehtonen, E., Lappi, O., Koirikivi, I., and Summala, H. (2014). Effect of driving experience on anticipatory look-ahead fixations in real curve driving. *Acid. Anal. Prev.* 70, 195–208. doi: 10.1016/j.aap.2014.04.002

Lehtonen, E., Lappi, O., Kotkanen, H., and Summala, H. (2013). Look-ahead fixations in curve driving. *Ergonomics* 56, 34–44. doi: 10.1080/00140139.2012.739205

Lemonnier, S., Brémond, R., and Baccino, T. (2015). Gaze behavior when approaching an intersection: dwell time distribution and comparison with a quantitative prediction. *Trans. Res. F Traffic Psychol. Behav.* 35, 60–74. doi: 10.1016/j.trf.2015.10.015

Macadam, C. C. (2003). Understanding and modeling the human driver. *Veh. Sys. Dyn.* 40, 101–134. doi: 10.1076/vesd.40.1.101.15875

Mann, D. L., Spratford, W., and Abernethy, B. (2013). The head tracks and gaze predicts: how the world's best batters hit a ball. *PLoS ONE* 8:e58289. doi: 10.1371/journal.pone.0058289

Mars, F. (2008). Driving around bends with manipulated eye-steering coordination. *J. Vis.* 8, 1–11. doi: 10.1167/8.11.10

Mars, F., and Navarro, J. (2012). Where we look when we steer with or without active steering wheel control. *PLoS ONE* 7:e43858. doi: 10.1371/journal.pone.0043858

Mourant, R. R., and Rockwell, T. H. (1972). Strategies of visual search by novice and experienced drivers. *Hum. Fact.* 14, 325–335. doi: 10.1177/001872087201400405

Raviv, D., and Herman, M. (1991). "A new approach to vision and control for road following," in *Proceedings of the IEEE Workshop on Visual Motion* (Princeton, NJ), 217–225. doi: 10.1109/WVM.1991.212804

Regan, D., and Gray, R. (2000). Visually guided collision avoidance and collision achievement. *Trends Cogn. Sci.* 4, 99–107. doi: 10.1016/S1364-6613(99)01442-4

Ripoll, H., and Fleurance, P. (1988). What does keeping one's eye on the ball mean? *Ergonomics* 31, 1647–1654. doi: 10.1080/00140138808966814

Salvucci, D. D., and Gray, R. (2004). A two-point visual control model of steering. *Perception* 33, 1233–1248. doi: 10.1068/p5343

Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., and Ward, J. L. (1967). The attentional demand of automobile driving. *Highway Res. Rec.* 195, 15–33.

Shinar, D., McDowell, E. D., and Rockwell, T. H. (1977). Eye movements in curve negotiation. *Hum. Factors* 8, 237–238. doi: 10.1177/001872087701900107

Shinoda, H., Hayhoe, M., and Shrivastava, A. (2001). What controls attention in natural environments? *Vision Res.* 41, 3535–3545. doi: 10.1016/S0042-6989(01)00199-7

Spiers, H. J., and Barry, C. (2015). Neural systems supporting navigation. *Curr. Opin. Behav. Sci.* 1, 47–55. doi: 10.1016/j.cobeha.2014.08.005

Spiers, H. J., and Maguire, E. A. (2007). Neural substrates of driving behaviour. *Neuroimage* 36, 245–255. doi:10.1016/j.neuroimage.2007.02.032

Spiers, H. J., and Maguire, E. A. (2008). The dynamic nature of cognition during way finding. *J. Environ. Psychol.* 28, 232–249. doi: 10.1016/j.jenvp.2008.02.006

Steinman, R. M. (1986). The need for an eclectic, rather than systems, approach to the study of the primate oculomotor system. *Vis. Res.* 26, 101–112. doi: 10.1016/0042-6989(86)90074-X

Summala, H., Nieminen, T., and Punto, M. (1996). Maintaining lane position with peripheral vision during in-vehicle tasks. *Hum. Factors* 38, 442–451. doi: 10.1518/001872096778701944

Tatler, B. W., Hayhoe, M. M., Land, M. F., and Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *J. Vis.* 11, 1–23. doi: 10.1167/11.5.5

Tatler, B. W., and Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 596–610. doi: 10.1098/rstb.2010.0188

Tsimhoni, O., and Green, P. (2001). "Visual demand of driving and the execution of display-intensive in-vehicle tasks," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 1586–1590.

Underwood, G., Chapman, P., Bowden, K., and Crundall, D. (2002). Visual search while driving: skill and awareness during inspection of the scene. *Transportation Res. Part F: Traffic Psychol. Behav.* 5, 87–97.

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., and Crundall, D. (2003). Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics* 46, 629–646. doi: 10.1080/0014013031000090116

Vickers, J. N. (2016). Origins and current issues in Quiet Eye research. *Curr. Iss. Sport Sci.* 1, 1–11. doi: 10.15203/CISS_2016.101

Wann, J., and Land, M. (2000). Steering with or without the flow: is the retrieval of heading necessary? *Trends Cogn. Sci.* 4, 319–324. doi: 10.1016/S1364-6613(00)01513-8

Wann, J. P., and Swapp, D. (2000). Why you should look where you are going. *Nat. Neurosci.* 3, 647–648. doi: 10.1038/76602

Wann, J. P., and Wilkie, R. (2004). "How do we control high speed steering?" in *Optic Flow and Beyond,* eds L. M. Vaina, S. A. Beardsley, S. K. Rushton (Norwell, MA: Kluwer Academic Publishers), 371–389.

Wilkie, R., Wann, J. P., and Allison, R. (2008). Active gaze, visual look-ahead, and locomotor control. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1150–1164. doi: 10.1037/0096-1523.34.5.1150

Wilson, M., Stephenson, S., Chattington, M., and Marple-Horvat, D. E. (2007). Eye movements coordinated with steering benefit performance even when vision is denied. *Exp. Brain Res.* 176, 397–412. doi: 10.1007/s00221-006-0623-3

Wolfe, J. M., and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* 5, 1–7. doi: 10.1038/nrn1411

Zhao, H., and Warren, W. H. (2015). On-line and model-based approaches to the visual control of action. *Vis. Res.* 110, 190–202. doi: 10.1016/j.visres.2014.10.008

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership