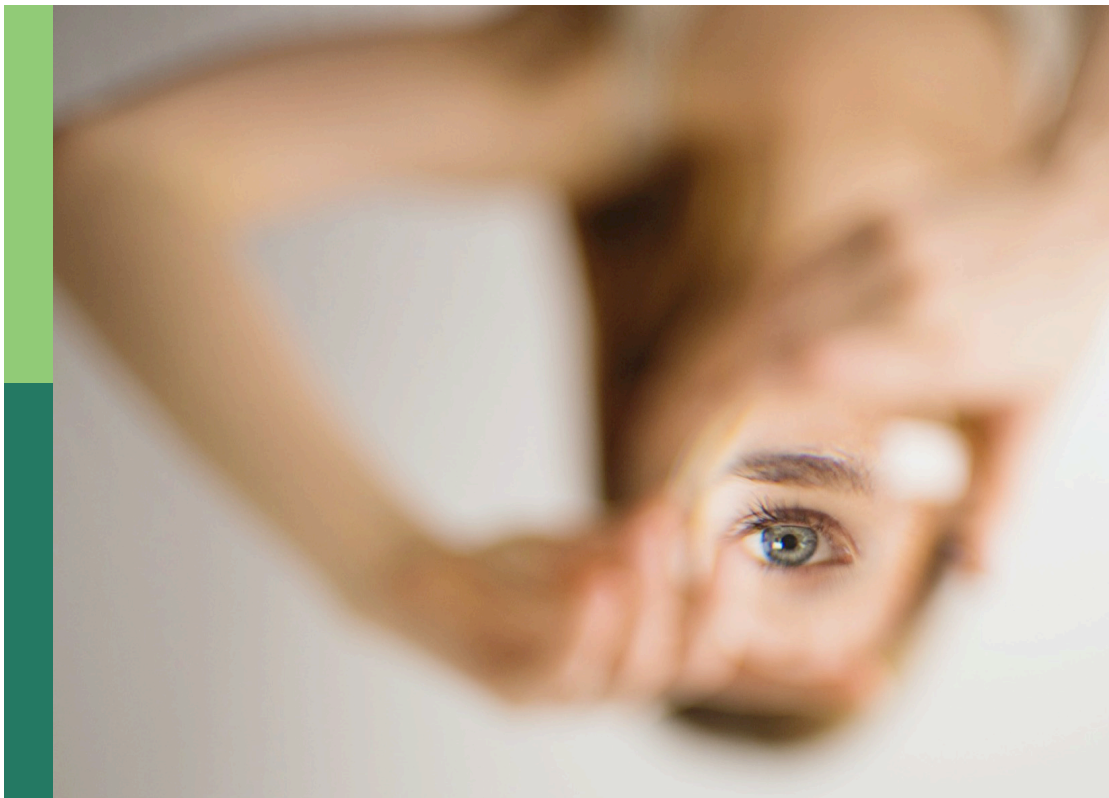
frontiers | Research Topics

Towards a basic standard methodology for international research in psychology

Edited by
Miguel Ángel Carrasco, Fco. Pablo Holgado-Tello, José Antonio Lozano Lozano, Susana Sanduvete-Chaves and Salvador Chacón-Moscoso

Published in
Frontiers in Psychology



ISSN 1664-8714
ISBN 978-2-83252-131-1
DOI 10.3389/978-2-83252-131-1

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: frontiersin.org/about/contact

Towards a basic standard methodology for international research in psychology

Topic editors

Miguel Ángel Carrasco — National University of Distance Education (UNED), Spain
Fco. Pablo Holgado-Tello — National University of Distance Education (UNED), Spain
José Antonio Lozano Lozano — Universidad Autónoma de Chile, Chile
Susana Sanduvete-Chaves — Sevilla University, Spain
Salvador Chacón-Moscoso — Sevilla University, Spain

Topic coordinator

Milagros Ocalin Sánchez Hernández — National Autonomous University of Nicaragua, León, Nicaragua

Citation

Carrasco, M. Á., Holgado-Tello, F. P., Lozano, J. A. L., Sanduvete-Chaves, S., Chacón-Moscoso, S., eds. (2023). *Towards a basic standard methodology for international research in psychology*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83252-131-1

Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 24 March 2023
DOI 10.3389/fpsyg.2023.1170108

OPEN ACCESS

EDITED AND REVIEWED BY
Pietro Cipresso,
University of Turin, Italy

*CORRESPONDENCE
Milagros Ocalin Sánchez Hernández
✉ msanchez5627@alumno.uned.es
Fco. Pablo Holgado-Tello
✉ pfholgado@psi.uned.es

SPECIALTY SECTION
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

RECEIVED 20 February 2023
ACCEPTED 06 March 2023
PUBLISHED 24 March 2023

CITATION
Sánchez Hernández MO, Holgado-Tello FP, Carrasco MA, Chacón-Moscoso S, Sanduvete-Chaves S and Lozano-Lozano JA (2023) Editorial: Towards a basic standard methodology for international research in psychology. *Front. Psychol.* 14:1170108. doi: 10.3389/fpsyg.2023.1170108

Editorial: Towards a basic standard methodology for international research in psychology

Milagros Ocalin Sánchez Hernández[1]*, Fco. Pablo Holgado-Tello[2]*, Miguel Angel Carrasco[3], Salvador Chacón-Moscoso[4,5], Susana Sanduvete-Chaves[4] and José Antonio Lozano-Lozano[5,6]

[1]Department of Psychology, National Autonomous University of Nicaragua at León (UNAN-León), León, Nicaragua, [2]Department of Methodology of Behavioral Science, National University of Distance Education (UNED), Madrid, Spain, [3]Department of Psychology of Personality, Assessment and Treatment, National University of Distance Education (UNED), Madrid, Spain, [4]Department of Experimental Psychology, Universidad de Sevilla, Seville, Spain, [5]Department of Psychology, Universidad Autónoma de Chile, Santiago, Chile, [6]Institute of Biomedical Sciences, Universidad Autónoma de Chile, Santiago, Chile

KEYWORDS

International Psychology, psychological research, methodology, data analysis, research designs, measurement

Editorial on the Research Topic
Towards a basic standard methodology for international research in psychology

There is a need to promote the creation of referential sources to help the improvement of global scientific practice. This could answer new challenges such as open science development and the application of research focused on diversity and inclusiveness. Consequently, leading to international psychological research that ensures the methods applied in analysis across countries and cultural groups are reliable and valid.

In this Research Topic, we address the role of methodology in International Psychology through the compilation of theoretical and empirical articles including original research and systematic reviews that explore psychological phenomena with a global focus, and from different areas of Psychology. This encompasses Clinical, Developmental, Educational, Cross cultural, Sport, and Work and Organizational Psychology.

We consider these articles are contemporary and innovational because the authors highlight and apply methodological issues related to the measurement of psychological phenomena, research designs, data collection techniques, and data analysis techniques across different countries.

Among the concrete contributions to this Research Topic, there are two articles that explore symptoms and the diagnostic criteria of psychological disorders. In Spain, findings on the early detection and prevention of psychosis are highlighted in Ceballos-Munuera et al.. They found a jointly partial mediation of Aberrant Salient and Disorganized Dimension between Ideas of Reference and the Psychotic Dimension across a continuum from general population to clinically diagnosed patients. Evidencing that a set of vulnerabilities (unusual thought content) could lead to a high-risk general pathological state and proneness to psychosis in particular. In Chile, Fioravante et al. found statistically significant differences between groups of school-age children with an Attention Deficit Hyperactivity Disorder (ADHD) diagnosis and without ADHD. This result indicated the importance of appropriate criteria and procedures to establish a diagnosis and implement effective interventions.

Regarding the effect of contextual events throughout life and mental health, there are some studies focusing on the changes and the experience of COVID pandemic. Díaz-García et al. found there was a variation in lifestyles of Spanish elders, but despite knowledge about the clinical aspects of SARS-CoV-2 and the application of recommended preventive measures, authors found certain relationships remained unchanged such as the one with family and friends. Adding to this, Carr et al. explore whether or not the COVID-19 pandemic was impacting pivotal links between living wages and employee attitudes and subjective wellbeing in twin cohorts of low-waged workers across New Zealand. They found the need of considering subjective wellbeing in the context of a crisis for employee livelihoods and lives.

To explore Development and Educational Psychology phenomena, some authors used a more theoretical approach with the use of Systematic Reviews Research. Alarcón-Espinoza et al. identify for research on emotional self-regulation in childhood and adolescence that most authors apply mixed methodologies and there are problems related to the recording of the duration and sequence of behaviors, highlighting the use of guidelines as guides for future research. While Tronchoni et al. try to highlight pedagogical concerns related to the use of reading in university teaching. Through this work, the authors verified the lack of research on proposals regarding how this pedagogical strategy evolves for a more interactive learning practice.

Furthermore, some studies focus more on theoretical-methodological operationalization of psychological phenomena. Smith et al. explore the process of what people from different cultures do when they faced with complex problems by presenting two computer-simulated dynamic problems in which they were required to act quickly or cautiously. Participants were asked to think aloud in their native language while working on the tasks and the protocols were digitally recorded, transcribed, and coded by coders from each country in terms of the steps involved in complex problem solving and dynamic decision making. Only those protocols with more than 15 transitions were analyzed. Results highlight the importance of process analyses in different tasks and show how cultural background guides people's decisions under uncertainty.

Mohamed et al. study the application of Taguchi algorithm (applied mostly in engineering) for the analysis of relationships of depression and obesity indices with other indicators as socioeconomic, screen time, sleep time, fitness use, and nutrition. Results found that, although Taguchi method can estimates correlations even undetected by applying other common statistics techniques with psychological data, scholars need to be cautious with statistical methods for measuring and estimating their research variables.

García-Franco et al. present a method for the study and assessment of personality disorders, especially those of the dramatic and emotional type. It consists in the application of a Bayesian network whose parameters have been obtained by the Delphi method of consensus from a group of experts. The result is a probabilistic graphical model that represents the psychological variables related to the personality disorders together with their relations and conditional probabilities, which allow identifying the symptoms with the highest diagnostic potential. They discuss the need to validate the model in the clinical population along with its strengths and limitations.

Complementing previous studies, Arias-Pujol et al. analyze the psychotherapist-patient interaction in psychoanalytic psychotherapy in a single case of a 4-year-old boy with a diagnosis of severe autism spectrum disorder. They used an *ad hoc* observation instrument combining a field format and a category system. They also estimated the results by applying a polar coordinate analysis that allows to obtain an inter-relational map of the connections detected between the established focal behavior and the different categories. Particularly, the result of this study show which therapist behaviors are most useful for promoting social interaction in a child with severe autism.

Regarding the Psychometric contributions, Lee et al. validate the Korean version of the Student-Athletes' Motivation toward Sports and Academics Questionnaire (SAMSAQ) using exploratory structural equation modeling (ESEM). The results of this study illustrated that the SAMSAQ-KR appears to be a robust and reliable instrument.

In line with it, some articles demonstrate the application of new psychometric methods. Ünlü studies self-determination theory (SDT) by applying sets and relations instead of a common numerical approach. The applied technique is an inductive item tree analysis, which is an established method of Boolean analysis of questionnaires. The underlying models were computed within each of the intrinsic, identified, introjected, and external regulations, in autonomous and controlled motivations, and the entire motivation domain. In future studies, the approach of this article could be employed to develop adaptive assessment and training procedures in SDT contexts.

Liang et al. proposed a missing data model for not-reached items in cognitive diagnosis assessments. They simulated the model by estimating model parameters using Bayesian Markov chain Monte Carlo method. The model improved diagnostic feedback results and produced accurate item parameters when the missing data mechanism was non-ignorable. They also study the applicability of the model using a dataset from the 2018 Program for International Student Assessment computer-based mathematics cognitive test.

Cui et al. propose a stochastic approximation expectation maximization (SAEM) algorithm to estimate a model in which Ramsay-curve item response theory is incorporated in a three-parameter normal ogive (RC-3PNO) with non-normal latent trait distributions. They based this proposal considering that in real testing the assumption of the normality of latent traits in the estimation of item response models may be untenable. The simulation studies reveal that the SAEM algorithm produces more accurate item parameters for the RC-3PNO model than for a conventional one (3PNO), especially when the latent density is not normal, as in the cases of a skewed or bimodal distribution. The authors also demonstrate the application of the proposed algorithm using a PISA 2018 test dataset.

Finally, we would like to emphasize that the editors greatly appreciate the contributions received from the authors, which were compiled in this Research Topic.

We also hope it would be of interest to readers and a source of ideas and motivation for the development of future work in Methodology for International Research in Psychology.

Author contributions

MS, FH-T, and MC took the lead in writing the manuscript. SS-C, SC-M, and JL-L provided critical feedback and helped shape content of the manuscript. All authors reviewed the final version of the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

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.

Publisher's note

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 12 April 2022
doi: 10.3389/fpsyg.2022.878331

Aberrant Salience and Disorganized Symptoms as Mediators of Psychosis

*Celia Ceballos-Munuera[1], Cristina Senín-Calderón[2], Sandra Fernández-León[3], Sandra Fuentes-Márquez[4] and Juan Fco. Rodríguez-Testal[1]**

[1] Personality, Evaluation and Psychological Treatment Department, University of Seville, Seville, Spain, [2] Department of Psychology, University of Cádiz, Cádiz, Spain, [3] Penitenciary Psychiatric Hospital, Seville, Spain, [4] Clinical Mental Health Management Unit, Hospital Juan Ramón Jiménez, Huelva, Spain

Introduction: Ideas of reference (IR) are frequent in psychopathology, mainly in psychotic disorders. The frequency of IR and preoccupation about them are related to the psychotic dimension, and to a lesser extent, to negative or emotional disorganized dimensions. Aberrant salience (AS), has been proposed as an indicator of the onset of psychosis, particularly of schizophrenia. This study analyzed the mediating role of AS, disorganized symptoms and preoccupation about IR in the relationship between IR and the psychotic dimension.

Method: The sample consisted of 330 participants (116 university students and 214 clinically active patients), 62.4% of whom were women aged 18–79. The Referential Thinking Scale, the Aberrant Salience Inventory, and the Brief Psychiatric Rating Scale were administered.

Results: Evidence of a partial mediation model showed that the relationships between IR and the psychotic dimension were mediated jointly by AS and the disorganized dimension, and preoccupation about IR no longer had a role. This relationship was significantly influenced by participant age. The variables in the model explained 54.16% of the variance.

Conclusion: The model proposed enabled a set of vulnerabilities (unusual thought content) to be predicted that could lead to a high-risk general pathological state and proneness to psychosis in particular. These findings are discussed with regard to early detection and prevention of psychosis.

Keywords: ideas of reference, preoccupation, psychosis, aberrant salience, psychopathology

OPEN ACCESS

Edited by:
Miguel Ángel Carrasco, National University of Distance Education (UNED), Spain

Reviewed by:
María Roncero, University of Valencia, Spain
Marian Perez-Marin, University of Valencia, Spain
Almudena Carneiro-Barrera, University of Granada, Spain

****Correspondence:***
Juan Fco. Rodríguez-Testal
testal@us.es

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *17 February 2022*
Accepted: *17 March 2022*
Published: *12 April 2022*

Citation:
Ceballos-Munuera C, Senín-Calderón C, Fernández-León S, Fuentes-Márquez S and Rodríguez-Testal JF (2022) Aberrant Salience and Disorganized Symptoms as Mediators of Psychosis. Front. Psychol. 13:878331. doi: 10.3389/fpsyg.2022.878331

INTRODUCTION

A large part of the effort devoted to the study of psychotic disorders has been directed at early identification of psychotic experiences and attenuated psychotic symptoms, thereby establishing rates of transition to psychosis, functioning profiles, and intervention designs (Colizzi et al., 2020; Fusar-Poli et al., 2020; Raballo et al., 2020).

In a longitudinal study, Addington et al. (2015) emphasized the importance of unusual thought content, particularly of suspicion and persecutory ideas. However, unusual thought content also shapes other very common important psychotic experiences, such as ideas of reference and aberrant salience (Miller et al., 2003; Yung et al., 2005).

Ideas of reference (IR) are a type of self-referential processing defined as self-attributions about what happens in the social environment (Senín-Calderón et al., 2017). They are considered one of the subclinical psychotic experiences grouped under the conceptual umbrella of schizotypy, which are relatively frequent in the general population (Lenzenweger et al., 1997; Rodríguez-Testal et al., 2019). Studies have suggested that IR are clinically important because their frequency, stability and severity are related to psychosis (Reininghaus et al., 2016; Marshall et al., 2019; Rodríguez-Testal et al., 2019; Cicero et al., 2020).

Research on IR has tried to establish its relationship with other symptoms and diagnoses, in addition to psychotic disorders, based on differences in their content (observation, communication, guilt or shame, communication media, causality, etc.) or their emotional effects (pleasant or unpleasant) (Wing et al., 1974; Lenzenweger et al., 1997; Startup and Startup, 2005; Cicero and Kerns, 2011; Senín-Calderón et al., 2014). Recently, (Wong et al., 2021) proposed considering IR according to their relationship with events in the setting which could bring them on (e.g., severe social events or a pandemic), between those that could be similarly interpreted by others, and those which are characteristic of people who experience them more strongly. They called the first IR attenuated and the second exclusive, characterizing a continuum of severity up to delusional intensity.

IR has been a basic symptom in proposals for Clinical High Risk assessment (Gross et al., 2008; Schultze-Lutter et al., 2010) as an attenuated psychotic symptom (Yung et al., 2005), as alterations in self-experience (Parnas et al., 2005), on a continuum leading up to delusional activity as delusional ideas of reference (Wong et al., 2012), and with regard to delusional ideas of persecution (Green et al., 2008).

Some IR parameters, such as the distress they cause, preoccupation about them, or specific content can be considered clinically significant (Green et al., 2008; Wong et al., 2012, 2021). Some initial data suggest that preoccupation about the presence of IR differentiates diagnostic categories more clearly, particularly those characterized by psychotic symptomatology (Senín-Calderón et al., 2016). Although some results suggest more importance of IR (Pelizza et al., 2019), there is an outstanding lack of studies on the role of parameters such as preoccupation about them, or the processes involved in these parameters.

According with van Os (2009), aberrant or anomalous salience (AS) has been proposed as a characteristic indicator of onset of psychotic disorders, particularly of schizophrenia. Kapur (2003), suggested that it consists of a motivational and attentional change or alteration toward stimuli from the setting, in such a way that neutral and irrelevant stimuli become abnormally salient. AS would then be the alteration of the natural motivational process toward novelty/reward attributed to the dopaminergic dysregulation traditionally linked by some authors to the onset of psychosis (Kapur et al., 2005; Raballo et al., 2019). From this perspective we might wonder whether the presence of IR, as a social cognition process, could be related to later increase in contextual ambiguity, perplexity, the still uncertain changes in meanings, which identify AS, and were classically called the trema stage (Conrad, 1958). Thus, AS may be a condition mediating the appearance of the first IR and later abnormal significance [the apophenia stage (Conrad, 1958), leading to the crystallization of delusion (Stanghellini et al., 2020)].

Therefore, AS could suggest a level of disorganization in cognitive functions, as observed in studies analyzing unusual thought content (Wilcox et al., 2014; Addington et al., 2015). In fact, Sass et al. (2018) allude to perceptual dysintegration, relating AS and other deficits that affect control and monitorization of cognitive processes. Thus, the emergence of AS could be related to more general disorganization and could be connected with other parameters of IR such as preoccupation about their content. This set of variables would be globally related to positive symptomatology.

Both IR (i.e., self-referencing) and AS can be included among the self-disorders (Mishara et al., 2016), although these cognitive processes may therefore appear at different moments in the development of psychosis. Self-disorders or anomalous subjective experiences (affectation of first-person experience, of its limits, its internal processes, etc.), are considered closely related to schizophrenia and AS, specifically, as a better predictor of psychosis than attenuated psychotic symptoms (Feyaerts et al., 2021). The IR would be an example of non-specific self-disorders included among the self-disorders as a trait (of schizotypy) (Lenzenweger et al., 1997), while processes like AS are a state related to schizophrenia (Blain et al., 2020).

Identification of the mediators in any model proposed is essential to research. Mediation analysis enables researchers to go beyond mere relationships between variables by studying the possible effects of these relationships (MacKinnon et al., 2020). Although it is clear that a causal relationship between the variables cannot really be established, it does suggest a possible direction in the relationship between clinical variables and health (MacKinnon and Luecken, 2008). Another reason for studying mediators is the individual differences in the relationships between biological, psychological, behavioral and social factors. Mediation analyses can thereby suggest the development of efficient and effective intervention programs based on identification and later focus on the critical components of the program (O'Rourke and MacKinnon, 2018). In this sense and from a dimensional approach (van Os and Reininghaus, 2016), this study attempts to establish a series of relationships (not causal) among different indicators possibly related to the onset of psychosis. Thus, we propose a relationship between experiencing IR and the psychotic dimension of symptoms as a whole, and a tentative multiple mediation model to analyze this relationship. The hypothesis posed was that AS experiences, the disorganized dimension and preoccupation about IR sequentially mediate the relationship between the IR, as frequency of unusual thought content, and the

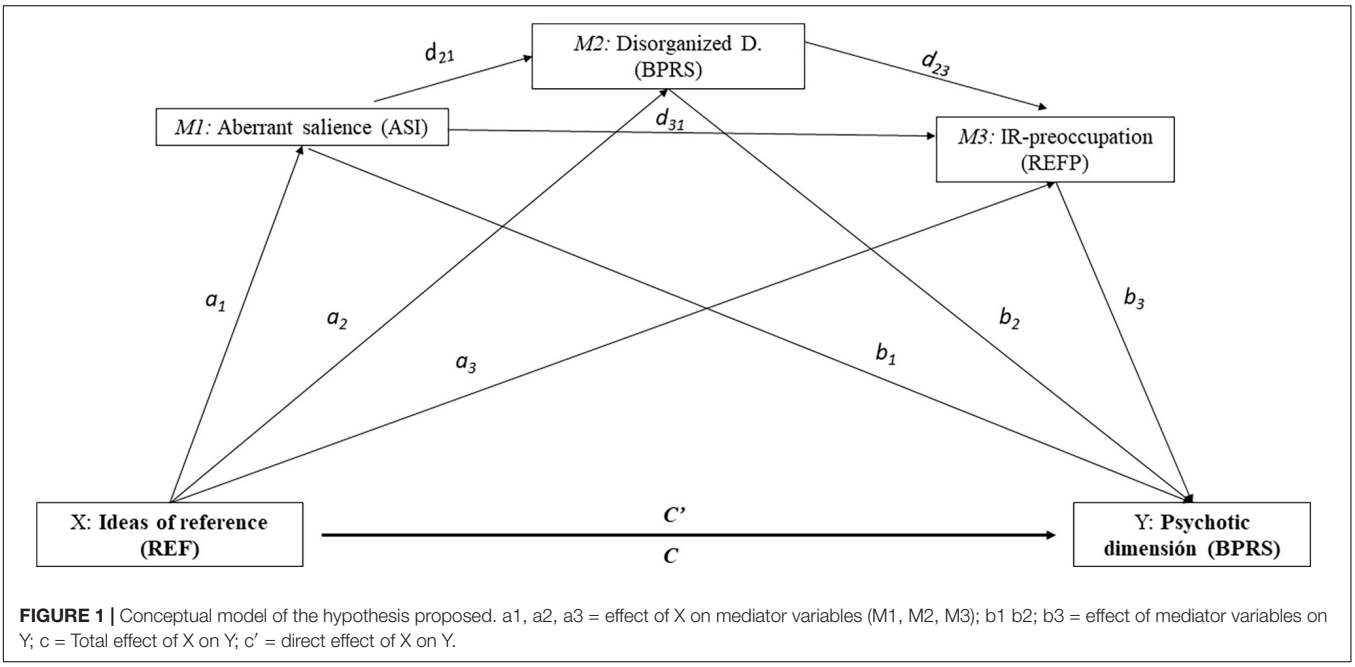


FIGURE 1 | Conceptual model of the hypothesis proposed. a1, a2, a3 = effect of X on mediator variables (M1, M2, M3); b1 b2; b3 = effect of mediator variables on Y; c = Total effect of X on Y; c′ = direct effect of X on Y.

appearance of a set of positive manifestations. The study's hypotheses are organized in the theoretical model shown in **Figure 1**.

MATERIALS AND METHODS

Participants

The study sample was made up of 330 participants (116 university students and 214 clinically active patients), aged 18–79, (*M* = 29.53; *SD* = 12.72), 62.4% of whom were women. The majority of participants were single (77.6%), 16.4% married or with a partner, 1.5% widowed, and 4.5% separated/divorced. In the group of students, 53.3% were women, the average age was 20.69 (*SD* = 3.45), and the average Social Class Index (Hollingshead, 1975) was 38.83 points (*SD* = 20.30; middle class; range 11–77); 4.3% were taking medication at the time of evaluation (anxiolytics) and 13.8% had a psychopathological history. In the group of patients, 46.7% were women, the average age was 34.32 (*SD* = 13.33), and the average Social Class Index was 46.41 points (*SD* = 22.40; low class; range 11–77). In this group, 67.8% were on medication at the time of the first interview, 77.5% had a psychopathological history and 35% had more than one diagnosis. **Table 1** describes the sociodemographic characteristics of the participants and clinical diagnoses of the patients. The diagnoses were made by healthcare professionals (the patients' clinical psychologists) with long clinical experience following the DSM-IV-TR classification (American Psychiatric Association [APA], 2000).

Instruments

Referential Thinking Scale (REF, (Lenzenweger et al., 1997). Validated Spanish version by Rodríguez-Testal et al. (2019). It evaluates the interpretation of casual situations where the person may think others are staring at, laughing at or taking notice of him or her. It consists of 34 items with a "true/false" answer format. Preoccupation about IRs is scored by adding a Likert-type scale from 0 to 5 to each item, where 0 shows not preoccupied about IR at all, and 5 very much preoccupied (Senín-Calderón et al., 2016). This test has various factors (laughter/comments, attention/appearance, guilt/shame; songs/communication media, and reactions/changes), but in this study, only the sum of its items in a global REF score was used, and for preoccupation (REFP) when IR was found to be present. The Spanish validation found adequate

TABLE 1 | Sample sociodemographic characteristics and clinical diagnoses.

	Students (*n* = 116)	Patients (*n* = 214)
Sex (% women)	53.3%	46.7%
Age (*M, SD*)	20.69 (3.45)	34.32 (13.33)
Age range (years)	19–47	18–79
Diagnoses (*n*)	–	Depressive D. = 42 Adjustment D. = 10 Anxiety D. = 52 Schizophrenia and other Psychotic D. = 65 Bipolar D. = 13 Personality D.[1] = 19 Other[2] = 13

[1]*Personality disorders (main diagnosis): Schizoid = 1; Paranoid = 3; Schizotypy = 5; Histrionic = 1; Borderline = 5; Dependence = 1; Unspecified personality D. = 3.* [2]*Other: Dissociative = 1; Somatoform D. = 3; Eating behavior D. = 1; Sexual D. = 1; Addictive D. = 1; Neurodevelopment D. = 2.*

psychometric properties with ordinal α for the total score that varied from 0.94 (adolescents) to 0.97 (patients and adult general population). Internal consistency for this study had a Cronbach's α = 0.88 for IR frequency, and 0.92 for preoccupation.

*Aberrant Salience Inventory (*ASI, (Cicero et al., 2010), validated in Spanish by Fernández-León et al. (2019). This scale was used to measure proneness to psychosis and evaluate the assignment of significance to usually irrelevant internal and external stimuli. It has 29 items with a dichotomous (true/false) answer format with five dimensions: Heightened cognition, impending understanding, heightened emotionality, increased significance, senses sharpening. The total score is found by adding up the items answered affirmatively. The authors reported a Cronbach's a = 0.89 for the complete scale and evidence of validity compared to other measures of proneness to psychosis. The Spanish validation had adequate internal consistency (Cronbach's α = 0.95) and evidence of validity compared to other measures of positive and negative symptoms. In this study, the Cronbach's a = 0.90.

Brief Psychiatric Rating Scale (BPRS; Lukoff et al., 1986). Spanish validation by Peralta Martín and Cuesta Zorita (1994). This is an other-report scale with 24 items that evaluate positive (psychotic and disorganized) and negative emotional symptoms. The items are rated from 1 (absence of psychopathology) to 7 (extreme psychopathology). The 18-item Spanish version had a Cronbach's a reliability of 0.59 to 0.70, and test-retest reliability r = 0.70. This study used only the psychotic dimension (Items 6. Suspiciousness, 7. Delusions, 8. Grandiosity, 9. Hallucinations) and the disorganized dimension (Items 11. Conceptual disorganization, 12. Excitement, 21. Bizarre behavior, 22. Elation, 23. Hyperactivity, 24. Distractedness). Internal consistency in this study was Cronbach's a = 0.85 for the psychotic dimension and Cronbach's a = 0.82 for the disorganized dimension.

Procedure

All participants were informed of the objectives of the study and gave their written consent for participation. The sample was collected by incidental sampling. The participants from the general population were university students who voluntarily agreed to fill in the study tests in class, and later were evaluated in an individual interview. These students received points toward their grade for participating at both evaluation times. The group of patients was collected by incidental sampling at a private psychology clinic and various public hospitals. They filled in the tests in writing, and at another time, were individually interviewed. This study followed the precepts of the Declaration of Helsinki and was approved by the Ethics Committee of the Junta de Andalucía [Andalusian Government] (PI 010/16).

Data Analysis

All data analyses were performed with SPSS.19 software. The theoretical mediation model proposed was tested following the recommendations of MacKinnon et al. (2004), using multiple mediation analysis with three mediators (AS, disorganized dimension and preoccupation about IR) using the PROCESS macro v3.5 (Hayes, 2017; Model 6) with bootstrapping. The coefficients were estimated from 10000 bootstrap samples and all analyses were accepted with a level of significance of $p < 0.05$.

RESULTS

Descriptive Analyses

Before applying the mediation model, the relationships between the variables to be entered were tested. The results, in **Table 2**, show significant positive association between all the variables in the study. As observed in **Table 2**, differences were found between the ages of patients and non-patients [$t_{(261.626)} = 14.10$; $p < 0.001$], so this variable was entered in the mediation model as a covariate.

Mediation Analysis

According to Baron and Kenny (1986), four conditions must be met to establish a theoretical mediation model: The IV or predictor, must be related to the DV or criterion; the predictor must be related to the mediator variables; these mediators have to be related to the criterion variable once the effect of the IV is controlled for. Finally, the effect of the predictor variable on the criterion variable must be reduced when the effect of the mediator variables is controlled for. The results of the multiple mediation analysis are shown in **Table 3**.

The presence of IR had a significant total effect on the psychotic dimension. When AS experiences, disorganization and

TABLE 2 | Means, standard deviations, and Spearman correlations between study variables.

	1.	2.	3.	4.	5.	6.	*M*	*D.T*
1. REF	–	0.472**	0.563**	0.368**	0.876**	0.08**	5.466	5.456
2. BPRS- Psy		–	0.545**	0.712**	0.408**	0.434**	1.860	1.199
3. ASI			–	0.508**	0.535**	0.157**	10.888	6.767
4. BPRS-Dis				–	0.285**	0.482**	1.432	0.544
5. REFP					–	-0.049*	10.350	16.033
6. Age						–	29.53	12.72

*N = 330; 1 = IR; 2 = Psychotic dimension; 3 = Aberrant salience; 4 = Disorganized dimension; 5 = Preoccupation about IR. *p < 0.05; **p < 0.01. (bilateral).*

TABLE 3 | Total, direct and indirect effects based on 10,000 bootstrap samples.

	B	*SE*	*t*	*p*	CI 95% (upper and lower)
Total effect (X→Y)	0.087	0.009	9.902	0.001	[0.070, 0.104]
Direct (X→Y) effect	0.027	0.013	2.055	0.041	[0.001, 0.053]
Age (covariate)	0.009	0.004	2.226	0.027	[0.001, 0.016]
Indirect effects					
−1: X→M1→Y	0.027	0.008			[0.015, 0.044]
−2: X→M1→M2→Y	0.016	0.005			[0.008, 0.026]
−3: X→M1→M3→Y	0.001	0.001			[−0.001, 0.001] (ns)
−4: X→M1→M2→M3→Y	0.001	0.001			[−0.001, 0.000] (*ns*)
−5: X→M2→Y	0.015	0.006			[0.005, 0.028]
−6: X→M2→M3→Y	0.001	0.001			[−0.001, 0.000] (ns)
−7: X→M3→Y	0.002	0.013			[−0.027, 0.026] (ns)

X = IR; Y = psychotic D; M1 = SA; M2 = Disorganized D.; M3 = Preoccupation about IR; ns = not statistically significant (as shown by a bootstrap confidence interval including 0); CI = Confidence interval. The coefficients are unstandardized.

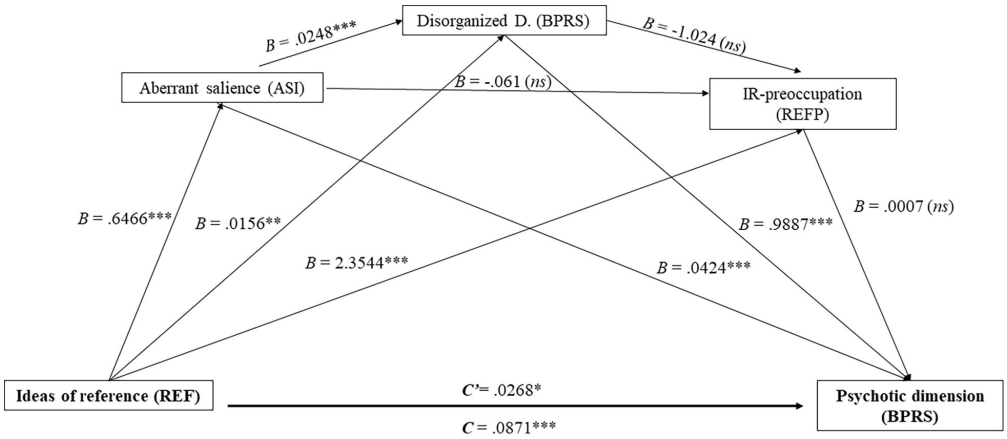


FIGURE 2 | Mediation model proposed and tested; c′ = direct effect of X; c = total effect of X on Y; on Y; $^{**}p < 0.01$; $^{***}p < 0.001$; *ns* = not statistically significant. Coefficients shown are unstandardized.

preoccupation about IR were added to the model, the size of the beta coefficient decreased, although still remaining significant. The total indirect effect was statistically significant ($B = 0.060$, IC 95% [0.031,0.090]). The results suggest that AS, disorganization and preoccupation about IR, partially mediated the relationship between presence of IR and the psychotic dimension. As shown in **Table 3**, the indirect effects of AS and disorganization on the psychotic dimension were significant, in interaction as well as alone. Age was significant ($B = 0.009$, IC 95% [0.001,0.016]) in the relationship between IR and the psychotic dimension, effect which was controlled for. The overall model represented 54.16% of the variance in the total score of the psychotic dimension, $R^2 = 0.54$, F (5, 324) = 76.56, $p < 0.001$, with a large effect size (Cohen's $f^2 = 1.18$). The final model is shown in **Figure 2**.

DISCUSSION

In spite of the strong presence of IR in human nature itself, and of being a type of mental activity historically characterized in both evolutionary development and the clinical context, there are few studies directed specifically at this cognitive process, and its mention has been relegated mainly to the delusional form (Senín-Calderón et al., 2014).

Similarly, AS, traditionally related to psychosis because it represents the emergence of alterations of the self (Henriksen and Parnas, 2019), can also be objectivized on the continuum of human experience (Fyfe et al., 2008), possibly because of its relationship with other processes as common as depersonalization or absorption (Sass et al., 2018; Fernández-León et al., 2020).

This study attempted to explore these two unusual thought contents, IR and AS, by examining their possible dependence, connection with disorganization processes and finally, their possible relationship with the psychotic dimension. In the mediation model proposed, the hypothesis is that AS experiences, product of the frequency of IR, lead to an increase in disorganization indicators, which in turn, could be associated with greater preoccupation about the presence of IR, and finally, with more symptoms characteristic of the psychotic dimension. Evidence was found for a mediation model, according to which

the relationships between IR and psychotic dimension symptoms were both mediated by AS and the disorganized dimension, and preoccupation about IR no longer had a role.

The findings of this study are consistent with others that have found a statistically significant positive relationship between the referential thinking scale (REF) and the ASI (Chun et al., 2019; Rodríguez-Testal et al., 2019), as well as the CAPE scale of proneness to psychosis (Fonseca-Pedrero et al., 2012), so it may be said that both tests are clearly related to psychosis, although the REF scale seems to be somewhat more specifically so, at least in the general population (Senín-Calderón et al., 2014). When these scales were used to evaluate IR and AS, it was demonstrated that the latter must participate in development of IR, although they are characterized in a wider context of psychotic disorganization (Cicero et al., 2010). It was also demonstrated that the assessment of preoccupation about IR is significantly related to measures of psychotic symptomatology, and also with emotional components that indicate clinically significant distress (Senín-Calderón et al., 2016).

The tradition of the continuum of unusual thought content, such as IR, joins the German tradition of Robert Gaupp (Mishara and Fusar-Poli, 2013) and Ernest Kretschmer (Rodríguez-Testal et al., 2016), who described aspects related to personality, which would place sensitivity (asthenia) at the base of sensitive delusion of reference, precipitated by life events (Kretschmer and Strauss, 1952). The excellent recent study by Wong et al. (2021) suggests that, although the IR are precipitated by a context, the exclusiveness with which they are experienced must be considered. These authors emphasize the mediating role of rumination between events experienced and the severity of the IR. Rumination about the event probably has this role. In brief, IR can occur during general cognitive functioning, but requires the mediation of rumination to become exclusive to psychosis. In this study, the mediating role of preoccupation, but specifically about the IR themselves, was not verified, although it was significantly correlated with all the measures. It is therefore possible that its role is rather of moderator, as proposed in other studies, although with negative affect as the predictor (Ludwig et al., 2020).

The reconsideration of alterations of the self as diagnostic criteria for schizophrenia (ICD-11) (World Health Organization [Who]., 2018) attributes extraordinary importance to AS. However, some indications suggest that this indicator of disorganization of the self is not exclusive to schizophrenia (Neumann et al., 2020), although it may be more prominent, continuous, even a trait of this disorder (Blain et al., 2020). Possibly, as suggested by Feyaerts et al. (2021), most specificity of these alterations of the self occurs when processes such as hyperreflexivity (exaggerated self-consciousness) and diminished self-presence (diminished intensity or vitality of oneself) emerge.

This study did not intend to analyze the different diagnostic categories or verify whether the large group of participants diagnosed with active schizophrenia were characterized to a greater extent by AS. Therefore, one limitation of this study is precisely that it assumed a continuum perspective, including from participants from the general population to different diagnoses, with and without psychotic symptoms. However, we think that this perspective enables the processes in its relationship to be understood without considering diagnostic differences, since the interview took place at a different time from when the tests were filled out. Another limitation to be mentioned is its cross-sectional design, and incidental recruitment of the sample, so results cannot be generalized. However, the sequence of the mediation model could make it possible to explore the relationship of relevant variables and an approach to causality (Hayes and Preacher, 2014), although understood as strictly provisional. Some very important variables for discussing psychosis are also lacking, from negative symptoms to consideration of high negative affect and low positive affect, or facets related to upbringing and stressful situations (Olvet et al., 2015; Fernández-León et al., 2020; Healy et al., 2020).

Summarizing, a model and relationships between variables has been proposed in a sequence that could enable an approach to psychosis that takes unusual thought content as a starting point. This proposal enables vulnerabilities to be explored, which due to their onset and stability in process severity, may be useful indicators in the early detection of the psychosis for prevention or early intervention. The results followed the perspective that considers schizotypy a latent variable of personality organization, represented by a subtle, subclinical psychotic phenomenology (e.g., perceptive aberrations or ideas of reference) which could lead to a variety of phenotypical results related to schizophrenia (Lenzenweger, 2018). Unlike Meyer et al. (2021), these characteristics of schizotypy (trait, such as ideas of reference) could be the antecedent of state processes, such as aberrant salience. According to these authors, it is possible for impaired self-processing to appear. Current studies have attempted to design a more meticulous evaluation of the high-risk groups, forming more homogeneous subgroups stratified by variables (Carrión et al., 2017). In view of the mediator role found in this study, some expressions of self-disorders, particularly aberrant salience, would make a more detailed analysis of the psychosis prodromes possible. This study also supports the need for design and validation of reliable instruments for identification and detection of referential thought (ideas of reference) in clinical and general populations, a project on which we are now working on. This proposal would enable exploration of vulnerabilities, which due to their onset and stability in the severity of the process, may be useful as indicators to be kept in mind for early detection of the psychotic process for prevention or early intervention.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Junta de Andalucía (Andalusian 156 Government) (PI 010/16). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

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

FUNDING

This work has been funded by the research group "Mental Alteration and Social Dysfunction" (CTS301), of the University of Seville.

REFERENCES

Addington, J., Liu, L., Buchy, L., Cadenhead, K. S., Cannon, T. D., Cornblatt, B. A., et al. (2015). North American prodrome longitudinal study (NAPLS 2): the prodromal symptoms. *J. Nerv. Ment. Dis.* 203, 328–335. doi: 10.1097/NMD.0000000000000290

American Psychiatric Association [APA] (2000). *Diagnostic and Statistical Manual of Mental Disorders, Text Revision. DSM-IV-TR*. Washington DC: APA.

Baron, R. M., and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51:1173. doi: 10.1037/0022-3514.51.6.1173

Blain, S. D., Longenecker, J. M., Grazioplene, R. G., Klimes-Dougan, B., and DeYoung, C. G. (2020). Apophenia as the disposition to false positives: a unifying framework for openness and psychoticism. *J. Abnorm. Psychol.* 129, 279–292. doi: 10.1037/abn0000504

Carrión, R. E., Correll, C. U., Auther, A. M., and Cornblatt, B. A. (2017). A severity-based clinical staging model for the psychosis prodrome: longitudinal findings from the New York recognition and prevention program. *Schizophr. Bull.* 43, 64–74. doi: 10.1093/schbul/sbw155

Chun, C. A., Brugger, P., and Kwapil, T. R. (2019). Aberrant salience across levels of processing in positive and negative schizotypy. *Front. Psychol.* 10:2073. doi: 10.3389/fpsyg.2019.02073

Cicero, D. C., and Kerns, J. G. (2011). Unpleasant and pleasant referential thinking: relations with self-processing, paranoia, and other schizotypal traits. *J. Res. Personal.* 45, 208–218. doi: 10.1016/j.jrp.2011.02.002

Cicero, D. C., Gawęda, Ł, and Nelson, B. (2020). The placement of anomalous self-experiences within schizotypal personality in a nonclinical sample. *Schizophr. Res.* 218, 219–225. doi: 10.1016/j.schres.2019.12.043

Cicero, D. C., Kerns, J. G., and McCarthy, D. M. (2010). The aberrant salience inventory: a new measure of psychosis proneness. *Psychol. Assess.* 22, 688–701. doi: 10.1037/a0019913

Colizzi, M., Lasalvia, A., and Ruggeri, M. (2020). Prevention and early intervention in youth mental health: is it time for a multidisciplinary and trans-diagnostic model for care? *Int. J. Ment. Health Syst.* 14:23. doi: 10.1186/s13033-020-00356-9

Conrad, K. (1958). La Esquizofrenia incipiente. *Madr. Fund. Arch. Neurobiol.* (in press).

Fernández-León, S., Rodríguez-Testal, J. F., Gutiérrez-López, M. L., and Senín-Calderón, C. (2020). Interpersonal violence and psychotic-like experiences: the mediation of ideas of reference, childhood memories, and dissociation. *Int. J. Environ. Res. Public. Health* 17:4587. doi: 10.3390/ijerph17124587

Fernández-León, S., Senín-Calderón, C., Gutiérrez-López, M. L., and Rodríguez-Testal, J. F. (2019). Spanish validation of the aberrant salience inventory in a general adolescent population. *Psicothema* 31, 210–217.

Feyaerts, J., Henriksen, M. G., Vanheule, S., Myin-Germeys, I., and Sass, L. A. (2021). Delusions beyond beliefs: a critical overview of diagnostic, aetiological, and therapeutic schizophrenia research from a clinical-phenomenological perspective. *Lancet Psychiatry* 8, 237–249. doi: 10.1016/S2215-0366(20)30460-0

Fonseca-Pedrero, E., Paino, M., Lemos-Giráldez, S., and Muñiz, J. (2012). Validación de la Escala para la evaluación comunitaria de las experiencias Psíquicas-42 (CAPE-42) en universitarios y pacientes con psicosis. [Validation of the community assessment psychic experiences-42 (CAPE-42) in Spanish college students and patients with psychosis.]. *Actas Esp. Psiquiatr.* 40, 169–176.

Fusar-Poli, P., Salazar de Pablo, G., Correll, C. U., Meyer-Lindenberg, A., Millan, M. J., Borgwardt, S., et al. (2020). Prevention of psychosis: advances in detection, prognosis, and intervention. *JAMA Psychiatry* 77:755. doi: 10.1001/jamapsychiatry.2019.4779

Fyfe, S., Williams, C., Mason, O., and Pickup, G. (2008). Apophenia, theory of mind and schizotypy: perceiving meaning and intentionality in randomness. *Cortex* 44, 1316–1325. doi: 10.1016/j.cortex.2007.07.009

Green, C. E. L., Freeman, D., Kuipers, E., Bebbington, P., Fowler, D., Dunn, G., et al. (2008). Measuring ideas of persecution and social reference: the Green et al . Paranoid Thought Scales (GPTS). *Psychol. Med.* 38, 101–111. doi: 10.1017/S0033291707001638

Gross, G., Huber, G., Klosterkötter, J., and Linz, M. (2008). *Bonn Scale for the Assessment of Basic Symptoms*. Aachen: Shaker Verlag.

Hayes, A. F. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis, Second Edition: A Regression-Based Approach*. New York, NY: Guilford Publications.

Hayes, A. F., and Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *Br. J. Math. Stat. Psychol.* 67, 451–470. doi: 10.1111/bmsp.12028

Healy, C., Coughlan, H., Clarke, M., Kelleher, I., and Cannon, M. (2020). What mediates the longitudinal relationship between psychotic experiences and psychopathology? *J. Abnorm. Psychol.* 129, 505–516. doi: 10.1037/abn0000523

Henriksen, M. G., and Parnas, J. (2019). "Delusional Mood," in *The Oxford Handbook of Phenomenological Psychopathology*, eds G. Stanghellini, M. Broome, A. Raballo, A. V. Fernandez, P. Fusar-Poli, and R. Rosfort (Oxford: Oxford University Press), 742–752. doi: 10.1093/oxfordhb/9780198803157.013.72

Hollingshead, A. B. (1975). *Five Factor Index of Social Position*. New Haven, CT: Yale University. Unpublished manuscript.

Kapur, S. (2003). Psychosis as a state of Aberrant Salience: a framework linking biology, phenomenology, and pharmacology in Schizophrenia. *Am. J. Psychiatry* 160, 13–23. doi: 10.1176/appi.ajp.160.1.13

Kapur, S., Mizrahi, R., and Li, M. (2005). From dopamine to salience to psychosis—linking biology, pharmacology and phenomenology of psychosis. *Schizophr. Res.* 79, 59–68. doi: 10.1016/j.schres.2005.01.003

Kretschmer, E., and Strauss, E. B. (1952). *Medizinische Psychologie. A Text-book of Medical Psychology. Translated. with an Introduction*, 2 Edn, ed. E. B. Strauss (London: Hogarth Press).

Lenzenweger, M. F. (2018). Schizotypy, schizotypic psychopathology and schizophrenia. *World Psychiatry* 17, 25–26. doi: 10.1002/wps.20479

Lenzenweger, M. F., Bennett, M. E., and Lilenfeld, L. R. (1997). The referential thinking scale as a measure of schizotypy: scale development and initial construct validation. *Psychol. Assess.* 9, 452–463. doi: 10.1037/1040-3590.9.4.452

Ludwig, L., Mehl, S., Schlier, B., Krkovic, K., and Lincoln, T. M. (2020). Awareness and rumination moderate the affective pathway to paranoia in daily life. *Schizophr. Res.* 216, 161–167. doi: 10.1016/j.schres.2019.12.007

Lukoff, D., Liberman, R. P., and Nuechterlein, K. H. (1986). Symptom monitoring in the rehabilitation of schizophrenic patients. *Schizophr. Bull.* 12, 578–603. doi: 10.1093/schbul/12.4.578

MacKinnon, D. P., and Luecken, L. J. (2008). How and for whom? Mediation and moderation in health psychology. *Health Psychol.* 27, S99–S100. doi: 10.1037/0278-6133.27.2(Suppl.).S99

MacKinnon, D. P., Lockwood, C. M., and Williams, J. (2004). Confidence Limits for the Indirect Effect: distribution of the Product and Resampling Methods. *Multivar. Behav. Res.* 39, 99–128. doi: 10.1207/s15327906mbr3901_4

MacKinnon, D. P., Valente, M. J., and Gonzalez, O. (2020). The correspondence between causal and traditional mediation analysis: the link is the mediator by treatment interaction. *Prev. Sci.* 21, 147–157. doi: 10.1007/s11121-019-01076-4

Marshall, C., Lu, Y., Lyngberg, K., Deighton, S., Cadenhead, K. S., Cannon, T. D., et al. (2019). Changes in symptom content from a clinical high-risk state to conversion to psychosis. *Early Interv. Psychiatry* 13, 257–263. doi: 10.1111/eip.12473

Meyer, M. S., Cohn, J. R., Hayashi, K., and Cicero, D. C. (2021). Momentary assessment of aberrant salience, anomalous self-experiences, and psychotic-like experiences. *Personal. Disord. Theory Res. Treat.* (in press). doi: 10.1037/per0000530

Miller, T. J., McGlashan, T. H., Rosen, J. L., Cadenhead, K., Ventura, J., McFarlane, W., et al. (2003). Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. *Schizophr. Bull.* 29, 703–715. doi: 10.1093/oxfordjournals.schbul.a007040

Mishara, A. L., and Fusar-Poli, P. (2013). The phenomenology and neurobiology of delusion formation during psychosis onset: jaspers, truman symptoms, and Aberrant Salience. *Schizophr. Bull.* 39, 278–286. doi: 10.1093/schbul/sbs155

Mishara, A., Bonoldi, I., Allen, P., Rutigliano, G., Perez, J., Fusar-Poli, P., et al. (2016). Neurobiological models of self-disorders in early schizophrenia. *Schizophr. Bull.* 42, 874–880. doi: 10.1093/schbul/sbv123

Neumann, S. R., Glue, P., and Linscott, R. J. (2020). Aberrant salience and reward processing: a comparison of measures in schizophrenia and anxiety. *Psychol. Med.* 51, 1507–1515. doi: 10.1017/S0033291720000264

O'Rourke, H. P., and MacKinnon, D. P. (2018). Reasons for testing mediation in the absence of an intervention effect: a research imperative in prevention and intervention research. *J. Stud. Alcohol Drugs* 79, 171–181. doi: 10.15288/jsad.2018.79.171

Olvet, D. M., Carrión, R. E., Auther, A. M., and Cornblatt, B. A. (2015). Self-awareness of functional impairment in individuals at clinical high-risk for psychosis: subjective impairment in CHR. *Early Interv. Psychiatry* 9, 100–107. doi: 10.1111/eip.12086

Parnas, J., Møller, P., Kircher, T., Thalbitzer, J., Jansson, L., Handest, P., et al. (2005). EASE: examination of anomalous self-experience. *Psychopathology* 38:236. doi: 10.1159/000088441

Pelizza, L., Poletti, M., Azzali, S., Paterlini, F., Garlassi, S., Scazza, I., et al. (2019). Anhedonia in adolescents at ultra-high risk (UHR) of psychosis: findings from a 1-year longitudinal study. *Eur. Arch. Psychiatry Clin. Neurosci.* 270, 337–350. doi: 10.1007/s00406-019-01018-9

Peralta Martín, V., and Cuesta Zorita, M. (1994). [Validation of positive and negative symptom scale (PANSS) in a sample of Spanish schizophrenic patients]. *Actas Luso Esp. Neurol. Psiquiatr. Cienc. Afines* 22, 171–177.

Raballo, A., Cicero, D. C., Kerns, J. G., Sanna, S., Pintus, M., Agartz, I., et al. (2019). Tracking salience in young people: a psychometric field test of the Aberrant Salience Inventory (ASI): RABALLO et al. *Early Interv. Psychiatry* 13, 64–72. doi: 10.1111/eip.12449

Raballo, A., Poletti, M., Preti, A., and McGorry, P. (2020). Clinical high risk for psychosis in children and adolescents: a meta-analysis of transition prevalences. *Schizophr. Res.* (in press). doi: 10.1016/j.schres.2020.03.063

Reininghaus, U., Kempton, M. J., Valmaggia, L., Craig, T. K. J., Garety, P., Onyejiaka, A., et al. (2016). Stress sensitivity, Aberrant Salience, and threat anticipation in early psychosis: an experience sampling study. *Schizophr. Bull.* 42, 712–722. doi: 10.1093/schbul/sbv190

Rodríguez-Testal, J. F., Bendala-Rodríguez, P., Perona-Garcelán, S., and Senín-Calderón, C. (2019). Examining the structure of ideas of reference in clinical and community samples. *Compr. Psychiatry* 93, 48–55. doi: 10.1016/j.comppsych.2019.06.006

Rodríguez-Testal, J. F., Perona-Garcelán, S., Senín-Calderón, C., García-Jiménez, M. M., Álvarez-garcía, P., and Núñez-Gaitán, M. C. (2016). Kretschmer revisitado: fatiga mental e ideas de referencia. Contribuciones desde la Teoría de la Sensibilidad al Refuerzo. *Rev. Psicopatol. Psicol. Clín.* 21, 25–43. doi: 10.5944/rppc.vol.21.num.1.2016.16372

Sass, L., Borda, J. P., Madeira, L., Pienkos, E., and Nelson, B. (2018). Varieties of self disorder: a bio-pheno-social model of schizophrenia. *Schizophr. Bull.* 44, 720–727. doi: 10.1093/schbul/sby001

Schultze-Lutter, F., Ruhrmann, S., Berning, J., Maier, W., and Klosterkotter, J. (2010). Basic symptoms and ultrahigh risk criteria: symptom development in the initial prodromal state. *Schizophr. Bull.* 36, 182–191. doi: 10.1093/schbul/sbn072

Senín-Calderón, C., Perona-Garcelán, S., Fuentes-Márquez, S., and Rodríguez-Testal, J. F. (2017). A mediation model for ideas of reference: the role of the gray model, self-consciousness, and emotional symptoms. *Psychol. Rep.* 120, 443–459. doi: 10.1177/0033294117693593

Senín-Calderón, C., Rodríguez-Testal, J. F., and Perona-Garcelán, S. (2016). Las ideas de referencia y la preocupación por su presencia: estudio sobre su relevancia para la caracterización de las psicosis. *Rev. Iberoam. Psicol. Salud* 7, 1–8. doi: 10.1016/j.rips.2015.10.004

Senín-Calderón, C., Rodríguez-Testal, J. F., and Perona-Garcelán, S. (2014). *El Pensamiento Referencial: Aspectos Psicopatológicos y del Desarrollo*. Charleston: CreateSpace Independent Publishing Platform.

Stanghellini, G., Palumbo, D., Ballerini, M., Mucci, A., Catapano, F., Giordano, G. M., et al. (2020). Abnormal bodily experiences detected by Abnormal Bodily phenomena questionnaire are more frequent and severe in schizophrenia than in bipolar disorder with psychotic features. *Eur. Psychiatry* 63:e49. doi: 10.1192/j.eurpsy.2020.49

Startup, M., and Startup, S. (2005). On two kinds of delusion of reference. *Psychiatry Res.* 137, 87–92. doi: 10.1016/j.psychres.2005.07.007

van Os, J. (2009). 'Salience syndrome' replaces 'schizophrenia' in DSM-V and ICD-11: psychiatry's evidence-based entry into the 21st century? *Acta Psychiatr. Scand.* 120, 363–372. doi: 10.1111/j.1600-0447.2009.01456.x

van Os, J., and Reininghaus, U. (2016). Psychosis as a transdiagnostic and extended phenotype in the general population. *World Psychiatry* 15, 118–124. doi: 10.1002/wps.20310

Wilcox, J., Briones, D., Quadri, S., and Tsuang, M. (2014). Prognostic implications of paranoia and thought disorder in new onset psychosis. *Compr. Psychiatry* 55, 813–817. doi: 10.1016/j.comppsych.2013.12.010

Wing, J., Cooper, J., and Sartorius, N. (1974). *The Description and Classification of Psychiatric Symptoms: An Instruction Manual for the PSE and CATEGO System*. Cambridge: Cambridge University Press.

Wong, G., Hui, C. L. M., Tang, J. Y. M., Chiu, C. P. Y., Lam, M. M. L., Chan, S. K. W., et al. (2012). Screening and assessing ideas and delusions of reference using a semi-structured interview scale: a validation study of the Ideas of Reference Interview Scale (IRIS) in early psychosis patients. *Schizophr. Res.* 135, 158–163. doi: 10.1016/j.schres.2011.12.006

Wong, S., Hui, C. L. M., Wong, C. S. M., Suen, Y. N., Chan, S. K. W., Lee, E. H. M., et al. (2021). Induced ideas of reference during social unrest and pandemic in Hong Kong. *Schizophr. Res.* 229, 46–52. doi: 10.1016/j.schres.2021.01.027

World Health Organization [Who]. (2018). *The ICD-11 Classification of Mental and Behavioural Disorders*. Geneva: WHO.

Yung, A. R., Yung, A. R., Pan Yuen, H., Mcgorry, P. D., Phillips, L. J., Kelly, D., et al. (2005). Mapping the onset of psychosis: the comprehensive assessment of at-risk mental states. *Aust. N. Z. J. Psychiatry* 39, 964–971. doi: 10.1080/j.1440-1614.2005.01714.x

Conflict of Interest: 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 | Frontiers in Psychology

ORIGINAL RESEARCH
published: 21 April 2022
doi: 10.3389/fpsyg.2022.853236

Validation of the Student Athletes' Motivation Toward Sports and Academics Questionnaire (SAMSAQ) for Korean College Student-Athletes: An Application of Exploratory Structural Equation Modeling

*Youngjik Lee[1], Jason Immekus[2], Dayoun Lim[3], Mary Hums[4], Chris Greenwell[4], Adam Cocco[4] and Minuk Kang[1]**

[1] Department of Physical Education, Kookmin University, Seoul, South Korea, [2] Department of Educational Leadership, Evaluation and Organizational Development, University of Louisville, Louisville, KY, United States, [3] Department of Sports and Leisure Studies, Far East University, Eumseong-gun, South Korea, [4] Department of Health & Sport Sciences, University of Louisville, Louisville, KY, United States

OPEN ACCESS

Edited by:
Salvador Chacón-Moscoso, Seville University, Spain

Reviewed by:
Sai-fu Fung, City University of Hong Kong, Hong Kong SAR, China
Eric Hall, Elon University, United States

***Correspondence:**
Minuk Kang
minuk@kookmin.ac.kr

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: 12 January 2022
Accepted: 16 March 2022
Published: 21 April 2022

Citation:
Lee Y, Immekus J, Lim D, Hums M, Greenwell C, Cocco A and Kang M (2022) Validation of the Student Athletes' Motivation Toward Sports and Academics Questionnaire (SAMSAQ) for Korean College Student-Athletes: An Application of Exploratory Structural Equation Modeling. Front. Psychol. 13:853236. doi: 10.3389/fpsyg.2022.853236

The purpose of this study was to validate the Korean version of the Student-Athletes' Motivation toward Sports and Academics Questionnaire (SAMSAQ) using exploratory structural equation modeling (ESEM). A total of 412 (men 77%; women 23%) South Korean collegiate student-athletes competing in 27 types of sports from 13 different public and private universities across South Korea were analyzed for this study. ESEM statistical approach was employed to examine the psychometric properties of SAMSAQ-KR. To assess content validity, the SAMSAQ-KR was inspected by a panel of content subject experts. The Athletic Identity Measurement Scale was used to obtain convergent validity. The results of this study illustrated that the SAMSAQ-KR appears to be a robust and reliable instrument.

Keywords: college student-athletes, motivations, validation, ESEM, international psychology

INTRODUCTION

College student-athletes typically find it difficult to balance their responsibilities between undertaking higher education classes and performing at a competitive level in their respective sports. In general, college student-athletes spend around 20–30 h on the activities that are related to their sport (e.g., training, competitive matches, strength and conditioning, and so on) to maintain their level of competitiveness (Guidotti et al., 2013). This dedication toward athletic performance may hinder the academic achievement of college student-athletes. In addition, it typically affects the student-athletes' life negatively, which include sustaining injuries, limited social interactions, and limited post-athletic career opportunities (e.g., struggling to find a job outside of sports) (Lopes Dos Santos et al., 2020). Thus, there is a growing concern and interest across various countries regarding the educational development of the dual career of college student-athletes (Gaston-Gayles, 2005; Lupo et al., 2015; Quinaud et al., 2020). For example, many European countries (e.g., Belgium, Germany) provide academic services (e.g., tutoring and time management

workshops) to support their college student-athletes' academic achievement (Aquilina and Henry, 2013). In the United States (US), the National Collegiate Athletic Association (NCAA) has started providing numerous forms of academic support, such as appointing academic advisors and tutors and implementing study hall programs, to engage student-athletes in their academic roles (Huml et al., 2014). Additionally, the NCAA has also implemented many legislative acts to enhance their student-athletes' academic development (e.g., limiting in-season practice to 20 h per week) (Huml, 2018).

The lack of focus on education for college student-athletes has become one of the major problems in South Korean sports society as well. In South Korean school sports culture, there is a great emphasis on student-athletes' athletic achievement (Lee and Yu, 2021). The roots of this athletic culture can be found in South Korea's history. For example, South Korea has historically used sports to bolster the country's image, placing tremendous pressure on their athletes to achieve Olympic glory (Lee, 2006; Lim and Huh, 2009; Heo, 2011). Due to this culture, South Korean student-athletes are required to devote most of their time to training for their sport. In other words, they are required to sacrifice proper academic opportunities for athletic achievement (Lee and Yu, 2021). In addition, the emphasis on athletic performance over academic achievement for student-athletes in South Korea has perpetuated a culture where student-athletes can openly neglect their academic responsibilities. For example, under the South Korean college sports system, high school student-athletes are able to become college student-athletes without proper academic preparation and credentials (e.g., high school grade point average (GPA) and college entrance test), as South Korean colleges only consider athletic achievements for admissions (Kim, 2011).

Although South Korean student-athletes devote considerable time and energy to their sports, only a handful of South Korean student-athletes actually go on to become professional athletes due to unpredictable variables, such as injuries and family circumstances (Ham, 2003; Huml et al., 2014; Kim et al., 2014). Furthermore, the competitive life span of a South Korean professional athlete is short, with the average age of retired professional athletes at only 23.8 years (Yoo, 2016). In other words, most South Korean college student-athletes will need to integrate back into society as a typical member of South Korean society early in their adulthood without the title of athlete that they lived with their whole life (Otto et al., 2019). However, because these individuals spend more time developing athletic skills than developing academic skills, upon retirement, these South Korean student-athletes struggle to integrate successfully back into society (Lee and Kwon, 2013).

To improve this existing situation that many student-athletes are facing, the Korea University Sports Federation (KUSF) was established as a governing and advisory body for college sports in South Korea in 2010. A total of 110 public and private institutions across the nation are currently members of the KUSF, which is approximately 25% of all the South Korean institutions. The KUSF operates and manages the U-league that offers a total of six sports, including baseball, basketball, ice hockey, soccer, soft tennis, and volleyball. In 2017, in an attempt to mitigate the gap between college student-athletes' athletic and academic accomplishments, the KUSF implemented a minimum GPA requirement for their member institutions. Based on the requirement, student-athletes who earned a GPA of less than 2.0 in the previous semester were not eligible to participate in the next semester's U-league.

Student-Athletes' Motivation Toward Sports and Academics

It is important to monitor student-athletes' motivation toward academics and athletics, since motivation can be viewed as being central to their participation in the respective activities (Pestana et al., 2018). To examine the academic and athletic motivation of the college student-athletes in the US, Gaston-Gayles developed and validated the Student-Athletes' Motivation toward Sports and Academics Questionnaire (SAMSAQ), a 30-item scale to examine college student-athletes' motivation toward sports and academics (Gaston-Gayles, 2005). The instrument was developed within the framework of the expectancy-value theory (Weiner, 1986), the self-efficacy theory (Bandura, 1977), and the attribution theory (Weiner, 1986), which posits that a students' motivation to perform their academic and athletic tasks. The multidimensional instrument includes three subscales, student-athletic motivation (SAM; eight items), career athletic motivation (CAM; five items), and academic motivation (AM; 16 items), with responses recorded on a six-point Likert scale (1 = Very strongly disagree to 6 = Very strongly agree). The mean score of all the items included in the respective subscales provides the final score for the subscale, and the SAMSAQ's factor analytic results have supported the scale's theoretical three-factor structure. To determine internal validity in particular, Gaston-Gayles conducted exploratory factor analysis (EFA) with a sample size of 153 (Gaston-Gayles, 2005). Initially, a total of 30 items were considered for the analysis, but three items were eliminated due to low factor loadings, low item-to-total correlations, and low reliability.

The SAMSAQ has been validated in different contexts since college student-athletes' imbalance between academics and athletics is one of the growing concerns of international sports societies. Societal contexts where the SAMSAQ has been validated include Italy (Guidotti et al., 2013), the United Arab Emirates (Fortes et al., 2010), Portugal (Quinaud et al., 2020), and South Korea (Park et al., 2015). Previous reports of the inconsistent factor structures of the SAMSAQ suggest that the instrument may be sensitive to socio-cultural contexts (Corrado et al., 2012; Guidotti et al., 2013; Lupo et al., 2015). The validation of the scale in South Korea allowed for a better understanding of the South Korean college student-athletes' motivation toward academics and athletics. However, there remain two key issues in the 2015 version that need to be addressed. First, a relatively small number of participants from a single sport were used for the analysis. Second, confirmatory factor analysis (CFA) was employed as the main analysis for the study. However, due to low factor loadings, multiple items were deleted from the original instrument.

It is essential to utilize both EFA and CFA when identifying a new factor structure for a new version of the SAMSAQ, as it takes into consideration the socio-cultural contexts in which the study was conducted. Therefore, most of the previous SAMSAQ validation studies employed both EFA and CFA for their analysis (Fortes et al., 2010; Guidotti et al., 2013; Lupo et al., 2015; Quinaud et al., 2020). However, traditional factor analyses (i.e., CFA and EFA) also have their own shortcomings.

Exploratory Structural Equation Modeling

The factor analytic procedures provide a powerful, flexible approach to examine the internal structure of the SAMSAQ (Brown, 2015). In general, EFA and CFA are the most popular approaches to investigate the factor structure of psychological instruments. EFA is a data-driven approach to determining the number of empirical factors underlying a set of scale items, while CFA represents a model-based approach to testing an instrument's factor structure based on *a priori* information to specifying the relationship between observed (e.g., items) and latent (e.g., motivation) variables. For instance, even though an EFA model may be empirically supported, it could still report unacceptable model–data fit with CFA. In addition, CFA is typically associated with problems related to the goodness of fit, measurement invariance across groups, differential item functioning, and differentiation of factors (Marsh et al., 2009). CFA's overly restrictive independent cluster model (ICM), requiring items to load on only one factor and constraining non-target factor loadings to zero, typically results in an unacceptable model–data fit (Marsh et al., 2011). Furthermore, requiring non-target loadings to be zero can result in higher than expected factor correlations (Asparouhov and Muthén, 2009; Marsh et al., 2010). Thus, these problems and issues have led researchers to question the appropriateness of ICM-CFA models in applied research, including psychological instruments.

Exploratory structural equation modeling is a single-procedure factor analytic approach that combines the features and advantages of EFA and CFA to test an instrument's multidimensionality (Asparouhov and Muthén, 2009). Specifically, ESEM determines the number of factors according to factor interpretability, model–data fit statistics (e.g., comparative fit index (CFI)), and inspection of model parameters (e.g., factor loadings). In terms of a confirmatory approach, ESEM utilizes a target rotation where target loadings are unconstrained factor loadings, whereas non-target loadings are cross-loadings specified close to zero (Marsh et al., 2014). In summary, ESEM yields an interpretable factor analysis that provides a model-based approach, capturing key components of EFA by allowing cross-loadings that fully identify where the items fall, as well as the measures of the model–data fit like CFA (Morin et al., 2020). Therefore, ESEM can be viewed as a robust and flexible model-based factor analytic approach to examine a scale's multidimensional structure with its exploratory and confirmatory approaches (Asparouhov and Muthén, 2009; Marsh et al., 2010; Morin et al., 2020).

Beyond testing an instrument's internal structure, ESEM also allows researchers to examine the group differences in the underlying latent factor means and differential item functioning (DIF) detection by using an ESEM-based factor within a multiple-indicator multiple-cause approach (MIMIC) modeling (Jöreskog and Goldberger, 1975; Muthén, 1989; Holland and Wainer, 1993; Marsh et al., 2014). In terms of DIF, it is identified as the item's statistical properties (e.g., difficulty and discrimination) that differ between the groups (e.g., demographics) based on equal standing on the latent trait (e.g., student-athletes' motivations). Unlike multiple-group CFA, the MIMIC model investigates the relationship between the background of participants (e.g., gender) and the factors of ESEM and the value of DIF among item intercepts (Marsh et al., 2014). Within the sports and physical education research, the MIMIC model may be suitable for the analyses of multiple groups to test the measurement invariance, as well as using typically small sample sizes from applied research (Vandenberg and Lance, 2000).

Figure 1 shows a hypothetical MIMIC model where the latent factor is regressed onto an observed predictor identified by the solid lines, whereas the dotted line indicates the direct effect of the observed predictor on the indicator (e.g., item). In terms of predictor, it could be a categorical (e.g., gender) or continuous variable (e.g., age), with the regression coefficient representing the direction and strength of its relationship with the latent factor. In **Figure 1**, the solid line indicates the regression coefficient, which provides differences in the latent mean score if the predictor variable indicates gender. In terms of the dotted line, it illustrates that an item represents DIF. The MIMIC model identifies the detection of intercept differences, followed by predicting equality of factor loadings, factor variances, and co-variances across groups. It is important to note that a test of intercept differences needs to be conducted before comparing factor mean scores of the groups because intercept variance is considered as a premise for inspecting latent mean differences.

Purpose of the Study

With its advantages, there is a growing interest in using ESEM to examine an instrument's multidimensionality in various fields such as education and medical research field (Dicke et al., 2018; Karlgren et al., 2020; Pommier et al., 2020; Sancho et al., 2020; Neff et al., 2021). In the sports psychology literature as a whole, however, few studies have utilized ESEM to examine the psychometric properties of the instrument as a model-based analysis. Therefore, the purpose of this study was to validate the Korean version of the SAMSAQ using ESEM analysis. Specifically, this study will demonstrate the use of ESEM to investigate the psychometric properties of SAMSAQ based on Marsh et al.'s (2014) ESEM using the following guidelines: First, based on Gaston-Gayles (2005) three-factor model, both ICM-CFA and ESEM model analyses will be conducted to examine which model provides better model–data fit and investigate the comparability of model parameters such as factor loadings. Second, if both models show unacceptable model–data fit, a series of ESEM models will be analyzed using an exploratory approach, which includes various factors to investigate the

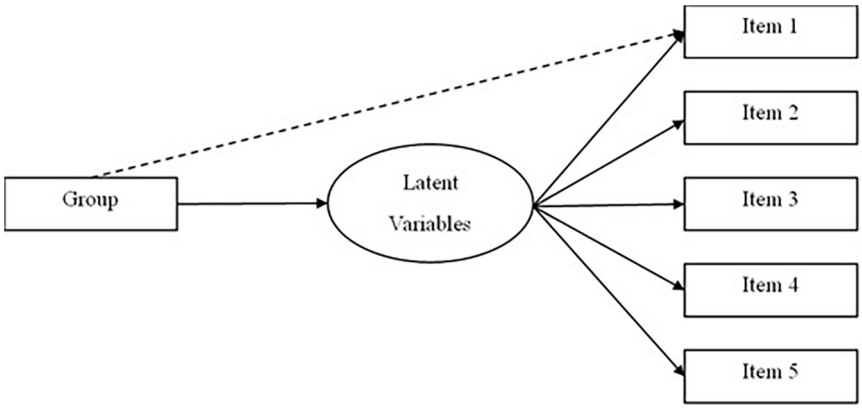


FIGURE 1 | Hypothetical multiple-indicator multiple-cause approach (MIMIC) model.

relationship between ESEM factors and all items. Among a series of ESEM models, a model that provides the best model–data fit and appropriate factor solution will be selected as a preferred model. Subsequently, based on the acceptable model–data fit and factor structure, the MIMC model analysis will be employed to investigate latent mean differences across the student-athletes' motivation dimensions and the presence of DIF among SAMSAQ items across gender. After examining the psychometric properties of the Korean version of the SAMSAQ, this study will also examine and provide current South Korean college student-athletes' motivation toward sports and academics across gender. Furthermore, this study contributes to the international sports psychology research field by providing rigorous guidance on how to conduct international research, including (a) how to collect research data from other countries, (b) how to translate an existing scale to other languages systemically, and (c) how to establish the content validity of the translated scale while considering that other countries have different cultures.

MATERIALS AND METHODS

Participants

A total of 435 participants initially responded to the survey. However, 23 responses were removed due to incomplete data. Thus, the remaining 412 responses (men 77%; women 23%) from South Korean collegiate student-athletes competing in 27 types of sports from 13 different public and private universities across South Korea were analyzed for this study. Team sports (i.e., canoe, field hockey, handball, rowing, soccer, and water polo) student-athletes comprised 19% of the sample, whereas student-athletes from individual sports represented 81% of the sample (i.e., archery, badminton, bobsleigh, bowling, boxing, cycling, danceSport, fencing, golf, gymnastics, ice skating, rifle, ski, swimming, taekwondo, tennis, track and field, triathlon, weightlifting, and wrestling). Gender imbalance in South Korean college student-athletes affected the participant composition of this study as well. In particular, in 2020, there were a total of 13,048 college student-athletes in South Korea and 10,299 (80%) were men, whereas 2,749 (20%) were women (Korean Sport and Olympic Committee, 2020).

Instrumentation

Translation Process

First, two bilingual translators (i.e., two Korean doctoral students who are majoring in a sport-related major enrolled in American universities) translated the instrument from English to Korean. Second, for back translation, two bilingual translators (i.e., two Korean faculty members in a sport-related department at two Korean universities who earned their graduate degrees in the United States) who have not been exposed to the instrument were asked to translate the Korean instrument to English. Third, field testing (e.g., interviews with current South Korean college student-athletes) was conducted to ensure that participants could comprehend the questions. Finally, fine-tuning of the instrument included extensive discussions with all the translators if potential discrepancies or problems were identified during the field testing.

Student-Athletes' Motivation Toward Sports and Academics

The SAMSAQ is a 30-item scale to examine the United States college student-athletes' motivation toward sports and academics (Gaston-Gayles, 2005). The instrument used a total of three different theories: expectancy-value, self-efficacy, and attribution. The instrument consists of three subscales: student-athletic motivation (e.g., It is important for me to do better than other athletes in my sport), career athletic motivation (e.g., My goal is to make it to the professional level or the Olympics in my sport), and academic motivation (e.g., The content of most of my courses is interesting to me), with responses provided on a six-point rating scale (*1 = Very strongly disagree* to *6 = Very strongly agree*). The SAMSAQ has reported acceptable levels of internal consistency ($\alpha > 0.80$) (Taber, 2018).

Athletic Identity Measurement Scale

The purpose of the Athletic Identity Measurement Scale (AIMS) is to assess participants' athletic identity (Brewer and Cornelius, 2001). AIMS will be used to examine convergent validity with the SAMSAQ. AIMS is a seven-item questionnaire with a seven-point rating scale, (*1* = *Strongly disagree* to *7* = *Strongly agree*). Participants' athletic identity is measured by a total composite score generated by a sum of the scores for the seven items. Higher AIMS scores indicate stronger athlete identification. AIMS has high internal consistency (α = 0.81–0.93) (Taber, 2018). Systemic translation strategies (Sousa and Rojjanasrirat, 2011) were applied to AIMS as well to achieve equivalence in meaning, which required a range of scale translation.

Procedures

After receiving approval from the University Institutional Review Board (IRB), the researchers developed online versions of the instrument and delivered the instrument to participants through Qualtrics. To reach out to the South Korean student-athletes, the researchers contacted the schools for approval to administer the survey to the student-athletes and to explain the purpose of the study. Upon receiving approval from the schools, the investigator sent a survey link to the coaches of the South Korean college sports teams. Coaches provided the survey link to student-athletes during their team meetings. Student-athletes who wanted to participate in the survey could access the survey on their own devices (e.g., cell phone, laptop, tablet, and so on). The survey contained a total of three sections, (a) demographics, (b) the SAMSAQ, and (c) the AIMS scale, and took approximately 35 min to complete. Prior to taking the survey, all the participants were informed about the research purpose and the option to discontinue the survey at any time. The main data collection phase occurred between August and December 2019.

Data Analysis

Content Validity

Content-related validity was gathered by a panel of experts comprising a former college student-athlete, a Korean professional athlete, a current college student-athlete, and two Korean faculty members with expertise in student-athlete research reviewing the translated version of the SAMSAQ. The panel of experts was provided a translated version of the SAMSAQ to provide feedback regarding item clarity, relevance, and comprehension. Through the content validity process, Item 10 (I chose/will choose my major because it is something I am interested in as a career) was removed since the South Korean college student-athletes are required to major in only sports-related fields, such as physical education, exercise science, and sport management.

Statistical Analysis

Descriptive statistics were used to inspect the students' average ratings and variability of responses. The factor structure of SAMSAQ-KR was guided by the results of Gaston-Gayles (2005) SAMSAQ-US. First, individual ICM-CFA and ESEM models based on Gaston-Gayles (2005) three-factor model were tested. For the ICM-CFA model, a simple model was used where items were loaded on their assigned motivation factors freely, and non-target loadings were constrained to zero. This involved fixing the factor variances to one (Kline, 2013). Parameter estimation was examined based on robust maximum likelihood (MLR) using Mplus 8.4 (Muthén and Muthén, 2005). MLR can be viewed as a suitable analysis to examine fit indices with non-normal, ordered categorical data and to estimate standard errors. Model–data fit was based on the chi-square (χ^2), the Comparative Fit Index (CFI), the Root Mean Squared Error of Approximation (RMSEA), and the Standardized Root Mean Square Residual (SRMR) index. Acceptable model–data fit criteria included CFI values greater than 0.95, RMSEA less than 0.06, and SRMR less than 0.08 (Hu and Bentler, 1999). Also, Akaike information criterion (AIC) and Bayesian information criterion (BIC) values were examined, with lower values indicating better model–data fit (Schwarz, 1978; Akaike, 1987). Coefficient omega (ω) was examined for factor analytic based estimate of reliability (McDonald, 1999).

TABLE 1 | Mean and standard deviation values for SAMSAQ items (*N* = 412).

Item	Mean	SD	Item	Mean	SD
1	4.53	1.33	17	2.98	1.37
2	5.03	1.15	18	4.11	1.33
3	4.58	1.03	19	4.02	1.34
4	4.70	1.15	20	4.73	1.35
5	4.73	1.21	21	3.41	1.29
6	3.46	1.26	22	4.17	1.33
7	4.36	1.07	23	5.12	0.98
8	4.66	1.17	24	4.73	1.08
9	3.73	1.47	25	4.44	1.27
11	3.73	1.46	26	3.49	1.23
12	4.37	1.04	27	4.80	1.08
13	4.87	1.10	28	4.19	1.10
14	2.31	1.14	29	4.15	1.22
15	1.89	1.03	30	4.42	1.35
16	2.94	1.30			

Convergent Validity

Convergent validity is established when scores on a measure correlate with scores on another measure evaluating a similar construct (Carlson and Herdman, 2012). In this study, the Pearson product-moment correlation was employed to assess the convergent validity of the SAMSAQ-KR by comparing the correlation scores with the AIMS scale. Specifically, the correlation between the scores of athletic-related motivation factors from SAMSAQ-KR and AIMS was investigated.

RESULTS

The mean values of items varied considerably with values ranging from 1.89 (Item 15: "It is worth the effort to be an exceptional athlete in my sport") to 5.12 (Item 23: "I am confident that I can earn a college degree"), whereas standard deviation (SD) ranged from 0.98 (Item 23: "I am confident that I can earn a college

TABLE 2 | Model–data fit of exploratory structural equation modeling (ESEM) models.

Model	χ^2	*df*	CFI	RMSEA	(90% CIs)	SRMR
3-Factor	748.149*	322	0.90	0.057	(0.051–0.055)	0.04
4-Factor	558.758*	296	0.94	0.046	(0.040–0.052)	0.03
5-Factor	456.787*	271	0.96	0.041	(0.034–0.047)	0.03

$p < 0.01$; df: Degrees of freedom. CIs: Confidence intervals.

degree") to 1.47 (Item 9: "I have some doubt about my ability to be a star athlete on my team"). **Table 1** presents the descriptive statistics of the SAMSAQ-KR data for all the 412 participants.

A test based on Gaston-Gayles's three-factor ICM-CFA model (Gaston-Gayles, 2005) indicated unacceptable model–data fit, where $\chi^2(249) = 1{,}568.708$, $p < 0.01$; RMSEA = 0.116 (90% CI: 0.110–0.112); CFI = 0.64; and SRMR = 0.04. Furthermore, a three-factor ESEM model also reported unacceptable model–data fit, where $\chi^2(322) = 748.149$, $p < 0.01$; RMSEA = 0.057 (90% CI: 0.051–0.055); CFI = 0.90; and SRMR = 0.162. Consequently, a series of ESEM model analyses were conducted to identify the number of empirical factors. As reported in **Table 2**, the results supported a five-factor solution, where χ^2 (271) = 456.787, $p < 0.01$, RMSEA = 0.041 (90% CI: 0.034–0.047), CFI = 0.96, and SRMR = 0.03.

The EFA factor loading (see **Table 3**) results showed that the items predominantly loaded onto a specific factor. Five items (e.g., Item 12 and Item 25) did not load on any factor, and one item indicated substantive cross-loadings on two factors (i.e., Item 9). Also, Item 10 (I chose/will choose my major because it is something I am interested in as a career) was eliminated since South Korean college student-athletes are mandated to major in only sports-related fields. Therefore, a total of 23 items from the original version of SAMSAQ were retained and loaded above 0.40 onto one of the five factors.

Based on the results of factor loading and acceptable model–data fit, five factors were investigated with a review of all items significantly loading on each factor. Factor 1 was identified as academic achievement motivation and consisted of three items ($\omega = 0.70$) that indicated participants' academic performance, such as GPA, grades, and graduation. Factor 2 was labeled athletic motivation and included six items ($\omega = 0.79$) that assessed students' motivation toward their sport. Factor 3 was labeled learning outcome motivation and comprised three items ($\omega = 0.77$) that dealt with students' learning experiences through their classes and application of that knowledge and skills outside the school upon their graduation. Factor 4 was labeled as career athletic motivation and consisted of three items ($\omega = 0.85$) that included the items representing the participants' desire to compete at the professional and Olympic levels. Factor 5 was labeled as academic motivation and included a total of seven items ($\omega = 0.75$) that indicated participants' motivation toward their academic-related tasks.

Table 4 reports the correlation among the ESEM-based factors that ranged from negligible, –0.02 (e.g., Factor 4 and Factor 5) to moderate, 0.50 (e.g., Factor 2 and Factor 4).

TABLE 3 | Item pattern coefficients for final exploratory structural equation modeling (ESEM) solution.

	Factors				
Item	**1**	**2**	**3**	**4**	**5**
1	**0.95**	– 0.17	– 0.02	– 0.01	– 0.07
2	0.17	**0.47**	– 0.00	0.34	0.00
3	0.27	0.01	**0.61**	0.01	0.09
4	**0.43**	0.00	0.23	– 0.10	0.06
5	0.09	**0.49**	– 0.00	0.25	– 0.00
6	– 0.02	– 0.19	0.15	– 0.11	**0.49**
7	0.17	0.01	**0.56**	0.00	– 0.04
8	0.10	0.39	0.23	0.06	– 0.02
9	– 0.16	– 0.04	0.04	**0.55**	**0.52**
11	0.15	– 0.07	0.05	– 0.14	**0.49**
12	0.11	0.05	0.00	– 0.07	0.00
13	0.18	**0.52**	0.01	0.29	0.00
14	– 0.00	**0.57**	0.17	0.29	0.02
15	0.14	**0.65**	0.01	0.16	0.08
16	– 0.10	– 0.22	0.17	0.04	**0.69**
17	0.00	– 0.33	0.37	– 0.15	**0.47**
18	– 0.09	0.35	– 0.05	0.29	– 0.15
19	0.07	– 0.07	– 0.02	**0.83**	– 0.14
20	0.03	0.33	– 0.01	**0.52**	– 0.00
21	0.29	– 0.31	0.03	0.14	**0.48**
22	0.09	0.07	0.03	**0.78**	– 0.04
23	**0.41**	0.31	0.12	0.03	0.12
24	0.19	0.39	0.26	0.15	0.05
25	0.05	0.34	0.02	0.07	– 0.21
26	0.03	– 0.01	0.04	– 0.10	**0.64**
27	0.13	**0.57**	– 0.01	0.34	0.02
28	0.05	– 0.02	**0.65**	0.18	– 0.02
29	0.12	0.21	– 0.03	– 0.18	– 0.34
30	0.12	0.06	0.18	– 0.16	**0.68**

Bolded values indicate loadings above 0.40. Factor 1: Academic Achievement Motivation; Factor 2: Athletic Motivation; Factor 3: Learning Outcome Motivation; Factor 4: Career Athletic Motivation; Factor 5: Academic Motivation.

TABLE 4 | Exploratory structural equation modeling (ESEM) factor correlations.

	Factors				
	1	**2**	**3**	**4**	**5**
1	1				
2	0.24	1			
3	0.43*	0.07	1		
4	0.13	0.50*	0.02	1	
5	0.16	– 0.00*	0.05*	– 0.02*	1

$p < 0.05$. Factor 1: Academic Performance Motivation; Factor 2: Athletic Motivation; Factor 3: Learning Outcome Motivation; Factor 4: Career Athletic Motivation; Factor 5: Academic Motivation.

The results of MIMIC modeling reported that the inclusion of direct effects from gender to the items would not improve the model–data fit. Furthermore, the regression coefficients from gender to each of the SAMSAQ-KR factors were non-significant.

TABLE 5 | Descriptive statistics and Pearson product-moment correlations between the athletic motivation and career athletic motivation (SAMSAQ-KR) and Athletic Identity Measurement Scale (AIMS).

Variables	Mean	SD	1	2	3
1. AIMS	4.82	0.88	1.00		
2. AM	4.04	0.45	0.62**	1.00	
3. CAM	4.31	1.17	0.59**	0.64**	1.00

** $p < 0.01$.

TABLE 6 | Mean and standard deviation values for motivation scores by gender.

	Academic Achievement Motivation	Athletic Motivation	Learning Outcome Motivation	Career Motivation	Academic Motivation
Male	4.6 ± 1.0	4.3 ± 1.3	3.9 ± 0.9	4.3 ± 1.2	3.6 ± 1.2
Female	4.9 ± 0.9	3.7 ± 1.3	4.5 ± 0.8	4.0 ± 1.2	4.0 ± 1.1

Table 5 reports convergent validity results based on the correlation between SAMSAQ-KR's athletic-related motivation factor (i.e., athletic motivation and career athletic motivation) scores and AIMS. The results indicated that the correlation between athletic-related motivation factors scores and AIMS was moderate (0.59–0.64).

DISCUSSION AND CONCLUSION

The present study aimed to validate the Korean version of the SAMSAQ using ESEM analysis. Compared to the previous studies, which examined South Korean college student-athletes' motivation toward one sport with a relatively small group of participants, the current study used a larger sample size (n = 412) from 27 different sports types for the analysis. Additionally, ESEM is seldom used in sports psychology research. This study estimated South Korean college student-athletes' motivation toward academics and athletics, accounting for gender. To ensure content validity, a panel of experts (i.e., both former and current college student-athletes, a current professional athlete, and Korean faculty members with expertise in student-athlete research) reviewed the translated version of the SAMSAQ. In addition, convergent validity was obtained by comparing correlation scores with the AIMS scale. In conclusion, the SAMSAQ-KR proved to be a robust scale with good psychometric properties.

Overall, South Korean student-athletes recognized the importance of academics. This change may be attributed to the recent effort of the government of South Korea to help their college student-athletes academically. For example, the KUSF was established as a governing body of South Korean college student-athletes. The main aim of the KUSF is to facilitate student-athletes' academic enhancement. For example, the KUSF recently made and implemented the minimum required GPA for student-athletes, supported many colleges by investing in resources for designing academic advisor programs, and conducted seminars for student-athletes regarding academic importance and future career. With the establishment and academic policy implementation of the KUSF, South Korean college student-athletes have many opportunities to recognize the importance of balancing their athletic and academic activities.

Compared to the previous cross-cultural studies that examined student-athletes' motivation toward athletics and academics used SAMSAQ, the results of this study demonstrated that the SAMSAQ-KR version was characterized by a different factor structure compared to the versions followed by other countries. For example, most previous studies sustained the three-factor structure by utilizing the CFA (Fortes et al., 2010; Guidotti et al., 2013; Lupo et al., 2015; Quinaud et al., 2021). However, the results of this study illustrated that SAMSAQ-KR is composed of a total of five factors: athletic motivation, academic motivation, career athletic motivation, academic achievement motivation, and learning outcome motivation. The results also confirmed that each country has unique socio-cultural contexts and different relationships between the academic and athletic environments in college sport (Fortes et al., 2010; Lupo et al., 2017; Quinaud et al., 2021).

Among all the factors, academic achievement motivation and learning outcome motivation were conceptually equivalent to academic motivation factor in Gaston-Gayles (2005)'s original version of SAMSAQ, since these new factors are closely related to academic-related motivation, such as earning a high GPA (Item 1) and focusing on class content (Item 28). In particular, in SAMSAQ-KR, the academic achievement motivation factor was affected by the recent academic environments of both South Korean high school and college student-athletes. In 2013, the ministry of education in South Korea made and implemented the new rules to enhance the academic performance of South Korean student-athletes. One of the main rules was the minimum required GPA, which mandates high school student-athletes to meet the required GPA to play their sports. In addition, in 2017 (as mentioned earlier), the KUSF was established to manage and support South Korean college student-athletes' academic performance. KUSF's initial new rules included minimum required GPA as well. Thus, South Korean college student-athletes were able to recognize the importance of not only their academic performance but also their academic achievement during their stay in both high school and college.

Similar to the previous studies that investigated the college student-athletes' motivation by gender (Gaston-Gayles, 2005; Sherry and Zeller, 2014), the present study illustrated that South Korean female college student-athletes showed higher motivation toward all academic-related factors compared to the male student-athletes (see **Table 6**). This observation might be due to South Korea's male-dominated sports-related career environment. In 2020, for example, there were a total of 7,512 professional athletes in South Korea and 4,834 (64%) were male athletes, whereas 2,678 (36%) were female athletes (Korean Sport and Olympic Committee, 2020). In addition, in terms of the coaching positions, the number of coaches from all levels of sports, including both non-professional (e.g., school

sports) and professional levels, in South Korea was 20,593. Of them, the number of male coaches was 16,938 (82%), whereas the number of female coaches was only 3,665 (18%) (Korean Sport and Olympic Committee, 2020). In other words, male college-graduated athletes have more opportunities to retain their sports career after their graduation in the sports-related field, such as professional athletes and coaches. However, female athletes would need to look for future careers that are not related to sports due to a lack of resources and opportunities for college-graduated female athletes. Thus, female college student-athletes might be more focused on academics rather than athletics to earn career opportunities outside of sports after their graduation.

The findings of this study could provide practical implications for the administrators of South Korean college athletics. For example, as discussed earlier, the KUSF has been trying to support and enhance the academic performance of South Korean college student-athletes. However, there is no suitable scale for examining student-athletes' academics and athletic motivation. Thus, the administrators of South Korean college athletics will be able to use the SAMSAQ-KR scale to analyze and evaluate South Korean college student-athletes' motivation toward athletics and academics. This will allow administrators to evaluate the impact of the academic policies on the student-athletes by examining their motivation toward academics and athletics. In addition, the faculty members might also use this scale to develop their class materials. For example, the results of this study illustrated that the participants are highly oriented toward learning outcome motivation, which is related to student-athletes' learning experiences through their classes and application of the knowledge and skills outside of school premises upon their graduation. Therefore, the faculty members should recognize/understand student-athletes' desire to acquire practical knowledge and skills through the classes and try to make an effort to develop their classes by reflecting student-athletes' learning outcome motivation. Furthermore, the present study confirmed that South Korean female college student-athletes lay greater emphasis on academics than their male counterparts due to reduced opportunities in sports-related careers after their graduation. Therefore, a different approach, depending on gender, will be needed at the time of making new academic policies for South Korean college student-athletes.

Despite the contributions, the current study also has some limitations that need to be considered. Even though the number of college student-athletes analyzed in this study met the criteria for psychometric analysis, it would be better if future studies include more student-athletes from other sports for the analysis. In addition, although considering the imbalance in gender representation in South Korea's college sports, that male student-athletes were 10,001 (79%) out of 12,695, the sample of the current study was comprised of mostly male student-athletes. Thus, further research should be conducted with more female college student-athlete for a better interpretation of SAMSAQ-KR.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because additional studies are undergoing using the dataset. Requests to access the datasets should be directed to YL, youngjik.lee@kookmin.ac.kr.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Louisville Human Subjects Research Institutional Review Board. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YL, MK, and JI conceptualized the study. YL, JI, CG, and AC analyzed the data. YL and DL drafted the manuscript. MH reviewed and edited the manuscript. All authors were involved in designing the study and procedures, read and agreed to the published version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.853236/full#supplementary-material

REFERENCES

Akaike, H. (1987). "Factor analysis and AIC," in *Selected Paper of Hirotugu Akaike*, eds E. Parzen, K. Tanabe, and G. Kitagawa (Berlin: Springer), 317–332. doi: 10.1007/BF02294359

Aquilina, D., and Henry, I. (2013). "Promoting Student–Athlete Interests in European Elite Sport Systems," in *Routledge Handbook of Sport Policy*, eds I. Henry and L. M. Ko (London, United Kingdom: Routledge), 251–266. doi: 10.4324/9780203807217

Asparouhov, T., and Muthén, B. (2009). Exploratory structural equation modeling. *Struct. Equ. Model.* 16, 397–438. doi: 10.1080/10705510903008204

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychol. Rev.* 84, 191–215. doi: 10.1037//0033-295x.84.2.191

Brewer, B. W., and Cornelius, A. E. (2001). Norms and factorial invariance of the athletic identity measurement scale. *Acad. Athl. J.* 15, 103–113.

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. New York, USA: The Guilford Press.

Carlson, K. D., and Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organ. Res. Methods* 15, 17–32. doi: 10.1177/1094428110392383

Corrado, L. A., Tessitore, L., Rauter, S., and Topic, M. D. (2012). Motivation for a dual-career: italian and Slovenian student-athletes. *Kinesiol. Slov.* 18, 47–56.

Dicke, T., Marsh, H. W., Riley, P., Parker, P. D., Guo, J., and Horwood, M. (2018). Validating the Copenhagen Psychosocial Questionnaire (COPSOQ-II) using set-ESEM: identifying psychosocial risk factors in a sample of school principals. *Front. psychol.* 9:584. doi: 10.3389/fpsyg.2018.00584

Fortes, P. C., Rodrigues, G., and Tchantchane, A. (2010). Investigation of academic and athletic motivation on academic performance among university students. *Int. J. Trade Econ. Financ.* 1, 367–372. doi: 10.7763/IJTEF.2010.V1.65

Gaston-Gayles, J. L. (2005). The factor structure and reliability of the student athletes' motivation toward sports and academics questionnaire (SAMSAQ). *J. Coll. Stud. Dev.* 46, 317–327. doi: 10.1353/csd.2005.0025

Guidotti, F., Minganti, C., Cortis, C., Piacentini, M., Tessitore, A., and Capranica, L. (2013). Validation of the Italian version of the student athletes' motivation toward sport and academics questionnaire. *Sport Sci. Health.* 9, 51–58. doi: 10.1007/s11332-013-0145-x

Ham, C. (2003). The study on direction for improvement of intercollegiate athletics. korean. *J. Soc. Aerobic Exerc.* 7, 73–85.

Heo, H. (2011). Recognitions about elite athletes of sports culture and human rights by socio-demographic variables. *J. Korean Soc. Sociol. Sport.* 23, 145–161. doi: 10.22173/jksss.2010.23.4.145

Holland, P. W., and Wainer, H. (1993). *Differential Item Functioning*. New York, USA: Routledge, doi: 10.4324/9780203357811

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit Indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Huml, M. R. (2018). A factor structure examination of athletic identity related to NCAA divisional differences. *J. Coll. Stud. Dev.* 59, 376–381. doi: 10.1353/csd.2018.0035

Huml, M. R., Hancock, M. G., and Bergman, M. J. (2014). Additional support or extravagant cost?: student-athletes' perceptions on athletic academic centers. *J. Issues Intercoll. Athl.* 7, 410–430.

Jöreskog, K. G., and Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *J. Am. Stat. Assoc.* 70, 631–639. doi: 10.1080/01621459.1975.10482485

Karlgren, K., Lakkala, M., Toom, A., Ilomäki, L., Lahti-Nuuttila, P., and Muukkonen, H. (2020). Assessing the learning of knowledge work competence in higher education–cross-cultural translation and adaptation of the Collaborative Knowledge Practices Questionnaire. *Res. Papers Educ.* 35, 8–22. doi: 10.1080/02671522.2019.1677752

Kim, D. H. (2011). What made them give up their studies?: a study on problems and solutions of academic management for athletic meritocrat in university Korean. *J. Sport Sci.* 22, 2155–2171. doi: 10.24985/kjss.2011.22.3.2155

Kim, D. W., Kim, S. Y., and Kim, D. H. (2014). A narrative inquiry on the university experiences of student-athletes to study.korea*n.J. Phys. Educ.* 53, 329–340.

Kline, R. (2013). "Exploratory and Confirmatory Factor Analysis," in *Applied Quantitative Analysis in Education and the Social Sciences*, eds Y. Petscher, C. Schatschneider, and D. Compton (New York, USA: Routledge), 183–217. doi: 10.4324/9780203108550

Korean Sport and Olympic Committee (2020). *Statistics*. SouthKorea: Korean Sport and Olympic Committee.

Lee, H., and Yu, T. (2021). Exploring terminology of student-athlete for transition: implications for physical education.korean. *J. Sport Pedagogy* 28, 129–148. doi: 10.21812/kjsp.2021.4.28.2.129

Lee, J. (2006). The change of value on physical activity followed by social transitions in Korea. *J. Korean Philos. Soc. Pro. Sport Dance* 14, 1–12.

Lee, Y. G., and Kwon, H. I. (2013). Exploration of outcomes for the system development of learning right policy.Korean. *J. Phys. Educ.* 52, 527–538.

Lim, S., and Huh, J. (2009). A critical review of the literature on sports and physical education policies in the third Republic of South Korea.Korean. *J. Sport Law.* 12, 105–130. doi: 10.19051/kasel.2009.12.1.105

Lopes Dos Santos, M., Uftring, M. M., Stahl, C. A., Lockie, R. G., Alvar, B., Mann, J. B., et al. (2020). Stress in academic and athletic performance in collegiate athletes: a narrative review of sources and monitoring strategies. *Front. Sports Act. Living*. 2:42. doi: 10.3389/fspor.2020.00042

Lupo, C., Guidotti, F., Gon?alves, C. E., Moreira, L., Topic, M. D., Bellardini, H., et al. (2015). Motivation towards dual career of European student-athletes. *Eur. J. Sport Sci.* 15, 151–160. doi: 10.1080/17461391.2014.940557

Lupo, C., Mosso, C. O., Guidotti, F., Cugliari, G., Pizzigalli, L., and Rainoldi, A. (2017). Motivation toward dual career of Italian student-athletes enrolled in different university paths. *Sport Sci. Health* 12, 485–594. doi: 10.1007/s11332-016-0327-4

Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J., and Nagengast, B. (2011). Methodological measurement fruitfulness of exploratory structural equation modeling (ESEM): new approaches to key substantive issues in motivation and engagement. *J. Psychoeduc. Assess.* 29, 322–346. doi: 10.1177/0734282911406657

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A., Trautwein, U., et al. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychol. Assess.* 22, 471–491. doi: 10.1037/a0019227

Marsh, H. W., Morin, A., Parker, P. D., and Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annu. Rev. Clin. Psychol.* 10, 85–110. doi: 10.1146/annurev-clinpsy-032813-153700

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A., et al. (2009). Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. *Struct. Equ. Model.* 16, 439–476. doi: 10.1080/10705510903008220

McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. New Jersey, USA: Lawrence Erlbaum Associates, Inc.

Morin, A. J., Myers, N. D., and Lee, S. (2020). "Modern Factor Analytic Techniques. Bifactor Models, Exploratory Structural Equation Modeling (ESEM), and Bifactor-ESEM," in *Handbook of Sport Psychology*, eds G. Tenenbaum and R. Eklund (New York, USA: John Wiley & Sons, Inc), 1044–1073. doi: 10.1002/9781119568124.ch51

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika* 54, 557–585. doi: 10.1007/BF02296397

Muthén, L. K., and Muthén, B. O. (2005). *Mplus User's Guide: Statistical Analysis with Latent Variables*. Los Angeles, CA, USA: Muthén

Neff, K. D., Tóth-Király, I., Knox, M. C., Kuchar, A., and Davidson, O. (2021). The development and validation of the state self-compassion scale (long-and short form). *Mindfulness* 12, 121–140. doi: 10.1007/s12671-020-01505-4

Otto, M. G., Martinez, J. M., and Barnhill, C. R. (2019). How the perception of athletic academic services affects the overall college experience of freshmen student-athletes. *J. Athl. Dev. Exp.* 1, 40–52. doi: 10.25035/jade.01.01.05

Park, S., Hong, S., and Lee, M. (2015). Validation of the student athletes' motivation towards sports and academics questionnaire to korean student-athletes. *J. Exerc. Rehabil.* 11, 220–227. doi: 10.12965/jer.150202

Pestana, E. R., de Carvalho, W. R., Nunes, L. A., da Silva, Almeida, F. D., and Salvador, E. P. (2018). Sports practice and factors associated with school performance in grade and high school: comparison between athletes and non-athletes. *Sport Sci. Health* 14, 639–644. doi: 10.1007/s11332-018-0478-6

Pommier, E., Neff, K. D., and Tóth-Király, I. (2020). The development and validation of the Compassion Scale. *Assessment* 27, 21–39. doi: 10.1177/1073191119874108

Quinaud, R. T., Fernandes, A., Gonçalves, C., and Carvalho, H. (2020). Student-athletes' motivation and identity: variation among Brazilian and Portuguese university student-athletes. *Psychol. Rep.* 123, 1703–1723. doi: 10.1177/0033294119892885

Quinaud, R. T., Gonçalves, C. E., Possamai, K., Morais, C. Z., Capranica, L., and Carvalho, H. M. (2021). Validity and usefulness of the student-athletes' motivation toward sport and academics questionnaire: a Bayesian multilevel approach. *PeerJ* 9:e11863. doi: 10.7717/peerj.11863

Sancho, P., Pinazo-Hernandis, S., Donio-Bellegarde, M., and Tomás, J. M. (2020). Validation of the University of California, Los Angeles Loneliness Scale (version 3) in Spanish older population: an application of exploratory structural equation modelling. *Aust. Psychol.* 55, 283–292. doi: 10.1111/ap.12428

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1007/978-3-319-10470-6_18

Sherry, M., and Zeller, K. (2014). Gender and motivation: a study of the athletic and academic motivations of division I female college basketball players. *Women's Stud.* 43, 73–92. doi: 10.1080/00497878.2014.852425

Sousa, V. D., and Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J. Eval. Clin. Prac.* 17, 268–274. doi: 10.1111/j.1365-2753.2010.01434.x

Taber, K. S. (2018). The use of cronbach's alpha when developing and reporting research instruments in science education. *Res. Sci. Educ.* 48, 1273–1296. doi: 10.1007/s11165-016-9602-2

Vandenberg, R. J., and Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ. Res. Methods* 3, 4–69. doi: 10.1177/109442810031002

Weiner, B. (1986). *An Attributional Theory of Motivation and Emotion*. Germany: Springer, 159–190. doi: 10.1007/978-1-4612-4948-1_6

Yoo, K. (2016). *The average age of retired South Korean professional athletes," in Korea News1*. Available online at: https://www.news1.kr/articles/?258592 (accessed Novemeber 12, 2021).

Conflict of Interest: 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 | Frontiers in Psychology

ORIGINAL RESEARCH
published: 17 May 2022
doi: 10.3389/fpsyg.2022.828081

Pandemic or Not, Worker Subjective Wellbeing Pivots About the Living Wage Point: A Replication, Extension, and Policy Challenge in Aotearoa New Zealand

Stuart C. Carr[1], Jarrod Haar[2], Darrin Hodgetts[1], Harvey Jones[1], James Arrowsmith[3], Jane Parker[3], Amanda Young-Hauser[1] and Siautu Alefaio[1]*

[1]School of Psychology, Massey University, Auckland, New Zealand, [2]Department of Management, Auckland University of Technology (AUT), Auckland, New Zealand, [3]School of Management, Massey University, Auckland, New Zealand

OPEN ACCESS

Edited by:
José Antonio Lozano Lozano, Universidad Autónoma de Chile, Chile

Reviewed by:
Jefferson Staduto, Universidade Estadual do Oeste do Paraná, Brazil
Marcin Wnuk, Adam Mickiewicz University, Poland
Ari Saptono, Jakarta State University, Indonesia

***Correspondence:**
Stuart C. Carr
s.c.carr@massey.ac.nz

Specialty section:
This article was submitted to Organizational Psychology, a section of the journal Frontiers in Psychology

Received: *02 December 2021*
Accepted: *25 April 2022*
Published: *17 May 2022*

Citation:
Carr SC, Haar J, Hodgetts D, Jones H, Arrowsmith J, Parker J, Young-Hauser A and Alefaio S (2022) Pandemic or Not, Worker Subjective Wellbeing Pivots About the Living Wage Point: A Replication, Extension, and Policy Challenge in Aotearoa New Zealand. Front. Psychol. 13:828081. doi: 10.3389/fpsyg.2022.828081

Recent pre-pandemic research suggests that living wages can be pivotal for enhancing employee attitudes and subjective wellbeing. This article explores whether or not the present COVID-19 pandemic is impacting pivotal links between living wages and employee attitudes and subjective wellbeing, with replication indicating robustness. Twin cohorts each of 1,000 low-waged workers across New Zealand (NZ), one pre-(2018), and one present-pandemic (2020) were sample surveyed on hourly wage, job attitudes, and subjective wellbeing as linked to changes in the world of work associated with the pandemic (e.g., job security, stress, anxiety, depression, and holistic wellbeing). Using locally estimated scatter-point smoothing, job attitudes and subjective wellbeing scores tended to pivot upward at the living wage level in NZ. These findings replicate earlier findings and extend these into considering subjective wellbeing in the context of a crisis for employee livelihoods and lives more generally. Convergence across multiple measures, constructs, and contexts, suggests the positive impacts of living wages are durable. We draw inspiration from systems dynamics to argue that the present government policy of raising legal minimum wages (as NZ has done) may not protect subjective wellbeing until wages cross the living wage Rubicon. Future research should address this challenge.

Keywords: minimum wage, living wage, pandemic, job attitude, wellbeing, decent work

INTRODUCTION

According to the International Labour Organization (ILO), just prior to COVID-19 the number one challenge in and for the world of work [International Labour Organization (ILO), 2019a] was addressing poor in-work conditions and unliveable wages. According to the ILO in 2019, 3.3 billion people experienced these conditions [International Labour Organization (ILO), 2019b], which was 19 times more than the global unemployment rate (172 million). Two-thirds of the world's entire workforce was thereby working in conditions

that were informal, lacking a proper job description, employment contract, protection in case of injury, regular hours, paid leave provision, social protection, and/or a regular liveable wage. The remaining third were supposed to be protected by a formal, legal Minimum wage floor. However, a survey of 14,000 workers across 14 different countries and economies found that almost two thirds of workers, many in so-called higher-income economies with formal jobs, were "struggling" to make ends meet [International Trade Union Confederation (ITUC), 2018]. Thus the world of work immediately prior to the COVID-19 pandemic could be characterized by in-work precariousness and poverty wages.

Since 2020, the COVID-19 virus has disrupted the whole world of work, and underscored the need for wages worldwide to keep pace much more with people's everyday needs for decent work conditions that protect their subjective wellbeing, including living wages [International Labour Organization (ILO), 2020, 2021, 2022]. Unlike minimum wages (Smith, 2015), living wages tend to be voluntary rather than statutory and to aim higher than bare subsistence, including affording some disposable income for people to participate with dignity in social life, enjoy occasional treats, and have some financial reserves to buffer them when crises strike. In New Zealand for example, living wages typically include meeting not only material needs like housing and food, but also social needs such as living with dignity and socio-economic inclusion (King and Waldegrave, 2012, 2014). In May 2020, needs like these were recognized in a "wellbeing budget," which included a commitment to improving wage conditions as part of a broader strategy to promote and protect the wellbeing of the population (New Zealand Treasury, 2020).

Addressing the issue of poverty wages is now very urgent both globally and locally. In 2022, according to the Director-General of the ILO, Guy Ryder, "the global employment and social outlook remains uncertain and fragile" (2022, p. 3), with unemployment being projected to rise (to over 200 m) in 2022, working hours to drop (currently by 2%) and the most vulnerable occupations and smaller organizations within them, hit hardest of all (*ibid*, p. 11). NZ is an example of such a country, yet research on the potential *benefits*, e.g., to subjective and societal wellbeing of raising wages in NZ remains sparse (Carr, 2022). The gap that this paper addresses, and the novelty of its contribution, is to explore the links between wage levels and subjective wellbeing, in the context of a public health crisis.

Pre-pandemic, work research on wage and wellbeing has mainly focused on "job attitudes" like job satisfaction and affective commitment, which are organizationally focused, rather than on workers' wider subjective wellbeing (e.g., Kuvaas, 2006; Judge et al., 2010). **Figure 1** presents three theoretical relationships between wage value and job attitudes/subjective wellbeing (Carr et al., 2016). The simplest linkage is linear (black line). Linear implies that there will be a discrete value of wage at which the criteria of subjective wellbeing and job attitudes switch from being negative to positive (≈). Linear linkages have though proved at best disappointingly weak (Young et al., 2014). One reason could be that the link itself may not be linear, for example because wage increments matter more economically, and thus psychologically, at lower than higher wage levels (George and Brief, 1989).

One way in which wage may matter more for lower waged roles is shown in **Figure 1** as a blue dotted line. This arc is based on a theory of diminishing marginal returns (Carr et al., 2016). Any wage is a good wage, and especially so at lower levels, where any kind of increment, most of all from zero to one, would make a just noticeable difference. Any resulting curvilinear relationship would thereby be more consistent with reforming minimum, rather than reaching living wages, with minimum cost and thereby lowered risk of job losses during economic and health crises like the COVID-19 pandemic (Law, 2020).

In contrast, working poverty trap theory predicts that workers paid any wage that was less than a viable or living wage, regardless of level, will be stuck at consistently low or negative levels of job attitudes and subjective wellbeing (Carr et al., 2016). This stasis in negative terrain is captured as a flattened section of the curve in **Figure 1** at the bottom left of the continuous blue line. Tracking left to right, the solid blue curve begins to rise and eventually crosses the threshold between negative and positive job attitudes/subjective wellbeing (≈) at the top right in **Figure 1**. Any living wage would thereby need to be set at or near to the monetary value on the *x*-axis where the curve begins to cross ≈ on the *y*-axis. Only from there onwards would it produce diminishing marginal returns.

The last two linkages have not yet—to the best of our knowledge—been fully tested against each other, or been corroborated with systematic evidence. This issue is very important to resolve because of the competing predictions about where to set wages if working poverty is to be avoided and subjective wellbeing enhanced. In **Figure 1**, diminishing marginal returns (the dotted line) would suggest that there is no working poverty trap. Instead, it suggests that any increase in the minimum wage would make a noticeable difference to job attitudes and worker subjective wellbeing. More importantly still perhaps, the point at which people's attitudes and subjective wellbeing change valence, from negative to positive, unhappiness to happiness and ill-being to well-being (≈) could be set lower than it would need to be set if there was a working poverty trap (sigmoidal function and solid blue curve).

Of the three theoretical linkages in **Figure 1**, the available evidence to date, though tentative, is most consistent with the solid blue, sigmoidal curve. Samples of low-waged workers from South Africa and New Zealand, conducted prior to the pandemic, have yielded a consistent pattern (Carr et al., 2018, 2019; Haar et al., 2018). Workers at or near the current Minimum wage, in each respective country, reported struggling to make ends meet, as well as tending to report negative job attitudes (job [dis]satisfaction, [dis]empowerment, and [lack of] occupational pride), and expressing a subjective sense of [un]fairness and [poor] quality of life. Only above the current respective living wage did dissatisfaction, disempowerment, and so on tend to change valence to satisfaction, empowerment, and so forth. Nevertheless, the samples in this study have been relatively small and localized, with measures that were also non-standardized. Such concerns led Carr et al. (2018)

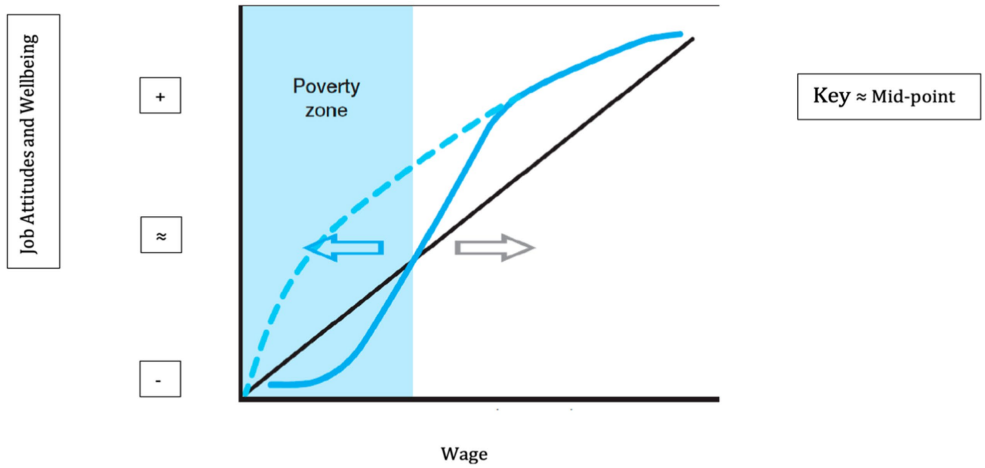


FIGURE 1 | Theoretical links between wage and wellbeing/job attitudes. Adapted with permission from Carr et al. (2016). Wellbeing, in the figure, refers in this paper to subjective wellbeing.

to call for more systematic study of the three hypothetical links in **Figure 1**.

A subsequent study, conducted in NZ, included a larger, nationwide sample ($N > 1{,}000$) and standardized measures (Carr et al., 2019). This study also replicated findings from Carr et al. (2018). It focused on job attitudes (job satisfaction, work engagement, meaningful empowerment, affective commitment, organizational citizenship behaviors, and work life balance). Among these, only job satisfaction and work life balance have links to subjective wellbeing, specifically happiness at work (Fisher, 2009); with a potential for spill-over into respondents' lives more generally (Haar et al., 2014). Subsequent qualitative analysis of individual 'outliers' found signs of mental health issues arising from precarious wage conditions, including stress, anxiety, and depression (Carr et al., 2019). These findings in turn led Carr et al. (2019) to call for further exploration of potential linkages between living wages and mental health.

Public health research on minimum wages has found a range of likely state-level benefits from state-level increases on population wellbeing, including physical (Lenhart, 2017) and mental health (Leigh et al., 2019). Examples would include reductions from diseases of the circulatory system that lead to premature deaths from conditions such as strokes (Tsao et al., 2016) and suicide (Kaufman et al., 2020). However, societal-level studies cannot tell us directly if state-level improvements are only among minimum wage workers or also for people paid more from wage inflation, including among those paid a living wage. Moreover, everyday forms of mental wellbeing at and through work (e.g., job stress, experiences of job insecurity, anxiety, and depression) remain largely unconnected within the relatively new field of wage and subjective wellbeing (Leigh et al., 2019). Triggers for stress-induced strains on subjective wellbeing have included concerns about wages, especially for those below the median income (Chuluun et al., 2016). Additionally, employees who perceive their organization as unfair and feel job insecure may also be at higher risk of emotional exhaustion and work stress (Kausto et al., 2005).

Subjective wellbeing concerns like these, which are due to precarious work and wage conditions, have almost certainly been amplified by the COVID-19 pandemic. In NZ for example, a nation-wide survey of over 3,000 people has found that almost 40 percent of households experienced a significant drop in work-related income (Galicki, 2020). One-in-four households were caught in arrears on at least one payment (including consumer loans, utility bills, and housing costs). One in 10 had missed a rent or mortgage payment. Forty-one percent either agreed or strongly agreed that thinking about their financial situation made them anxious. Anxiety like this, about unmet financial needs, can contribute to a range of related mental health issues, including anxiety and depression (Shaw and Gupta, 2001; Stride et al., 2007). The question then is could a living wage[1] provide any social protection for people's mental health, by helping to meet people's financial needs in NZ during a pandemic?

Pandemics are inherently and intrinsically threatening to both physical and mental wellbeing. This threat renders an assessment of the links between wage and subjective wellbeing, which Carr et al. (2019) were already calling for prior to COVID-19, even more timely and relevant. Pandemic conditions have not only threatened jobs and worker wellbeing. They have also changed the conditions of work itself. In NZ for instance,[2] overall employment effects of COVID-19 were moderated by substantial public policies directed at subsidizing wages to protect jobs and assisting selected industries and small businesses. There were programs directed at health care, vulnerable groups, and easing risks of mortgage defaults and the eviction of renters. At the same time, there were also substantial declines in hours worked, which reduced

[1]In NZ, the campaign Living wage rate is adjusted each year in response to cost of living, by the New Zealand Family Centre Social Policy Unit (King and Waldegrave, 2012, 2014).

[2]We are grateful to an informal ILO reviewer for this helpful section, which is included almost verbatim.

annual wages and incomes for many households. These COVID-19 policies and labor market changes conceivably had direct effects on the wellbeing of low-income workers as well as an indirect effect on their observed job attitudes and subjective wellbeing, causing potentially untold compositional changes in the unobserved characteristics of the low-income workers who remained employed and the jobs that they held. In other words, COVID-19 has been a great disruptor.

Disruptors are not only negative or destructive, however. They also bring unique learning opportunities. In research, one of these is the opportunity to see if any given function will replicate even under radically changed circumstances. Any finding that a function in **Figure 1** remained similar over radically different circumstances will "strongly attest to its durability across time" (Gergen, 1973, p. 315), by signaling robustness in that function (Schupbach, 1998). Thus a replication of the cusped curve in **Figure 1**, under disrupted pandemic conditions, *would give confidence that a living wage is pivotal for subjective wellbeing*.

Policy wise, this pandemic has also re-ignited heated and often fractious debate in NZ regarding the need for a living wage (Waldegrave, 2020). This debate is not unique to NZ (e.g., Wood, 1997; Kantor, 2016). On the one hand, advocates of living wages present any wage increase as the right thing to do for social inclusion and shared prosperity (Krieble, 2020). On the other hand, advocates of wage restraint, including some influential employer lobby groups, insist that raising wages to living wage levels, especially during a pandemic and associated economic crisis, will only lead to job losses, more job insecurity, and even less social inclusion (Law, 2020). To date, however, this debate has largely overlooked both job attitudes and worker subjective wellbeing, which are important considerations that have been linked to work productivity, at both individual (*per capita*) and unit (aggregated organizational and sub-organizational) levels (respectively, Harter et al., 2002 and Harrison et al., 2006). Evidence on how wage relates to worker job attitudes and subjective wellbeing might therefore cast light on a heated debate about the social *and* economic wisdom of wage increases, precisely at a time of crisis for sustaining livelihoods (United Nations, 2021).

Briefly, the overall objective in this study was to explore whether the sigmoidal relationship or alternatively either of the other two competing relationships in **Figure 1** would replicate in NZ. Replicability was assessed in two main ways: (a) across two cohorts, one sampled before and one during the current pandemic; and (b) across two sets of variables, one focused on different job attitudes and the other more on humanitarian features of people's everyday subjective wellbeing.

The overarching aim of this study was to assess if the concept of a living wage has practical and policy implications. Taking an evidence-based approach, we wanted to know whether there was any consistently identifiable, actual material wage value whereabouts workers in NZ would tend to report feeling better, not only about their work attitudes, but also in their wider subjective wellbeing.

MATERIALS AND METHODS

Participants

Participants were drawn from two national cohorts of lower income workers, one during March/early April 2018 ($n_1 = 1{,}011$) and another in September/October 2020 ($n_2 = 1{,}027$). All respondents had to be in paid employment with an annual personal income before tax of under NZ$60,000. A professional survey organization, Qualtrics, was engaged to draw two samples from across NZ, one in 2018 and a second in 2020 (Haar et al., 2018). The Qualtrics system has an estimated time for surveys, and removes respondents who complete the survey too quickly or too slowly. It also assures that one respondent only can complete the survey. This approach to data collection has grown and provided useful samples for researchers (Ferguson et al., 2014; Kaplan et al., 2016; Vitell et al., 2016). We utilized this approach specifically because Qualtrics can target income-level within their respondent recruitment. It pays respondents for their time, but the nature of this arrangement is proprietary.

In the 2018 cohort, by income level, the modal reported annual income, expressed in brackets of NZ$20,000 s, was NZ$40–60,000 band (39% of sample), followed by NZ$20–40,000 (36%), and then up to NZ$20,000 (25%). Most workers were paid hourly (71%) rather than being salaried (29%), 86% with one job, and working full-time (51%). Our lower-waged sample was skewed toward female workers (69%), with the modal age category being 36–45 years (19%). Ethnically, the majority reported as "NZ European" (62%) with the next largest category identifying as Māori (11%). By sector, the majority worked in the private sector (68%) rather than in either public service (17%) or in civil society organizations like Non-Government Organizations (NGOs) and charities (15%). These proportions have already been found to be reasonably representative of the lower-end of the wage spectrum and economy across NZ at the time (Carr et al., 2019).

In the 2020 cohort, the modal income level was NZ$40–60,000 (50%), followed by NZ$20–40,000 (34%) and up to NZ$20,000 (17%). A majority of workers (66%), as in 2018, were paid hourly (66%), 86% with one job, and working full-time (53%). As in the 2018 cohort, there was a skew toward female workers in this low-waged sample (65%), with the same modal age category (of 36–45 years, 22%). The two top ethnicities were NZ European and Māori (respectively, 65 and 12%). Private sector work was once more predominant (67%) over public service (20%) and civil society (13%).

Measures

In addition to a range of standard demographic items (Haar et al., 2018; Carr et al., 2019), we focused in this paper on three particular sets of variables, reflecting (i) wage level, (ii) job attitudes, and (iii) subjective wellbeing. These were examined separately over the two cohorts in 2018 and 2020.

Wage

Hourly pay and annual income, number of paying jobs, and full or part-time employment (*cf.*, Carr et al., 2019). In 2020 only, we asked whether during the preceding maximum level 4,

full lockdown people had been able to work from home, had a pay cut or bonus, experienced a temporary layoff, and cuts to pay (including hours) and whether these were back to normal. We also asked if cost-of-living (during lockdown) went up or down.

Job Attitudes

From **Table 1**, we measured Job satisfaction using three-item (Judge et al., 2005; $\alpha = 0.91$, 0.91); work engagement on nine-item omnibus measure of Schaufeli et al. (2001)($\alpha = 0.92$, 0.92); career satisfaction with three items of Greenhaus et al. (1990) ($\alpha = 0.85$, 0.87); meaningful work with three items from Spreitzer (1995) ($\alpha = 0.93$, 0.91); affective commitment with three items from Meyer et al. (1993); ($\alpha = 0.78$, 0.76); organizational citizenship behaviors (OCBs) using four items from Lee and Allen (2002) ($\alpha = 0.84$, 0.82); and work-life balance with three items from Haar (2013) ($\alpha = 0.88$, 0.86). In 2018, these measures were subjected to one combined confirmatory factor analysis (CFA; Haar et al., 2018). We used (1) the comparative fit index (CFI $\geq$ 0.95), (2) the root mean square error of approximation (RMSEA $\leq$ 0.08), and (3) the standardized root mean residual (SRMR $\leq$ 0.10). The combined CFA confirmed that our measures were each internally coherent, distinctive from each other and relatively free of common method bias (for details, Haar et al., 2018).

TABLE 1 | Key psychological measures.

Construct	Source	Exemplar item
Job Attitudes		
Job Satisfaction	Judge et al., 2005	"I find real enjoyment in my work"
Work Engagement	Schaufeli et al., 2001	"I am proud of the work that I do"
Career Satisfaction	Greenhaus et al., 1990	"I am satisfied with the success I Have achieved in my career/work"
Meaningful Work	Spreitzer, 1995	"My job activities are meaningful to me"
Affective Commitment	Meyer et al., 1993	"I would be very happy to spend the rest of my career with this organization"
Organizational Citizenship Behavior (OCB)	Lee and Allen, 2002	"I assist others with their duties"
Work-Life Balance	Haar, 2013	"I manage to balance the demands and personal/family life equally well"
Subjective Wellbeing		
Job Security	Armstrong-Stassen, 2001	"I am worried about being laid off"
Job Stress	Stanton et al., 2001	"Overall, how would you rate your stress from 0 to 10?"*
Anxiety	Axtell et al., 2002	"Calm—Never…. Always[five-points]"*
Depression	Axtell et al., 2002	"Optimistic—Never-Always…"*
Holistic Wellbeing	Tomlyn and Cummins, 2011	"How satisfied are you with Your health?"

*Reverse-scored.

Subjective Wellbeing

Job security was measured using a three-item measure from Armstrong-Stassen (2001) ($\alpha = 0.88$, 0.92); job stress using a single item in Stanton et al. (2001); ["Overall, how would you rate your stress from 0 (no stress) through 5 (neutral) to 10 (extreme stress)?"]; anxiety and depression with three items each from Axtell et al. (2002) (respectively, $\alpha = 0.92$, 0.92; $\alpha = 0.92$, 0.91). Holistic subjective wellbeing was assessed on a wide-ranging physical-mental-spiritual 10-item measure in Tomlyn and Cummins (2011) ($\alpha = 0.91$; 0.92).

Measurement Models

Using Analysis Of Moment Structures (AMOS) version 26, we conducted a pair of CFAs for each cohort on the job attitudes and subjective wellbeing constructs. We followed Williams et al. (2009) regarding assessing model fit: (1) the comparative fit index (CFI $\geq$ 0.95), (2) the root-mean-square error of approximation (RMSEA $\leq$ 0.08), and (3) the standardized root mean residual (SRMR $\leq$ 0.10). Overall, from **Table 2**, the hypothesized measurement model was the best fit for the data meeting all minimum thresholds job attitudes We ran alternative CFAs (combining various constructs) and these all resulted in poorer fit models (all $p < 0.001$; Hair et al., 2010).

Procedure

This project was funded by the Royal Society of NZ (RSNZ) Marsden Fund (17-MAU/137). Ethical approval was obtained from Massey University Human Ethics Committee (MUHEC) (2018). All participants were assured of confidentiality and remained anonymous to the researchers. As noted under "Participants," the survey was designed by the authors and distributed *via* a private research company, Qualtrics. We utilized this approach because Qualtrics can target income level in respondent recruitment and because their respondents are already familiar with surveys. Respondent familiarity may have introduced a familiarity bias (with survey forms), but this is offset by the possibility of using multiple items (more familiar

TABLE 2 | Results of confirmatory factor analysis (CFA).

	χ^2	*df*	CFI	RMSEA	SRMR
Job Attitudes					
2018	1301.9*	328	0.95	0.05	0.05
2020	1257.9*	328	0.96	0.05	0.05
Subjective Wellbeing					
2018	896.5*	161	0.95	0.07	0.04
2020	1012.1*	161	0.95	0.07	0.04

*$p < 0.001$.

to such panels), increasing reliability and validity (Carr et al., 2019). During piloting, lack of familiarity was identified as a barrier to participation by lower-income groups (Haar et al., 2018).

A first survey took place before the pandemic began, in the first quarter of 2018. A second survey took place after the maximum level, full "level 4" lockdown from March 31, 2020 to April 27, 2020 had ended and life (including work life) had pretty much returned to normal. These twin survey cohorts hence straddled pre-pandemic and pandemic.

RESULTS

Following Carr et al. (2018), we first used curve estimation to explore the best-fitting function (line, logarithmic curve, or sigmoid/cubic, as in **Figure 1**) to the data for each job attitude and subjective wellbeing measure, as a function of hourly wage. In both cohorts, hourly wage was clearly the modal means of payment, not annual salary (cohort 1 $n = 722$; cohort 2 $n = 680$). Accordingly, we again focused mainly on this form of payment. In 2020, we used current (COVID-affected) rather than normal hourly rate (which differed for $n = 25$ respondents). Within cohort 1 and 2, there were potentially distorting wage outliers that ranged from NZ $0/h at one extreme to $2,050/h at the other. In order to maintain consistency, we selected cases who were paid anywhere from legal adult minimum hourly wage at the time hour up to and including $40/h (≈$60,000/annum for 30 h/week).

Across job attitudes, there was a pattern that replicated across the changed circumstances in cohorts 1 and 2 (**Table 3**). First, job attitudes covaried significantly ($p < 0.01$) with hourly wage. Second, the link was non-linear (the only partial exception being OCB). Third, the form of curve was predominantly sigmoidal not logarithmic, resembling the solid curve in **Figure 1** (Carr et al., 2016). There were three instances where the precise nature of the curve was inconclusive, at this stage.

Variances explained in **Table 3**. Though well within normal limits in psychological research, these were not high. This could be partly due to the cusped nature of sigmoidal functions, to which curve estimation may not be sensitive. To explore that possibility further, we applied Locally Estimated Scatterplot Smoothing (LOESS). LOESS is more sensitive than curve estimation to points of inflexion (Carr et al., 2019).

From **Figure 2**, there was a consistent pattern, with the lowest-waged workers tending to score relatively low on job attitudes. This is seen in a clear wage-subjective wellbeing tail, in which there was little or no increase in job attitude score for increments in wage. In other words, there was a consistent working poverty trap. Comparing the two time periods left and right in **Figure 2**, the curve consistently pivoted from flat to zero to positive gradient, flat to upward, and peaked first about $NZ20 per hour (the NZ Living wage in 2018 and 2020, respectively, was $20.55 and $22.10 per hour). Thereafter, there tended to be a slight dip, followed by a curve indicative of diminishing marginal returns (just noticeable differences). Thus, for job attitudes, the data summarized in **Table 3** and **Figure 2** point convergently to the solid blue curve in **Figure 1**.

TABLE 3 | Curve estimations for Job Attitudes in cohorts 1 and 2.

Attitude	Cohort	Best fit	*F* value	*df*	Percent variance
Job Satisfaction	1	Cubic	10.16****	3,423	6.7
	2	Cubic	10.05****	2,445	4.3
Work Engagement	1	Cubic	7.42****	3,423	5.0
	2	Cubic	5.86***	2,445	2.6
Career satisfaction	1	Cubic	13.78****	3,423	8.9
	2	Cubic	14.14****	2,445	6.0
Meaningful work	1	Cubic	12.95****	3,423	8.4
	2	Cubic	9.07****	2,445	3.9
Affective commitment	1	Cubic	7.72****	3,423	5.2
	2	Cubic	7.88****	2,445	3.4
OCB	1	Cubic/Logarithmic/ Linear****		Tied	4.6
	2	Cubic/Linear****		Tied	4.6
Work Life Balance	1	Cubic	3.86**	3,423	2.7
	2	Cubic/Logarithmic****		Tied	5.0

***p < 0.01; ***p < 0.005; ****p < 0.001.*

Curve estimations for subjective wellbeing variables were increasingly distal from the workplace itself (**Table 4**). Accordingly, the percentages of variance explained were less than in **Table 3**. Non-linear relationships still predominated over linear linkages, although the precise nature of the curvature (logarithmic or cubic) was more often inconclusive. Again therefore, we utilized LOESS regression analysis, which were more capable of detecting any potential cusp-like inflexions in the curve, which had not been detected using curve estimation. LOESS curves charting the links between hourly wage on the one hand and subjective wellbeing indicators on the other, in each respective cohort (i.e., during 2018 and 2020) are presented in **Figure 3**.

From **Figure 3**, there was both pattern and a possible slight deviation from pattern. First, the relationships were cusped, with discernible mini peaks at around NZ$20 per hour. A visible exception to these consistent patterns was the curvature for job (in)security, which peaked nearer $25/h. The threshold for job security may be understandably higher during an economically uncertain pandemic.

Post hoc, we conducted a check on this level of uncertainty in 2020 (Materials and Methods). Almost half the full sample (47%) could not work from home during the full lockdown period, which rose to fully 76% for workers paid by the hour. One third of workers had experienced some form of pay cut during the lockdown itself. One-in-five had been temporarily laid off, among whom 40% had not received any wage subsidy. Only 13%–14% of cohort/hourly-paid workers had received a wage bonus. By the time of the survey that was taken in 2020, the overall pay rate was back-to-normal for 87/88% of the cohort/hourly-paid. Where it was not, waged hours were implicated. Among employees paid by-the-hour, approximately one-in-three reported a cut in their waged hours during lockdown, with the number of paid hours still not having returned to normal for 20 percent of them. During the lockdown, cost-of-living rose for 30% of the cohort, and fell for 26%, but remained the same for 40% of the overall sample.

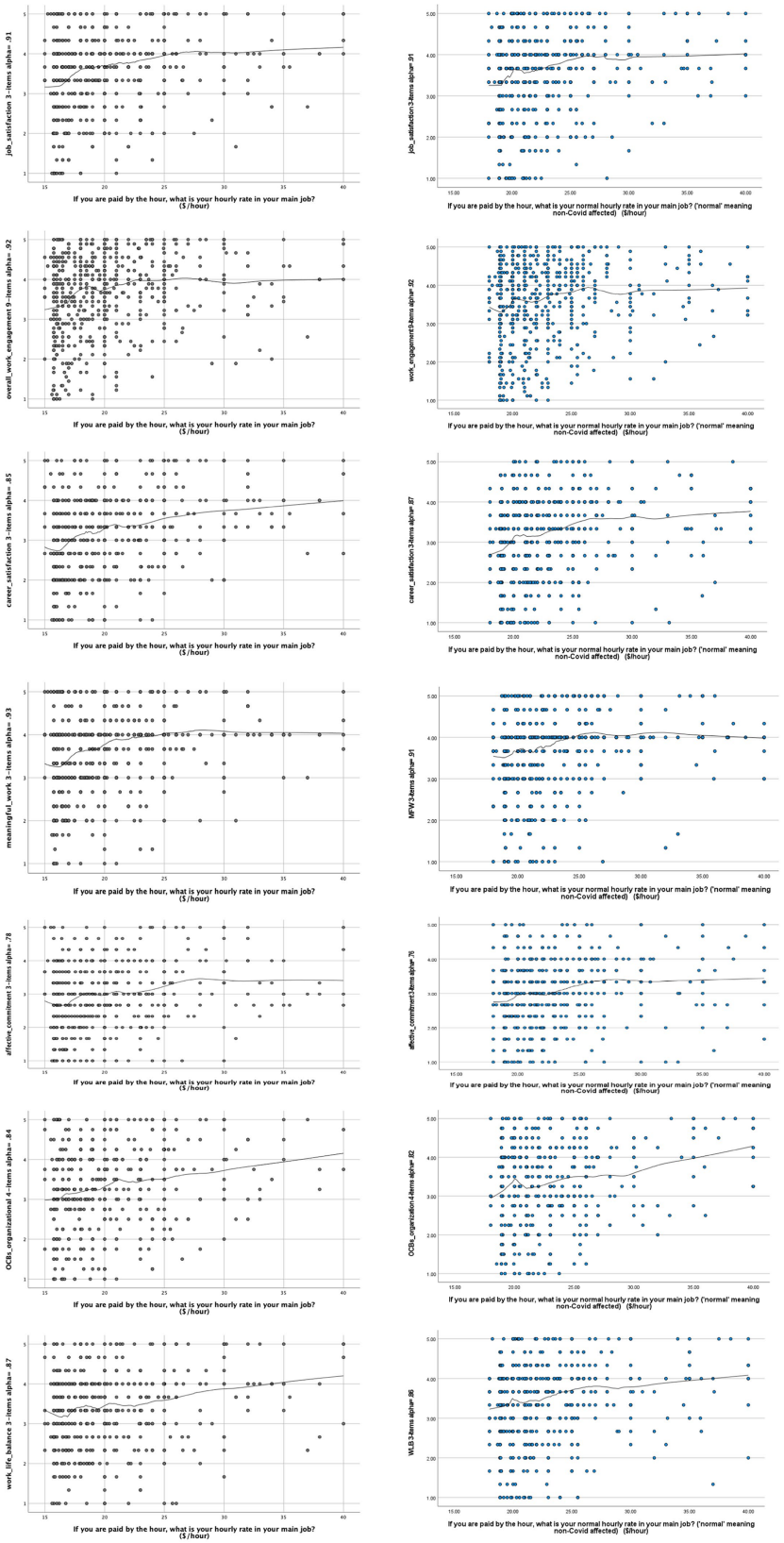


FIGURE 2 | Locally Estimated Scatterplot Smoothing (LOESS) curves for Job Attitudes ($N = 593 + 619 = 1,212$, tension parameter = 0.35). We reset the range of sampled wage values to include legal Minimum wage and above ($n = 21$ excluded).

FIGURE 3 | LOESS curves for Subjective Wellbeing ($N = 1,212$, tension parameter = 0.035–40).

TABLE 4 | Curve estimations for Subjective Wellbeing in cohorts 1 and 2.

Subjective Wellbeing facet	Cohort	Best fit	*F* value	*df*	Percent variance
Job (in) Security	1	Cubic	4.04**	3,574	2.1
	2	Cubic	3.77*	2,630	1.2
Job Stress	1	Cubic	2.36 ($p = 0.07$)	3,574	1.2
	2	Cubic/Logarithmic		Tied	1.1
Anxiety	1	Cubic	3.46*	3,574	1.8
	2	Linear/Logarithmic/ Cubic****		Tied	2.4
Depression	1	Cubic	5.94****	3,574	3.0
	2	Linear/Logarithmic/ Cubic*		Tied	1.4
Holistic wellbeing	1	Linear/Logarithmic/ Cubic****		Tied	3.6
	2	Cubic	10.68****	2,630	3.3

*$p<0.05$; **$p<0.01$; ****$p<0.001$.

DISCUSSION

In terms of theory, our findings were most consistent with the concept of a workplace poverty trap, meaning the need for wages to exceed a certain threshold value before job attitudes AND subjective wellbeing are properly safeguarded (Krieble, 2020). Looking back at **Figure 1**, the findings in this extended study are in general more consistent with the blue sigmoidal curve than with the concept of diminishing marginal returns (or just noticeable differences). According to this rival perspective, any job is a good job, and any wage is a good wage (Law, 2020). In our study, across both waves, and both different circumstances, just noticeable differences (JNDs) only became apparent not before, but rather *after* the living wage threshold was crossed (Carr et al., 2016). The consistency of this pattern, in which JNDs were more visible only well above the living wage and not below it, is indicative of robustness (Gergen, 1973; Schupbach, 1998). Replicability was sustained through a pandemic, and across from job attitudes into subjective wellbeing. Indeed, comparing **Figures 2, 3**, the poverty trap zone was visibly more clearly "under" the waterline (≈) in the case of subjective wellbeing than it was for job attitudes. This indicates that reaching the living wage was especially salient for subjective wellbeing.

The amount of variation that we observed about the LOESS curves (**Figures 2, 3**) was consistent with other studies in NZ (Carr et al., 2018, 2019). It likely reflects the everyday diversity of people's work and life circumstances (Carr et al., 2021), which includes work (and wider living) conditions other than hourly wage, such as average number of hours worked per week (Carr et al., 2021). In this study, we did not compute job attitudes and subjective wellbeing as a function of hours worked. We used an arguably cruder measure, mainly rate of pay per hour. This was nonetheless the pay modality specified in both the minimum and living wage in NZ. It was also the modal form of pay for our sample of relatively low-waged workers. Although a majority of respondents were employed on a full-time basis, almost half (Method) worked on part-time contracts, with variable hours. Variations in the number of hours worked per week may thereby account for at least some of the spread of data points about the curves marking central tendencies in **Figures 2, 3** (and across both hourly-paid and salaried workers). Future research may nonetheless explore hours worked as a moderator of links from pay (hourly and salaried) to job attitudes/subjective wellbeing.

Our research took place in a period when the government of NZ was focused on raising the legal minimum wage to more closely approximate the country's higher, but also voluntary living wage. The two surveys of cohorts took place approximately 2.5 years apart, from April 2018 to October 2020. During this time, the mandatory legal minimum wage rose from NZ$16.50 to NZ$18.90 per hour (an increase of 14.5%, reflected in the rightward deflection of the lowest wage point in **Figures 2, 3** from 2018 to 2020). The living wage rose from $20.55 to $22.20 per hour (an increase of 8%). Overall, between the first and second survey cohorts, the minimum wage made a net gain on the living wage in the order of more than 5%. This gain was exclusive of the rate of price inflation from 2018 to 2020, although the rate was relatively low, at just 1.6% *per annum* for both these years (Statista, 2021).

Despite this gain by the minimum on living wage, **Figures 2, 3** show that the NZ$ wage-value at which people tended to report subjective wellbeing *above* the waterline (≈) remained visibly *above* the actual minimum wage value, across both cohorts. Whether subsequent planned minimum wage increases (in 2021, after this study was over) will be sufficient to significantly reduce the poverty trap tail in **Figure 1** remains unknown until future tests can be conducted (forthcoming).

In the meantime, a focused wage reform alone may not be enough to eradicate (working) poverty. A wider approach could be required, in which a whole suite of policies, from wages to housing, are reset as part of a wider assemblage (Hopner et al., 2021). COVID-19 has not destroyed infrastructure and installed production capacity, in fact it is disorganizing production chains and changing many aspects of the labor market, as well as aggravating social and economic problems, especially for low-income workers. Policies aimed at social protection have been adopted in several low- and middle-income countries, such as conditional cash transfers, housing programs for the low-income population, free health care, etc., and could mitigate the problem of low-income workers.[3] These initiatives deserve more applied and evaluative research attention, focused on how they integrate vs. possibly undercut each other to affect people's everyday work-related subjective wellbeing (Hasdell, 2020).

With respect to wage reforms and addressing in-work poverty in NZ for example, the success of any particular reform may depend not only just on wage reform policy itself, but also on how well the government combats *other* social issues, like housing unaffordability, as part of a wider, more concerted push. For example, if rental prices simply rise ahead of any centrally-implemented rises in the legal minimum wage, then any increases in take-home wages will simply be eroded by the increased cost of housing rental. COVID-19 has put additional strain on people and livelihoods, as well as government coffers.

[3]We attribute this suggestion with appreciation to one of our expert peer reviewers, from whom it has been borrowed almost verbatim.

Thus a more systems-wide approach to tackling in-work poverty, going even wider than the current wellbeing budget (New Zealand Treasury, 2020) may still be required—and require to be evaluated in future research.

With respect to future research, our study focused on subjective wellbeing, even though the concept itself is broader, extending to physical health and a range of public health statistics other than those sampled in this study alone (Leigh et al., 2019). We have also conceptualized job attitudes as predominantly related to work performance, when in fact they overlap with subjective wellbeing, for example through the concept of happiness (Fisher, 2009). We chose the indicators we did because they are linked to workplace performance (Carr, 2022). Yet as Fisher (2009) and Haar et al. (2018) have pointed out, happiness in the work place, for instance job and occupational satisfaction often links to life satisfaction, i.e., happiness in wider society. In that sense, future research should sample different facets of wellbeing, and explore the possibility that occupational subjective wellbeing is a *mediator* between wage and wellbeing.[4]

In conclusion, our data were largely subjective, rather than econometric. The study aimed to explore job attitudes and aspects of subjective wellbeing, both inherently experiential and relevant, during a turbulent and disrupted time in NZ (work) history. From that humanitarian work psychology perspective (Carr et al., 2012), the replicability of previous findings, during a pandemic, points toward a certain robustness in the idea of a living wage being pivotal for eradicating poverty. Beyond that point, our findings have also posed a serious challenge to governments like ours in NZ. The challenge is, to not only keep raising minimum wages to meet living wage levels, but also perhaps to make more of a dynamic systems approach—that simultaneously regulates a suite of policy options in the world of work, housing, and other costs of living. This change should be done in partnership with key stakeholders from labor and management (Arrowsmith et al., 2020). It should somehow manage to include other material cost-of-living factors, like housing rental prices, from the wider NZ societal system.

Practical Implications

1. The government of NZ should keep raising the Minimum wage closer to, and preferably alongside the Living wage.
2. Raises to the Minimum wage will not work unless they are coupled to policy changes to freeze or restrain rents, and to link Minimum wage increases, dynamically, to rising inflation.
3. Evaluations of increases to the Minimum wage should include the health cost savings from boosted wellbeing, subjective, and objective, at work and in society.

[4]We are grateful to one of our expert peer reviewers for this guiding future research suggestion.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Massey University's Northern Human Ethics Committee [Massey University Human Ethics Committee (MUHEC), 2018]. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SC, JH, DH, HJ, JA, JP, AY-H, and SA conducted the study, contributed to the research and the writing of the article, led by SC, JH, and DH, and the article was revised and edited by the other research team members. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by School of Psychology Massey University New Zealand.

ACKNOWLEDGMENTS

The authors wish to thank our constructive and helpful expert peer reviewers, and editorial team, for your considered feedback and suggested improvements to the text, in an earlier version of this paper. This feedback has been adopted almost verbatim in places, and we are grateful for the improvements to the paper that have resulted.

REFERENCES

Armstrong-Stassen, M. (2001). Reactions of older employees to organizational downsizing: The role of gender, job level, and time. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 56, P234–P243. doi: 10.1093/geronb/56.4.P234

Arrowsmith, J., Parker, J., Carr, S., Haar, J., Young-Hauser, A., Hodgetts, D., et al. (2020). "Moving the minimum wage towards a 'living' wage: evidence from New Zealand" in *Labour is not a commodity' Today: The Value of Work and its Rules between Innovation and Tradition*. eds. E. Dagnino, A. Forsyth and M. Roiatti (Cambridge: Cambridge Scholars Publishing), 147–170.

Axtell, C., Wall, T., Stride, C., Pepper, K., Clegg, C., Gardner, P., et al. (2002). Familiarity breeds content: the impact of exposure to change on employee openness and wellbeing. *J. Occup. Organ. Psychol.* 75, 217–231. doi: 10.1348/09631790260098596

Carr, S. C. (2022). *Wage and Wellbeing*. New York: Springer.

Carr, S. C., Haar, J., Hodgetts, D., Arrowsmith, J., Parker, J., Young-Hauser, A., et al. (2019). An employee's living wage and their quality of work life: how important household size and income? *J. Sustain. Res.* 1:e190007. doi: 10.20900/jsr20190007

Carr, S. C., MacLachlan, M., and Furnham, A. (2012). *Humanitarian Work Psychology*. Basingstoke, UK: Palgrave-Macmillan

Carr, S. C., Maleka, M., Meyer, I., Barry, M. L., Harr, J., Parker, J., et al. (2018). How can wages sustain a living? By getting ahead of the curve. *Sustain. Sci.* 13, 901–917. doi: 10.1007/s11625-018-0560-7

Carr, S. C., Parker, J., Arrowsmith, J., and Watters, P. A. (2016). The living wage: theoretical integration and an applied research agenda. *Int. Labour Rev.* 155, 1–24. doi: 10.1111/j.1564-913X.2015.00029.x

Carr, S. C., Young-Hauser, A., Hodgetts, D. J., Schmidt, W., Moran, L., Haar, J., et al. (2021). Research update: how decent wages transform qualities of living. *J. Sustain. Res.* 3:e210012. doi: 10.20900/jsr20210012

Chuluun, T., Graham, C., and Myanganbuu, S. (2016). Who is happy in the land of eternal blue sky? Some insights from a first study of wellbeing in Mongolia. *Int. J. Wellbeing.* 6, 49–70. doi: 10.5502/ijw.v6i3.506

Ferguson, J. L., Ellen, P. S., and Bearden, W. O. (2014). Procedural and distributive fairness: determinants of overall price fairness. *J. Bus. Ethics* 121, 217–231. doi: 10.1007/s10551-013-1694-2

Fisher, C. D. (2009). Happiness at work. *Int. J. Manag. Rev.* 12, 384–412. doi: 10.1111/j.1468-2370.2009.00270.x

Galicki, C. (2020). Impact of COVID-19 on Financial Wellbeing. Auckland, NZ: Commission for Financial Capability.

George, J., and Brief, A. P. (1989). The economic instrumentality of work: An examination of the moderating effects of financial requirements and sex on the pay-life satisfaction relationship. *Acad. Manag. Proc.* 1989, 209–213. doi: 10.5465/ambpp.1989.4980879

Gergen, K. J. (1973). Social psychology as history. *J. Pers. Soc. Psychol.* 26, 309–320. doi: 10.1037/h0034436

Greenhaus, J. H., Parasuraman, S., and Wormley, W. M. (1990). Effects of race on organizational experiences, job performance evaluations, and career outcomes. *Acad. Manag. J.* 33, 64–86.

Haar, J. M. (2013). Testing a new measure of work-life balance: a study of parent and non-parent employees from New Zealand. *Int. J. Hum. Resour. Manag.* 24, 3305–3324. doi: 10.1080/09585192.2013.775175

Haar, J., Carr, S. C., Parker, J., Arrowsmith, J., Hodgetts, D., and Alefaio-Tugia, S. (2018). Escape from working poverty: steps toward sustainable livelihood. *Sustain. For.* 10, 41–44. doi: 10.3390/su10114144

Haar, J. M., Russo, M., Sune, A., and Ollier-Malaterre, A. (2014). Outcomes of work-life balance on job satisfaction, life satisfaction and mental health: a study across seven cultures. *J. Vocat. Behav.* 85, 361–373. doi: 10.1016/j.jvb.2014.08.010

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2010). *Multivariate Data Analysis: A Global Perspective* (*7th Edn*). Upper Saddle River: Pearson Prentice Hall Publishing.

Harrison, D. A., Newman, D. A., and Roth, P. L. (2006). How important are job attitudes? Meta-analytic comparisons of integrative behavioural outcomes and time sequences. *Acad. Manag. J.* 49, 305–325. doi: 10.5465/amj.2006.20786077

Harter, J. K., Schmidt, F. L., and Hayes, T. L. (2002). Business-level relationship between employee satisfaction, employee engagement, and business outcomes: a meta-analysis. *J. Appl. Psychol.* 87, 268–279. doi: 10.1037/0021-9010.87.2.268

Hasdell, R. (2020). *What We Know about Universal Basic Income: A Cross-Synthesis of Reviews*. Stanford, CT: Basic Income Lab.

Hopner, V., Hodgetts, D., Carr, S., Nelson, N., Chamberlain, K., and Ball, R. (2021). "Assembling the psycurity accord in response to the early COVID-19 outbreak in Aotearoa New Zealand" in *The Psychology of Global Crises and Crisis Politics: Intervention, Resistance, Decolonization*. eds. I. Strasser and M. Dege (Cham, Switzerland: Springer), 19–42.

International Labour Organization (ILO) (2019a). Poor working conditions are main global employment challenge. Geneva: ILO.

International Labour Organization (ILO) (2019b). World employment and social outlook—Trends 2019. Available at: https://www.ilo.org/global/research/global-reports/weso/2019/lang--en/index.htm (Accessed April 9, 2020).

International Labour Organization (ILO) (2020). Global Wage Report 2020/21: Wages and Minimum Wages in the Time of COVID-19. Geneva: ILO.

International Labour Organization (ILO) (2021). ILO Standards and COVID-19. Geneva: ILO.

International Labour Organization (ILO) (2022). World Employment and Social Outlook: Trends 2022. Geneva: ILO.

International Trade Union Confederation (ITUC) (2018). Global Poll: Governments' Failure to Address Low Wages and Insecure Jobs Threatens Trust in Politics and Democracy. London: YouGov/ITUC. Available at: https://www.ituc-csi.org/ITUC-Global-Poll-2018 (Accessed April 9, 2020).

Judge, T. A., Bono, J. E., Erez, A., and Locke, E. A. (2005). Core self-evaluations and job and life satisfaction: the role of self-concordances and goal attainment. *J. Appl. Psychol.* 90, 257–268. doi: 10.1037/0021-9010.90.2.257

Judge, T. A., Piccolo, R. F., Podsakoff, N. P., Shaw, J. C., and Rich, B. L. (2010). The relationship between pay and job satisfaction: a meta-analysis. *J. Vocat. Behav.* 77, 157–167. doi: 10.1016/j.jvb.2010.04.002

Kantor, B. (2016). Why a minimum wage will not help SA's poor. Sunday Times, August 9, 2021.

Kaplan, D. M., Berkley, R. A., and Fisher, J. E. (2016). Applicant identity congruence in selection decision making: implications for Alejandro and Consuela. *Hum. Resour. Manag.* 55, 39–51. doi: 10.1002/hrm.21657

Kaufman, J. A., Salas-Hernandez, L. K., Komro, K. A., and Livingston, M. D. (2020). Effects of increased minimum wages by unemployment rate on suicide in the USA. *J. Epidemiol. Community Health* 74, 219–224. doi: 10.1136/jech-2019-212981

Kausto, J., Elo, A. L., Lipponen, J., and Elovainio, M. (2005). Moderating effects of job insecurity in the relationships between procedural justice and employee well-being: gender differences. *Eur. J. Work. Organ.* 14, 431–452. doi: 10.1080/13594320500349813

King, P., and Waldegrave, C.. (2012). Report of an Investigation into Defining a Living Wage for New Zealand. Wellington: The Living Wage Campaign.

King, P., and Waldegrave, C. (2014). Living Wage Aotearoa Movement New Zealand 2014 Update. Wellington: The Living Wage Campaign.

Krieble, T. (2020). Gains for Everyone: Towards an Inclusive Growth Agenda for Aotearoa New Zealand. Wellington, NZ: Helen Clark Foundation and the New Zealand Institute of Economic Research.

Kuvaas, B. (2006). Work performance, affective commitment, and work motivation: the roles of pay administration and pay level. *J. Organ. Behav.* 27, 365–385. doi: 10.1002/job.377

Law, D. (2020). *Minimum Wages to the Maximum: The Risks of Lifting the Minimum Wage*. Wellington, NZ: Policy Point.

Lee, K., and Allen, N. J. (2002). Organizational citizenship behaviour and workplace deviance: the role of affect and cognitions. *J. Appl. Psychol.* 87, 131–142. doi: 10.1037/0021-9010.87.1.131

Leigh, J. P., Leigh, W. A., and Du, J. (2019). Minimum wages and public health: A literature review. *Prev. Med.* 118, 122–134. doi: 10.1016/j.ypmed.2018.10.005

Lenhart, O. (2017). The impact of minimum wages on population health: evidence from 24 OECD countries. *Eur. J. Public Health* 18, 1031–1039. doi: 10.1007/s10198-016-0847-5

Massey University Human Ethics Committee (MUHEC) (2018). Living wages: Transforming lives, transforming work? Auckland (New Zealand): Massey University, Reference NOR 18/04.

Meyer, J. P., Allen, N. J., and Smith, C. A. (1993). Commitment to organizations and occupations: extension and test of a three-component conceptualization. *J. Appl. Psychol.* 78, 538–551. doi: 10.1037/0021-9010.78.4.538

New Zealand Treasury (2020). *Wellbeing budget 2020: Rebuilding together*. Wellington: NZ Treasury

Schaufeli, W. B., Salanova, M., Gonzalez-Roma, V., and Bakker, A. B. (2001). The measurement of engagement and burnout: a two sample confirmatory factor analytic approach. *J. Happiness Stud.* 3, 71–92. doi: 10.1023/A:1015630930326

Schupbach, J. N. (1998). *Robustness, Diversity of Evidence, and Probabilistic Independence*. New York: Routledge.

Shaw, J. D., and Gupta, N. (2001). Pay fairness and employee outcomes: exacerbation and attenuation effects of financial need. *J. Occup. Organ. Psychol.* 74, 299–320. doi: 10.1348/096317901167370

Smith, L. (2015). Reforming the minimum wage. *Am. Psychol.* 70, 557–565. doi: 10.1037/a0039579

Spreitzer, G. M. (1995). Psychological empowerment in the workplace: dimensions, measurement, and validation. *Acad. Manag. Rev.* 38, 1442–1465.

Stanton, J. M., Balzer, W. K., Smith, P. C., Parra, L. F., and Ironson, G. (2001). A general measure of work stress: the stress in general scale. *Educ. Psychol. Meas.* 61, 866–868. doi: 10.1177/00131640121971455

Statista (2021). Statista. Available at: https://www.statista.com/statistics/375265/inflation-rate-in-new-zealand/ (Accessed March 4, 2021).

Stride, C., Wall, T.D., and Catley, N. (2007). *Measures of Job Satisfaction, Organisational Commitment, Mental Health and Job-Related Wellbeing: A Benchmarking Manual*. Chichester, UK: John Wiley & Sons.

Tomlyn, A. J., and Cummins, R. A. (2011). The subjective wellbeing of high-school students: validating the personal wellbeing index—school children. *Soc. Indic. Res.* 101, 405–418. doi: 10.1007/s11205-010-9668-6

Tsao, T. Y., Konty, K. J., Van Wye, G., Barbot, O., Hadler, J. L., Linos, N., et al. (2016). Estimating potential reductions in premature mortality in New York City from raising the Minimum Wage to $15. *Am. J. Public Health* 106, 1036–1041. doi: 10.2105/AJPH.2016.303188

United Nations (2021). Sustainable Development Knowledge Platform. New York, United Nations.

Vitell, S. J., King, R. A., Howie, K., Toti, J. F., Albert, L., Hidalgo, E. R., et al. (2016). Spirituality, moral identity, and consumer ethics: a multi-cultural study. *J. Bus. Ethics* 139, 147–160. doi: 10.1007/s10551-015-2626-0

Waldegrave, C. (2020). Why a real living wage should be the minimum. The New Zealand Herald, December 23, 2022.

Williams, L. J., Vandenberg, R. J., and Edwards, J. R. (2009). Structural equation modelling in management research: a guide for improved analysis. *Acad. Manag. Ann.* 3, 543–604. doi: 10.5465/19416520903065683

Wood, R. C. (1997). Rhetoric, reason and rationality: the national minimum wage debate and the UK hospitality industry. *Int. J. Hosp. Manag.* 16, 329–344. doi: 10.1016/S0278-4319(97)00029-7

Young, L., Milner, M., Edmunds, D., Pentsil, G., and Broman, M. (2014). The tenuous relationship between salary and satisfaction. *J. Behav. Stud.* 7, 1–9.

Conflict of Interest: 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 | Frontiers in Psychology

SYSTEMATIC REVIEW
published: 24 May 2022
doi: 10.3389/fpsyg.2022.884756

Emotional Self-Regulation in Everyday Life: A Systematic Review

Marina Alarcón-Espinoza[1], Susana Sanduvete-Chaves[2], M. Teresa Anguera[3], Paula Samper García[4] and Salvador Chacón-Moscoso[2,5]*

[1] Departamento de Psicología, Universidad de La Frontera, Temuco, Chile, [2] Departamento de Psicología Experimental, Universidad de Sevilla, Seville, Spain, [3] Faculty of Psychology, Institute of Neurosciences, University of Barcelona, Barcelona, Spain, [4] Department of Basic Psychology, University of Valencia, Valencia, Spain, [5] Departamento de Psicología, Universidad Autónoma de Chile, Santiago, Chile

Emotional self-regulation in childhood and adolescence constitutes a growing interest in the scientific community, highlighting in recent years the need to observe its development in their daily life. Therefore, the objective of this systematic review is to characterize publications referring to the development of emotional self-regulation of people under 18 years-old, in natural contexts. Based on the PRISMA guidelines, searches are carried out in the Web of Science, Scopus and PsycINFO databases, and in Google Scholar until May 2020. After reviewing the full text of 376 publications, 14 works are selected that are observed in their extrinsic, substantive and methodological characteristics based on the GREOM and MQCOM guidelines, by two independent evaluators. Most of the studies correspond to the last 20 years, increasing the interest in observing older children, in interaction with adults and/or in different cultures. They apply mixed methodologies, not always ascribing to a low intensity design. Strengths are observed regarding the collection and analysis of the quality of the data; and weaknesses related to the failure to record the duration and sequence of behaviors, highlighting the use of guidelines as guides for future research.

Keywords: evolutionary development, emotional regulation, observational methodology, natural contexts, childhood, adolescence

OPEN ACCESS

Edited by:
Antonio Zuffiano, Sapienza University of Rome, Italy

Reviewed by:
Cristina Senín-Calderón, University of Cádiz, Spain
África Borges, University of La Laguna, Spain

****Correspondence:***
Marina Alarcón-Espinoza
marina.alarcon@ufrontera.cl

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *27 February 2022*
Accepted: *25 April 2022*
Published: *24 May 2022*

Citation:
Alarcón-Espinoza M, Sanduvete-Chaves S, Anguera MT, Samper García P and Chacón-Moscoso S (2022) Emotional Self-Regulation in Everyday Life: A Systematic Review. Front. Psychol. 13:884756. doi: 10.3389/fpsyg.2022.884756

INTRODUCTION

Emotional self-regulation, referring to the understanding, acceptance, and modulation of emotional responses, is a process that children and adolescents carry out in order to adapt to their psychosocial environment, orienting themselves toward the achievement of their evolutionary goals and favoring their mental health (Van Lissa et al., 2019). The achievement of emotional self-regulation allows progress in the acquisition of greater autonomy, at the same time that it is related to the development of adequate self-esteem and feelings of self-efficacy that facilitate social and school adjustment. The emotional educational process is continuous and permanent throughout the life cycle (Gallardo Fernández and Saiz Fernández, 2016), favoring the individual in order to achieve emotional competence which allows to regulate their emotions.

Much of the research on emotional development has focused on the relationships between parents and children under the age of two and/or preschoolers. However, in the last decade there have been studies referring to understanding how families socialize the expression of emotions in their children's middle childhood and adolescence (Adrian et al., 2011; Bai et al., 2016), observing that the emotional regulation level that children reach at age 7 predicts the quality of positive

friendship at age 10, showing greater ability to express their emotions effectively, interpret emotions and respond to them appropriately (Blair et al., 2014).

Gallardo Fernández and Saiz Fernández (2016) have highlighted that the 21st century school has to assume responsibility for educating children's emotions as much or more than the family, highlighting that educators must be the main emotional leaders of the students. Furthermore, Bailey et al. (2016) highlight that children marked by effective interactions with their teachers have better socio-emotional and cognitive skills, highlighting that effective teachers can help children in the transition toward self-regulation of their emotions; and that the emotional and organizational support of the educational context can be particularly sensitive to the social-emotional functioning of children in the classroom.

Meanwhile, in the field of research, Sabatier et al. (2017) point out that, in the last 15 years of research in the field of emotional development, the findings regarding neurobiological and environmental elements that influence the acquisition of skills to manage emotions have been highlighted, with consensus that, with age, people improve in the control of their emotions. However, fewer studies were observed that analyze these regulatory processes during the adolescent period and many of these studies would correspond to western and developed countries. Also, there is a need to document the development of emotional regulation processes in different social and economic contexts.

Compas et al. (2017), based on a meta-analytic review, propose an agenda for future research that includes improving the conceptualization seeking integration between the various constructs that study the subject; prioritize the study of the development of emotional regulation capacities instead of the study of symptoms, and improve the methodology and research designs by approaching more ecological models that allow understanding these processes in real contexts and times.

Likewise, Adrian et al. (2011) affirm that the empirical evidence indicates that emotional regulation skills are developed in a dynamic and multifaceted system, observing that, although observational and longitudinal methodologies have been mostly used with children under 6 years of age, it is necessary to continue carrying out multimodal evaluations and research with multiple methods and multilevel assessments in school-age children and adolescents.

Along the same lines, Buckley et al. (2003), emphasize the need to research about the emotional development of children and adolescents, in a collaborative way with school personnel, thus being able to observe how they use different coping strategies in natural development contexts, an aspect that is also mentioned by Bai et al. (2016), who have also emphasized the need to know how children's spontaneous emotional expressions develop and maintain in uncontrolled environments of daily life, particularly within the family and during the school-age years.

Therefore, understanding natural contexts as all those contexts in which the behavior is habitual, and is not constrained by requirements that alter spontaneity (Craik, 2000; Bolger et al., 2003; Wilhelm et al., 2012), the objective of this systematic review has been to characterize publications referring to the development of emotional self-regulation in people under the age of 18 years, through relationship/communication guidelines, in natural contexts.

METHODS

Bibliographic searches were carried out in the Web of Science, Scopus and PsycINFO databases and in academic Google from its inception until May 2020 with the following keywords in title, keywords and/or abstract: ("emotional autoregulation") OR ("autorregulación emocional") OR ("emotional self-regulation") OR ("emotional selfregulation") OR ("emotional self regulation") OR ("competencias emocionales") OR ("emotional skills") OR ("emotional competences") OR ("regulación emocional") OR ("emotional regulation") OR ("educación emocional") OR ("emotional education") AND ("comunicación") OR ("communication") AND ("relaciones interpersonales") OR ("relationships") AND ("vida cotidiana") OR ("daily life").

As inclusion criteria of the studies selected to respond to the objective of the present investigation, the following were considered: (a) that their objective was to investigate self-regulation/emotional regulation; (b) primary studies, excluding theoretical works, systematic reviews, and meta-analysis; (c) that they observe daily relationship/communication patterns in natural contexts; (d) that the main participants were people under 18 years of age (regardless of whether parents or teachers were also involved) (e) who studied universal population (normal evolutionary development); (f) written in English or Spanish; (g) with access to the full text. For study selection, two investigators applied the criteria independently. Subsequently, intercoder reliability was calculated using the kappa coefficient (κ). Agreement was reached on the discrepancies found with the mediation of a third researcher.

Additionally, in order to expand the number of primary studies included, the references of the included texts were reviewed and the authors were written to in order to request new articles that could meet the criteria indicated.

The included works were reviewed in order to observe: (1) extrinsic characteristics: their institutional affiliation, type and year of publication, and country where the research was carried out; (2) substantive characteristics: referring to the characteristics of the sample, and the way to conceptualize, base and evaluate self-regulation/emotional regulation; and (3) methodological characteristics: recording the characteristics of the method explicitly declared by the authors; those observed according to *Guidelines Reporting Evaluations based on Observational Methodology GREOM* (Portell et al., 2015); and those observed according to *Methodological Quality Checklist for studies based on Observational Methodology (MQCOM*; Chacón-Moscoso et al., 2019).

The review of the articles was carried out independently by two researchers, who when applying the MQCOM guide, had to agree on their observations in the face of the discrepancies found with the arbitration of a third expert researcher. The degree of initial agreement was calculated with the coefficient κ.

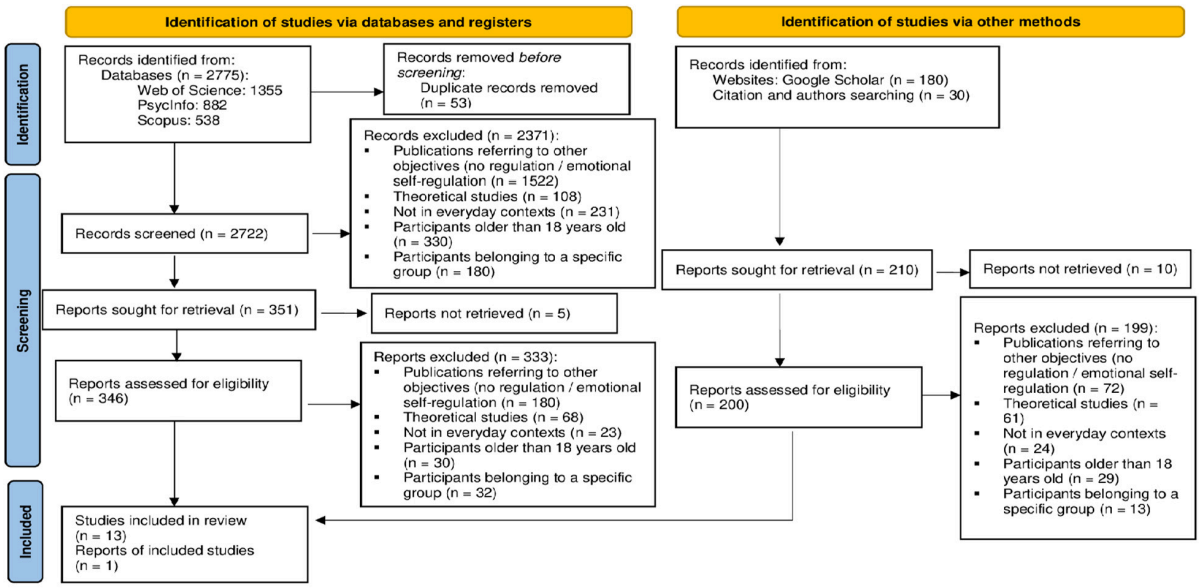


FIGURE 1 | Study selection flow chart.

TABLE 1 | Extrinsic characteristics.

References	Institutional affiliation	Publication type	Country of the experience
Bai et al. (2016)	University of California, Los Angeles; McLean Hospital, Belmont, Massachusetts and Harvard Medical School	Article	United States
Bailey et al. (2016)	Yale University; George Mason University	Article	United States
Fabes et al. (1999)	Department of Family Resources & Human Development, Arizona State University; Arizona State University	Article	United States
Feldman and Klein (2003)	Department of Psychology, Bar-Ilan University, Israel	Article	Israel
Gallegos et al. (2017)	University of Texas at Austin	Article	United States
Garner (2006)	New Century College, George Mason University, 4400 University Drive	Article	United States
Herndon et al. (2013)	George Mason University	Article	United States
Keller et al. (2004)	University of Osnabrück, Germany; Henning Jensen, University of Costa Rica, San Jose; Aristotle University of Thessaloniki, Greece	Article	Cameroon, Greece and Costa Rica
Kim et al. (2014)	The Pennsylvania State University	Article	United States
Kochanska et al. (2007)	University of Iowa	Article	United States
Lavelli et al. (2019)	University of Verona; University of Milano-Bicocca; University of Osnabrück, Germany and Hebrew University of Jerusalem	Article	Italy, Nigeria, Ghana and Cameroon
Roberts (2020)	York University, Toronto, Ontario, Canada	Article	Canada
Santa Cruz (2020)	Universidad Cesar Vallejo, Trujillo, Peru	Bachelor thesis	Peru
Silva et al. (2018)	Universidade do Minho, Portugal	Article	Portugal

RESULTS

Selection of Studies

Figure 1 presents the PRISMA flow chart (Page et al., 2021), with the selection process of the primary studies in the systematic review. The intercoder reliability in the selection obtained a $\kappa = 0.73$. Finally, 14 studies met the inclusion criteria.

Characteristics of the Studies

In the coded variables, the intercoder reliability reached $\kappa = 0.81$.

Regarding the extrinsic characteristics of the articles studied (see **Table 1**), there is a progressive increase in publications with the mentioned inclusion criteria, the first of which was observed in 1999. The institutional affiliation of the researchers corresponds mainly to North American universities (8), observing two works by George Mason University. Of the other works, two publications report the joint effort of researchers from different universities, both of whom are the same researcher from the University of Osnabrück in Germany.

Regarding the substantive characteristics of the reviewed papers (see **Table 2**), most of the articles look at preschool-age children. Regarding the ethnicity variable, it is observed that 8 of

TABLE 2 | Substantive characteristics.

Publication	Sample size			Age	Gender	Race or Nationality
	Under 18 years	Fathers and/or mothers	Teachers			
Bai et al. (2016)	31	29 mothers and 31 fathers	–	8–12 years Fathers M:41,5 (SD:5,6)	14 girls and 17 boys	European Americans (64.5%), Mestizos (19.4%), Asian Americans (9.7%), Latino (3.2%), African American (3.2%)
Bailey et al. (2016)	312	–	44	3–5 years old	155 boys and 157 girls	52% White, 31% African American, and 2% Asian, Native American, or Pacific Islander; 10% of parents did not report ethnicity
Fabes et al. (1999)	135	–	–	4–5 years old	77 boys, 58 girls	87% Caucasian; 7% Mexican American, 4% African American, 2% Asian
Feldman and Klein (2003)	90	90 mothers and 42 fathers	16	2 years	52 boys and 38 girls	Israelis
Gallegos et al. (2017)	125	125 mothers and 125 fathers	–	Last trimester of pregnancy, 8 and 24 months of children.	74 boys and 51 girls	84% White, 8% Hispanic, 2% African American, 6% biracial or other ethnicity
Garner (2006)	70	70 mothers	–	3–6 years old	52% male	African American
Herndon et al. (2013)	308	–	–	Children 3–5 years old	51.0% male children	57% Caucasian, with 33.6% from African American families. 15% Latino/Hispanic, 6.4% not reporting
Keller et al. (2004)	116	116 families	–	3–20 months	–	Cameroon, Greece and Costa Rica
Kim et al. (2014)	132	132 mothers	–	12 and 18 months	78 (54.2%) women and 54 (37.5%) men	86.1% White, 3.5% African American, 2.1% Asian, 4.9% Latino, and 2.8% Other
Kochanska et al. (2007)	112	112 fathers	–	9 months to 6 years	52 girls	97% White
Lavelli et al. (2019)	60	60 mothers	–	Babies and mothers	Italian: 50% girls; immigrants: 50% girls; Cameroon: 60% girls.	20 Italian mother and child dyads, 20 first-generation West African immigrant mothers and their Italian-born babies, and 20 Cameroonian dyads
Roberts (2020)	33	10 mothers and 26 fathers	–	Families with children from 4 to 6 years old	16 girls and 17 boys	85% Canadian English
Santa Cruz (2020)	73	–	–	3 years old	–	–
Silva et al. (2018)	33	33 families	–	12–18 years old	21 girls (63%)	97% Portuguese, 3% Brazilian

the studies make explicit mention of the race of the participants and that none of them refer to the comparison between different cultures or countries. Studies interested in observing cultural variables indicate the nationality and/or immigration status of the participants.

Regarding the objectives of interest of the reviewed works (see **Table 3**), the reference frameworks used to support the study of emotional regulation come from the last years of the 20th century. The authors, when proposing the objectives of their research, indicate more than one motivation, highlighting the interest in research regarding evolutionary development and interaction/communication within the family.

Regarding the observed methodological characteristics (**Table 4**), in three of the publications the authors describe their work as observational – naturalistic, five studies claim to be longitudinal and six are defined as descriptive. Four of the studies propose observation in the natural context as the only form of evaluation; while the remaining investigations indicate this modality among other possibilities, such as the completion of questionnaires or tasks designed to provoke certain emotions or behaviors. The instruments reported to observe self-regulation/emotional regulation are mostly *ad hoc* observation instruments. The dimensions that the authors are interested in observing refer mainly to the interaction between children and adults.

Regarding the methodological characteristics observed by GREOM –first part– (**Table 5**), five publications justify the choice of a low intensity observation method. Regarding the study units, four publications indicate and apply inclusion criteria. Regarding the observed sessions, eight articles indicate the period of time in which it has been observed, seven specify the number of observation sessions carried out, seven publications mention the period of time elapsed between the observations, and eleven inform the method used for sampling. All publications describe the observation instrument used, six justify it, ten provide access to the instrument and one provides access to the coding manual.

Regarding the primary recording parameters, all of them record frequency (GREOM second part, see **Table 6**), five record the duration of the behavior and three mention the behavioral

TABLE 3 | Specific objectives of interest.

References	Authors cited when defining Emotional Self-Regulation	Objective associated with observing self-regulation or emotional regulation									Self-regulation/emotional regulation assessment					
		Evolutionary development	Academic performance	Discipline or school adjustment	Family interaction/ communication	Ethical-moral/ civic development	Teaching methodologies	Cultural differences	Social development/ prosociality	Prevention mental health difficulties	Instrument with adequate psychometric characteristics	Consider evaluator training	Design tasks to evaluate	Use various evaluation methods	Collect information through Different contexts or informants	Analyze information through inter-judge agreement
Bai et al. (2016)	–Gross				X							X				X
Bailey et al. (2016)	–Pianta –Thompson			X								X	X	X	X	X
Fabes et al. (1999)	–Eisenberg and Fabes –Cummings and Cummings	X							X					X	X	X
Feldman and Klein (2003)	–Vygotsky, Feldman, Greenbaum and Yirmiya	X			X							X	X	X	X	X
Gallegos et al. (2017)	–Morris, Silk, Steinberg, Myers and Robinson	X			X							X	X	X	X	X
Garner (2006)	–Thompson	X			X	X		X	X			X			X	X
Herndon et al. (2013)	–Cole, Michel and Teti –Denham, Zinsser and Brown	X		X							X	X		X		
Keller et al. (2004)	–Kopp	X						X				X	X	X	X	X
Kim et al. (2014)	–Calkins and Leerkes –Kopp –Thompson	X			X					X			X	X	X	X
Kochanska et al. (2007)	–Kopp –Caspi and Shiner	X			X						X	X	X	X	X	X
Lavelli et al. (2019)	–Fogel –Camras, Shuster and Fraumeni	X			X										X	X
Roberts (2020)	–Kopp –Thompson	X			X				X			X		X	X	X
Santa Cruz (2020)	–Goleman –Perpiñán	X		X					X		X					X
Silva et al. (2018)	–Gross –Gilbert	X										X		X	X	

TABLE 4 | Methodological characteristics declared by the authors.

References	Design	Data collection technique or instrument	Observed dimensions
Bai et al. (2016)	Naturalistic observational	Video recordings taken in homes and community settings	Mutual display of positive emotion, touch and joint leisure
Bailey et al. (2016)	Descriptive	Classroom observations and teacher questionnaires: Preschool Learning Behaviors Scale (PLBS); Social Competence and Bahavior Evaluation (SCBE-30); Teaching Rating Scale of School Adjustment (TRSSA); Student - Teacher Relationship Scale (STRS) The Preschool Self-Regulation Assessment (PSRA); PSRA-AR; CLASS	School Adjustment, Executive Control, Emotional Regulation, Emotional and Organizational Support
Fabes et al. (1999)	Descriptive	Observations of infantile behaviors. Teacher questionnaires (CBQ). Social competence teacher qualifications: Scale of Perceived Social Competence for Children.	Intensity of peer interaction Negative Emotions Positive or constructive social interactions
Feldman and Klein (2003)	Observational naturalistic and Descriptive	Observation of the relational style of the adult: Coding Interactive Behavior (CIB). Observation of Mediational Interaction (OMI). Interviews with caregivers. Observation of child compliance and adult discipline: The Observer. Self-regulation observation of childhood emotion (cognitive tests). Quality observation of the caregiver's relational style in the group.	Child Compliance Self-regulatory compliance of mothers and caregivers Childhood cognition and the regulation of emotions
Gallegos et al. (2017)	Longitudinal	Video recordings of couple's discussion tasks. Observations of video interactions in games and routines: child care scales (ICS), classification method by criteria, emotional abstinence scale. Observation of co-parenting conflict (CFRS), verbal sparring scale (ICC = 0.74), a measure of co-parenting conflict Observation of the emotional regulation of children	Negative marital affect observed before birth Parental emotional withdrawal Coparenting conflict Child regulation
Garner (2006)	Naturalistic observational	Observations in home visits Observations in visits to preschool centers	Prosocial socialization behaviors Socialization behaviors of emotions Peer episodes that caused emotion dysregulation
Herndon et al. (2013)	Descriptive	Observation of socio-emotional behavior: Minnesota Preschool Affection Checklist (MPAC-R/S). Teacher Qualifications: SCBE-30; PLBS; STRS; TRSSA	Child socio-emotional behaviors Attitudes toward school Positive relationships among teachers Cooperative participation
Keller et al. (2004)	Longitudinal	Observation of breeding systems Self recognition: blush test. Observation of self-regulation: Fulfillment of requests and Fulfillment of the prohibition	Breeding systems Auto-recognition Self-regulation
Kim et al. (2014)	Longitudinal	Observation of emotional regulation Video observation at bedtime: Emotional availability scales (EAS). Child attachment security: Strange situation. Childish temperament: Revised Infant Behavior Questionnaire (IBQ-R) and Early Infant Behavior Questionnaire (ECBQ)	Mothers' emotional availability Child attachment security Emotional regulation strategies
Kochanska et al. (2007)	Longitudinal	Observation of mother-child dyad Positive Emotionality Laboratory Procedures for Children Children's Intellectual Functioning: WPPSI-R information scale Children's impulsiveness: CBQ	Positive emotionality Self-regulation Intellectual functioning of the child
Lavelli et al. (2019)	Longitudinal	Observation of emotional expression in social interaction. Observation of maternal behaviors.	Active child care Maternal gaze and facial behaviors
Roberts (2020)	Descriptive	Parenting Practices Q-sort Parenting Questionnaire Observation of interactions between peers and friendship networks. Questionnaire for teachers and observers Q-Sort. Observation at home: Individual focal samples	Episodes of childhood distress Parents' approach to distress Children's social competence
Santa Cruz (2020)	Descriptive	Checklist of children's behaviors in free play and emotional self-regulation	Self-regulation of emotions in students
Silva et al. (2018)	Descriptive	Daily life reports. Perceptions questionnaire on sampling week Emotional regulation questionnaire: ERQ-CA. Positive and negative affect status questionnaire Self-observation of emotional regulation strategies	Emotional regulation strategies

sequences. The information is recorded mainly through videos and observed by trained personnel. In relation to data quality control, thirteen studies report concordance analysis of the collected data. Regarding the analysis of the data carried out, all of them made explicit the type of analysis used and thirteen of them justified it.

Regarding the methodological characteristics observed through MQCOM (see **Table 7**), seven studies justify and support the observation methodology used based on the degree of perceptiveness of the information. In one of the investigations, software is used to record, control, and analyze the quality of the data, and in four investigations, the use of this tool

TABLE 5 | Methodological characteristics observed through the Guidelines for Reporting Evaluations based on Observational Methodology -GREOM- (first part).

References	Observation method justification	Description of expected results	Design description			Inclusion criteria indicated and applied	Times		Contexts			Observation instrument						
			Informs observational design	Justify observational design	Participants you want to observe	Sequential data	Observation of common contexts	Specify the observation period	Specify number of observation sessions	Specifies the periodicity between observation sessions	Specify method used for sampling	Indicate WHAT is observed	Indicate WHO is being observed	Indicate CIRCUMSTANCES observed	Describe the observation instrument	Justify the observation instrument used	Provide access to the observation instrument used	Provide access to the encoding manual used	
Bai et al. (2016)	X	X		X	Several	X	X	X	X	X	X	X	X	X	X	X	X	X	
Bailey et al. (2016)		X		Partial	Several	X	X		X	X			X	X	X	X		X	
Fabes et al. (1999)	X	X		Partial	Several	X	X		X	X	X	X	X	X	X	X			
Feldman and Klein (2003)		X		Partial	Several		X					X	X	X	X	X			
Gallegos et al. (2017)		X		Partial	Several		X						X	X	X	X		X	
Garner (2006)		X		Partial	Several	X	X	X	X		X	X	X	X	X	X			
Herndon et al. (2013)		X		Partial	Several		X		X	X	X	X	X	X	X	X	X	X	
Keller et al. (2004)		X		Partial	Several		X	X	X	X		X	X	X	X	X	X	X	
Kim et al. (2014)	X	X		Partial	Several		X					X	X	X	X	X		X	
Kochanska et al. (2007)		X		Partial	Several		X			X	X	X	X	X	X	X			
Lavelli et al. (2019)		X		Partial	Several	X	X	X	X		X	X	X	X	X	X	X	X	X
Roberts (2020)	X	X		Partial	Several		X					X	X	X	X	X		X	
Santa Cruz (2020)		X		Partial	Several		X						X	X	X	X	X	X	
Silva et al. (2018)	X	X		Partial	Several	X	X		X	X	X	X	X	X	X	X	X	X	

It is marked with an X when the criterion is met.

TABLE 6 | Methodological characteristics observed through Guidelines for Reporting Evaluations based on Observational Methodology -GREOM- (second part).

References	Primary recording parameters			Means of observation	Session acceptance criteria		Observer characteristics					Reliability	Flow of study units		Analysis	
	Frequency	Duration	Sequence		Justification of consistency between sessions	Justification of interruptions of the sessions	The observer is a close person	The observer has been trained	The observer is being evaluated	The observer receives a payment	Self-report		Report observation interruptions	Report withdrawals from participants	Data analysis used	Justify data analysis modality
Bai et al. (2016)	X	X		Video	X	X		X				X	X	X	X	X
Bailey et al. (2016)	X			Video				X				X			X	X
Fabes et al. (1999)	X			Video	X			X				X	X		X	X
Feldman and Klein (2003)	X	X		Video				X				X			X	X
Gallegos et al. (2017)	X			Video				X				X		X	X	X
Garner (2006)	X		X	Audio and pencil and paper				X				X		X	X	X
Herndon et al. (2013)	X			Pencil and paper				X				X			X	X
Keller et al. (2004)	X			Video				X				X	X	X	X	X
Kim et al. (2014)	X			Video				X				X		X	X	
Kochanska et al. (2007)	X	X		Video				X				X		X	X	X
Lavelli et al. (2019)	X	X	X	Video	X		X	X				X			X	X
Roberts (2020)	X	X		Video				X				X			X	X
Santa Cruz (2020)	X			Pencil and paper								X			X	X
Silva et al. (2018)	X		X	Mobile							X				X	X

It is marked with an X when the criterion is met.

TABLE 7 | Methodological characteristics observed using the Methodological Quality Checklist for studies based on Observational Methodology (MQCOM).

References	Reference to the observation methodology	Delimitation of the study objectives	Referenced theoretical framework	Observation unit criteria	Temporal criteria	Dimensionality criteria	Inclusion/exclusion criteria	Adequacy of the observation instrument	Coding manual	Software usage	Data type specification	Parameters specification	Session delimitation	Inter-observer reliability	Type of data analysis	Interpretation of results in the discussion
Bai et al. (2016)	0.5	1	1	1	1	1	1	1	1	0.5	–	0.5	1	1	1	1
Bailey et al. (2016)	1	0.5	1	1	0.5	1	1	1	1	0.5	–	0.5	0.5	1	1	1
Fabes et al. (1999)	0.5	1	1	1	0.5	1	0.5	1	0.5	0	–	0.5	1	1	1	1
Feldman and Klein (2003)	0.5	1	1	1	0.5	1	0.5	1	1	0	–	0.5	1	1	1	1
Gallegos et al. (2017)	0.5	1	1	1	1	1	1	1	0	0	–	0.5	1	1	1	1
Garner (2006)	0	1	1	1	0.5	1	1	1	1	0	1	0.5	0.5	1	1	1
Herndon et al. (2013)	0.5	0.5	1	0.5	0.5	1	0.5	1	1	0	–	0	0.5	1	1	1
Keller et al. (2004)	1	1	1	1	1	1	1	1	1	0	–	0.5	1	1	1	1
Kim et al. (2014)	0.5	1	1	1	1	1	1	1	0.5	0.5	–	0.5	1	1	1	1
Kochanska et al. (2007)	1	1	1	1	1	1	1	1	0.5	0	–	0.5	1	1	1	1
Lavelli et al. (2019)	1	1	1	1	0.5	1	1	1	1	1	1	1	0.5	1	1	1
Roberts (2020)	1	1	1	1	0.5	1	0.5	1	1	0.5	–	0.5	0.5	1	1	1
Santa Cruz (2020)	1	1	1	1	0.5	1	1	1	0	0	–	0	0.5	1	1	0.5
Silva et al. (2018)	1	1	1	0.5	0.5	1	1	1	0	0	0.5	1	1	0	1	1

1 = meets the criteria; 0.5 = partially complies; 0 = does not comply.

was partial. Regarding the type of parameters recorded, in 10 studies the secondary record derived from the recording of a single category (for example: frequency or duration) was observed, in two studies the primary record of a single category was observed, and in the other two investigations the dynamic or transition recording between different observation parameters was used.

All the investigations indicate having carried out some inferential analysis to analyze the data. In 13 publications, the results are interpreted based on the objectives of the study and the scientific literature, while in the other study the results are interpreted based solely on the objectives of the study.

DISCUSSION

In the last 30 years, there has been a growing interest in the study of emotional self-regulation in older children and in different contexts and cultures, as suggested by Adrian et al. (2011), Bai et al. (2016), and Chervonsky and Hunt (2019). Regarding the substantive characteristics of the works, it is observed that a large part of the studies consider adults linked to children or adolescents as participants, showing a greater interest in observing the interaction in emotional regulation processes.

The countries and universities that lead the research carried out, as pointed out by Sabatier et al. (2017), correspond mostly to territories with higher income and quality of life, many of which have some tradition in studies of the evolutionary development of children. The influence of North American authors such as Claire Koop, Ross Thompson, James Gross, Susan Calkins, Pamela Cole or Nancy Eisenberg is observed, which could be related to having considered only works written in English and Spanish, suggesting that future studies incorporate written works in other languages.

Regarding the methodological characteristics, as strengths it was observed that most of the studies use different techniques or instruments for data collection; that the instruments designed *ad hoc* have a theoretical basis, are applied by properly trained personnel and have data quality control. In all the primary documents, situations typical of daily life are studied and analyzed, observing in all of them the use of the observational methodology, although there are some variants and diverse denominations, for which it is estimated as a weakness, that more than half of the studies do not propose the choice of a low intensity methodological design, with which they do not necessarily consider the richness involved in observing behaviors of daily life and detailing observation parameters such as duration and sequence of behaviors, aspects that are deemed necessary to observe in future research. In this sense, considering that the observational methodology constitutes a contribution to studies referring to evolutionary development in daily life, there is a need to highlight, in the preparation of future research, the review of the guidelines proposed by Chacón-Moscoso et al. (2019) and Portell et al. (2015), in order to guarantee the methodological quality.

Finally, observing that the study of behaviors in daily life has been gaining space and value when questioning the impact of studies carried out in laboratories (Compas et al., 2017), it is observed that, although every day there are older and better technological instruments that allow observing daily life and with people who are willing to comment on their experiences, it is necessary to regulate the ethical scope of the use of social

networks in research, since they could affect the private and public life of the participants.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

MA-E and MTA: idea. MA-E and SS-C: literature review (state of the art). MA-E, SS-C, and SC-M: methodology, data analysis, and results. MA-E, MTA, PS, SS-C, and SC-M: discussion and conclusions. MA-E, PS, and MTA: newsroom (original draft). MTA, SS-C, and SC-M: final revisions. MTA and SC-M: project design and sponsorships. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

This work was made possible thanks to the National Research and Development Agency (ANID)/Scholarship Program DOCTORAL SCHOLARSHIPS CHILE/2016 – 72180000 to (MA-E). The collaboration of academics from the University of Seville was made possible thanks to the National Fund for Scientific and Technological Development Regular FONDECYT, ANID, Government of Chile (ref. number 1190945); the PID2020-115486GB-I00 grant funded by MCIN/AEI/ 10.13039/501100011033; and the Andalusian ERDF Operational Program 2014–2020, Government of Andalusia, Spain (ref. US-1263096). We thank Rafael Ricardo Verdugo Mora, who collaborated in the coding of articles.

REFERENCES

Adrian, M., Zeman, J., and Veits, G. (2011). Methodological implications of the affect revolution: a 35-year review of emotion regulation assessment in children. *J. Exp. Child Psychol.* 110, 171–197. doi: 10.1016/j.jecp.2011.03.009

Bai, S., Repetti, R. L., and Sperling, J. B. (2016). Children's expressions of positive emotion are sustained by smiling, touching, and playing with parents and siblings: a naturalistic observational study of family life. *Dev. Psychol.* 52, 88–101. doi: 10.1037/a0039854

Bailey, C. S., Denham, S. A., Curby, T. W., and Bassett, H. H. (2016). Emotional and organizational supports for preschoolers' emotion regulation: relations with school adjustment. *Emotion* 16, 263–279. doi: 10.1037/a0039772

Blair, B. L., Perry, N. B., O'Brien, M., Calkins, S. D., Keane, S. P., and Shanahan, L. (2014). The indirect effects of maternal emotion socialization on friendship quality in middle childhood. *Dev. Psychol.* 50, 566–576. doi: 10.1037/a0033532

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

Buckley, M., Storino, M., and Saarni, C. (2003). Promoting emotional competence in children and adolescents: implications for school psychologists. *School Psychol. Q.* 18, 177–191. doi: 10.1521/scpq.18.2.177.21855

Chacón-Moscoso, S., Anguera, M. T., Sanduvete-Chaves, S., Losada-López, J. L., Lozano-Lozano, J. A., and Portell, M. (2019). Methodological quality checklist for studies based on observational methodology (MQCOM). *Psicothema* 31, 458–464. doi: 10.7334/psicothema2019.116

Chervonsky, E., and Hunt, C. (2019). Emotion regulation, mental health, and social wellbeing in a young adolescent sample: a concurrent and longitudinal investigation. *Emotion* 19, 270–282. doi: 10.1037/emo0000432

Compas, B. E., Jaser, S. S., Bettis, A. H., Watson, K. H., Gruhn, M. A., Dunbar, J. P., et al. (2017). Coping, emotion regulation, and psychopathology in childhood and adolescence: a meta-analysis and narrative review. *Psychol. Bull.* 143, 939–991. doi: 10.1037/bul0000110

Craik, K. H. (2000). "The lived day of an individual: a person-environment perspective," in *Person-Environment Psychology: New Directions and Perspectives*, eds W. B. Walsh, K. H. Craik, and R. H. Price (New Jersey, NJ: Erlbaum), 233–266.

Fabes, R. A., Eisenberg, N., Jones, S., Smith, M., Guthrie, I., Poulin, R., et al. (1999). Regulation, emotionality, and preschoolers' socially competent peer interactions. *Child Dev.* 70, 432–442. doi: 10.1111/1467-8624.00031

Feldman, F., and Klein, R. S. (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

Gallardo Fernández, I. M., and Saiz Fernández, H. (2016). Emotions and communicative acts from the dramatization of everyday situations. An educational intervention proposalin primary education. *Revista Electrónica Interuniversitaria de Formación del Profesorado* 19, 219–229. doi: 10.6018/reifop.19.3.267361

Gallegos, M. I., Murphy, S. E., Benner, A. D., Jacobvitz, D. B., and Hazen, N. L. (2017). Marital, parental, and whole-family predictors of toddlers' emotion regulation: the role of parental emotional withdrawal. *J. Fam. Psychol.* 31, 294–303. doi: 10.1037/fam0000245

Garner, P. W. (2006). Prediction of prosocial and emotional competence from maternal behavior in African American preschoolers. *Cult. Divers. Ethnic Minor. Psychol.* 12, 179–198. doi: 10.1037/1099-9809.12.2.179

Herndon, K., Bailey, C., Shewark, E., Denham, S., and Bassett, H. (2013). Preschoolers' emotion expression and regulation: relations with school adjustment. *J. Genet. Psychol.* 174, 642–663. doi: 10.1080/00221325.2012.759525

Keller, H., Yovsi, R., Borke, J., Kärtner, J., Jensen, H., and Papaligoura, Z. (2004). Developmental consequences of early parenting experiences: self- recognition and self-regulation in three cultural communities. *Child Dev.* 75, 1745–1760. doi: 10.1111/j.1467-8624.2004.00814.x

Kim, B. R., Stifter, C., Philbrook, L., and Teti, D. M. (2014). Infant emotion regulation: relations to bedtime emotional availability, attachment security, and temperament. *Infant Behav. Dev.* 37, 480–490. doi: 10.1016/j.infbeh.2014.06.006

Kochanska, G., Aksan, N., Penney, S. J., and Doobay, A. F. (2007). Early positive emotionality as a heterogeneous trait: implications for children's self-regulation. *J. Pers. Soc. Psychol.* 93, 1054–1066. doi: 10.1037/0022-3514.93.6.1054

Lavelli, M., Carra, C., Rossi, G., and Keller, H. (2019). Culture-specific development of early mother–infant emotional co-regulation: Italian, Cameroonian, and West African immigrant dyads. *Dev. Psychol.* 55, 1850–1867. doi: 10.1037/dev0000696

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:71. doi: 10.1136/bmj.n71

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

Roberts, W. (2020). Parents' observed responses to children's emotional distress: relations with social competence in preschool. *Br. J. Dev. Psychol.* 38, 186–204. doi: 10.1111/bjdp.12315

Sabatier, C., Restrepo Cervantes, D., Moreno Torres, M., Hoyos De los Rios, O., and Palacio Sañudo, J. (2017). Emotion regulation in children and adolescents:

concepts, processes and influences. *Psicol. Caribe* 34, 75–90. doi: 10.14482/psdc.34.1.9778

Santa Cruz, E. L. (2020). *Juego Libre en los Sectores y Autorregulación de Emociones en Estudiantes de 3 años de la I. E. I. 183 - Ate 2019*. Tesis de Maestría en Psicología Educativa. Lima: Universidad César Vallejo.

Silva, E., Freire, T., and Faria, S. (2018). Concurrent and lagged relations between emotion regulation and affect in adolescents' daily life. *Span. J. Psychol.* 21, e67. doi: 10.1017/sjp.2018.61

Van Lissa, C. J., Keizer, R., Van Lier, P. A. C., Meeus, W. H. J., and Branje, S. (2019). The role of fathers' versus mothers' parenting in emotion-regulation development from mid–late adolescence: disentangling between-family differences from within-family effects. *Dev. Psychol.* 55, 377–389. doi: 10.1037/dev0000612

Wilhelm, F. H., Grossman, P., and Müller, M. I. (2012). "Bridging the gap between the laboratory and the real world," in *Handbook of Research Methods for Studying Daily Life*, eds M. R. Mehl and T. S. Conner (New Tork, NY: The Guilford Press), 210–234.

Conflict of Interest: 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 | Frontiers in Psychology

ORIGINAL RESEARCH
published: 13 June 2022
doi: 10.3389/fpsyg.2022.889673

Modeling Not-Reached Items in Cognitive Diagnostic Assessments

Lidan Liang[1,2,3], Jing Lu[1], Jiwei Zhang[4]* and Ningzhong Shi[1]*

[1] Key Laboratory of Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University, Changchun, China, [2] School of Mathematics and Statistics, Yili Normal University, Yining, China, [3] Institute of Applied Mathematics, Yili Normal University, Yining, China, [4] Faculty of Education, Northeast Normal University, Changchun, China

In cognitive diagnostic assessments with time limits, not-reached items (i.e., continuous nonresponses at the end of tests) frequently occur because examinees drop out of the test due to insufficient time. Oftentimes, the not-reached items are related to examinees' specific cognitive attributes or knowledge structures. Thus, the underlying missing data mechanism of not-reached items is non-ignorable. In this study, a missing data model for not-reached items in cognitive diagnosis assessments was proposed. A sequential model with linear restrictions on item parameters for missing indicators was adopted; meanwhile, the deterministic inputs, noisy "and" gate model was used to model the responses. The higher-order structure was used to capture the correlation between higher-order ability parameters and dropping-out propensity parameters. A Bayesian Markov chain Monte Carlo method was used to estimate the model parameters. The simulation results showed that the proposed model improved diagnostic feedback results and produced accurate item parameters when the missing data mechanism was non-ignorable. The applicability of our model was demonstrated using a dataset from the Program for International Student Assessment 2018 computer-based mathematics cognitive test.

Keywords: cognitive diagnosis assessments, missing data mechanism, not-reached items, Bayesian analysis, sequential model

OPEN ACCESS

Edited by:
Fco. Pablo Holgado-Tello, National University of Distance Education (UNED), Spain

Reviewed by:
Zhemin Zhu, Beihua University, China
Ni Bei, University of Washington, United States

***Correspondence:**
Jing Lu
luj282@nenu.edu.cn
Jiwei Zhang
zhangjw713@nenu.edu.cn

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *04 March 2022*
Accepted: *03 May 2022*
Published: *13 June 2022*

Citation:
Liang L, Lu J, Zhang J and Shi N (2022) Modeling Not-Reached Items in Cognitive Diagnostic Assessments. Front. Psychol. 13:889673. doi: 10.3389/fpsyg.2022.889673

INTRODUCTION

In educational and psychological assessments, examinees often do not reach the end of the test which may be due to test fatigue or insufficient time. The percentage of not-reached items in large-scale cognitive testing varies across individuals, items, and countries. According to the 2006 Program for International Student Assessment (PISA) study, an average of 4% of items are not reached (OECD, 2009). In the PISA 2015 (OECD, 2018) computer-based mathematics cognitive dataset, the percentage of not-reached items in Chinese Taipei is approximately 3%, and the percentage of not-reached items for the science cluster in a Canadian sample is 2% (Pohl et al., 2019). According to the PISA 2018 (OECD, 2021) computer-based mathematics cognitive data, the proportion of nonresponses for each item ranges from 0 to 17.3% in some countries, and the maximum percentage of not-reached items is as high as 5%. Thus, the missing proportion at the item level is relatively high. In addition, the percentage of nonresponses per nation (OECD countries) ranges from 4% to15% according to

the PISA 2006 study (OECD, 2009). Even though the overall proportion of item nonresponses is small, the rate of not-reached responses for a single item or specific examinee may be large.

Previous literature focused on missing data in the item response theory (IRT) framework, which has shown that simply ignoring nonresponses or treating them as incorrect leads to biased estimates of item and person parameters (Lord, 1974, 1983; Ludlow and O'Leary, 1999; Huisman, 2000). Often, Rubin (1976) missing data mechanisms are worth reviewing for statistical inference. The complete data include observed data and unobservable missing data, and there are three types of missing data mechanisms (Rubin, 1976; Little and Rubin, 2002): missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR). MCAR refers to the probability of missing data as independent of both observed and missing data. MAR refers to the probability of missing data as only dependent on observed data. NMAR refers to the probability of missing data as dependent on the unobserved missing data itself, which is not ignorable. In general, MCAR and MAR mechanisms do not affect the parameter estimations of interest or the followed-up inference, thus missing data can be ignored in these two specific missing data mechanisms. However, Rose et al. (2010, 2017) showed that the proportion of examinees' correct scores based on the observed item responses was negatively correlated with the item nonresponse rate, which suggests that simple questions are easy to answer, and numerous difficult items may be omitted. Item nonresponses may depend on the examinee's ability and the difficulty of the items, and therefore the ignorable missing data mechanism assumption (MCAR or MAR) becomes highly questionable. This leads to the development of measurement models that consider the NMAR mechanism. Specifically, several scholars have proposed multidimensional IRT (MIRT) models to handle missing responses (e.g., Holman and Glas, 2005; Glas and Pimentel, 2008; Pohl et al., 2019; Lu and Wang, 2020). For example, Glas and Pimentel (2008) used a combination of two IRT models to model not-reached items for speeded tests according to the framework of the IRT. Subsequently, Rose et al. (2010) proposed latent regression models and multiple-group IRT models for non-ignorable missing data. Debeer et al. (2017) developed two item response tree models to handle not-reached items in various application scenarios.

Recently, cognitive diagnosis (von Davier, 2008, 2018, 2014; Xu and Zhang, 2016; Zhan et al., 2018; Zhang et al., 2020) has received considerable attention from researchers because cognitive diagnostic test enables the evaluation of the mastery of skills or attributes of respondents and allows diagnostic feedback for teachers or clinicians, which in turn aids in decision-making regarding remedial guidance or targeted interventions. In addition, the cognitive diagnostic test has improved on traditional tests. General educational examinations only provide test or ability scores in large-scale testing. However, we can neither conclude that examinees mastered the knowledge nor understand why examinees answered questions incorrectly from a single score. Moreover, it is impossible to infer differences in knowledge state and cognitive structures between individuals with the same score. Thus, the information provided by traditional IRT is not suitable for the needs of individual learning and development. To date, numerous cognitive diagnostic models (CDMs) have been developed, such as the deterministic inputs, noisy "and" gate (DINA) model (de la Torre and Douglas, 2004; de la Torre, 2009); the noisy inputs, deterministic, "and" gate model (NIDA; Maris, 1999); the deterministic inputs, noisy "or" gate (DINO) model (Templin and Henson, 2006); the log-linear CDM (Henson et al., 2009); and the generalized DINA model (de la Torre, 2011). Subsequently, a higher-order DINA (HO-DINA) model (de la Torre and Douglas, 2004) was proposed to link latent attributes *via* higher-order ability. Furthermore, Ma (2021) proposed a higher-order CDM with polytomous attributes for dichotomous response data.

Numerous studies have focused on item nonresponses in IRT models (Finch, 2008; Glas and Pimentel, 2008; Debeer et al., 2017). However, only a few studies have discussed missing data in cognitive assessments. Ömür Sünbül (2018) limited missing data mechanisms to MCAR and MAR in the DINA model and investigated different imputation approaches for dealing with item nonresponses, such as coding item responses as incorrect and using person mean imputation, two-way imputation, and expectation-maximization algorithm imputation. Heller et al. (2015) argued that CDMs may have underlying relationships with knowledge space theory (KST), which has been explored in several previous studies (e.g., Doignon and Falmagne, 1999; Falmagne and Doignon, 2011). Furthermore, de Chiusole et al. (2015) and Anselmi et al. (2016) have developed models for KST to consider different missing data mechanisms (i.e., MCAR, MAR, and NMAR). However, in their work, missing response data may not have been handled effectively, which may have biased results. Shan and Wang (2020) introduced latent missing propensities for examinees in the DINA model. They also included a potential category parameter, which affects the tendency to miss items. However, they did not provide a detailed explanation of the category parameters. Moreover, their model did not distinguish the type of item nonresponses.

The confound of different types of missing data produces inaccurate attribute profile estimations, which consequently results in incorrect diagnostic classifications. To the best of our knowledge, there has been no model developed to date that describes not-reached items in cognitive diagnosis. Thus, a missing model for not-reached items is proposed to fill this gap in cognitive diagnosis assessments. Specifically, a higher-order DINA model is used to model responses and an IRT model to describe missing indicators, which is a sequential model with linear restrictions on item parameters (Glas and Pimentel, 2008). The model is connected by bivariate normal distributions between examinees' latent ability parameters and missing propensity parameters and between item intercept and interaction parameters.

The rest of this paper is organized as follows. First, an IRT model is introduced as a missing indicator model for not-reached items. Then, a higher-order DINA model is used for the observed responses and the correlation between person parameters. Second, the Markov chain Monte Carlo (MCMC) algorithm (Patz and Junker, 1999; Chen et al., 2000) is developed to estimate the model parameters of the proposed model. Simulation studies are conducted to

assess the performance of the proposed model for different simulation conditions. Third, a real dataset from the PISA 2018 (OECD, 2021) computer-based mathematics data is analyzed. Concluding remarks and future perspectives are provided thereafter.

MODEL CONSTRUCTION

A two-dimensional data matrix with element Y_{ij} is considered, where examinees are indexed as $i = 1, ..., N$ and items are indexed as $j = 1, ..., J$. If the ith examinee answers the jth item, the response is observed, and the Y_{ij} is equal to the observation y_{ij}, otherwise, it is missing data. For convenience, the sign "d" is used to mark the missing data and the relevant parameters.

Missing Data Model for Not-Reached Items

Glas and Pimentel (2008) proposed a sequential model with a linear restriction on the item parameters to model the not-reached items. Specifically, the missing indicator matrix $\mathbf{D}$ with element d_{ij} is given by:

$$d_{ij} = \begin{cases} 0, & \text{if } y_{ij} \text{ was observerd}, \\ 1, & \text{if } y_{ij} \text{ was not observer}. \end{cases} \tag{1}$$

where $d_{ij} = 1$ indicates that the ith examinee drops out the jth item. Because of the small overall proportion of not-reached responses, the appropriate model must have few parameters to be estimable (Lord, 1983). The one-parameter logistic model (1PLM; Rasch, 1960) is adopted to model the missing indicators, thus the dropping-out probability of examinee i on item j is:

$$p\left(d_{ij} = 1|\theta_i^d, \beta_j^d\right) = \frac{exp\left(\theta_i^d - \beta_j^d\right)}{1 + exp\left(\theta_i^d - \beta_j^d\right)}, \tag{2}$$

and

$$\beta_j^d = \eta_0 + \left(j - J\right)\eta_1, \tag{3}$$

where β_j^d represents the so-called item difficulty parameter for item j, and θ_i^d denotes the ith examinee's dropping-out propensity. Also, $\beta_j^d = \eta_0$ when $j = J$, where η_0 is the difficulty threshold of the last item, and η_1 models a uniform change in the probability as a function of the item position in the test. Usually, the parameter η_1 is negative, and hence it is more likely to drop out the test at later position items of the test.

Higher-Order Deterministic Inputs, Noisy "And" Gate Model

The DINA model describes the probability of the item response as a function of latent attributes, and the probability of the ith examinee responding to item j correctly is as follows:

$$p(Y_{ij} = 1) = g_j + \left(1 - s_j - g_j\right)\prod_{k=1}^{K} \alpha_{ik}{}^{q_{jk}}, \tag{4}$$

where s_j and g_j are the slipping and guessing probabilities of the jth item, respectively, $1 - s_j - g_j = IDI_j$ is the jth item discrimination index (de la Torre, 2008), and α_{ik} is the kth attribute of the ith examinee, with $\alpha_{ik} = 1$ if examinee i masters attribute k and $\alpha_{ik} = 0$ if examinee does not master attribute k. The Q matrix (Tatsuoka, 1983) is an $J \times K$ matrix, with q_{jk}, $q_{jk} = 1$ denoting that the attribute k is required for answering the jth item correctly and $q_{jk} = 0$ if the attribute k is not required for answering the jth item correctly.

Equation (4) can be reparameterized as the reparameterized DINA model (DeCarlo, 2011).

$$\beta_j = logit(g_j), \tag{5}$$

$$\delta_j = logit\left(1 - s_j\right) - logit\left(g_j\right). \tag{6}$$

In addition, $logit(x) = log(\frac{x}{1-x})$, thus Equation (4) can be reformed as,

$$logit(P(y_{ij} = 1)) = \beta_j + \delta_j \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}, \tag{7}$$

where β_j and δ_j are the item intercept and interaction parameter, respectively, and they are assumed to follow a bivariate normal distribution as follows:

$$\begin{pmatrix} \beta_j \\ \delta_j \end{pmatrix} \sim N(\begin{pmatrix} \mu_\beta \\ \mu_\delta \end{pmatrix}, \Sigma_\mathbf{I}), \Sigma_\mathbf{I} = \begin{pmatrix} \sigma_\beta^2 & \sigma_{\beta\delta} \\ \sigma_{\beta\delta} & \sigma_\delta^2 \end{pmatrix}. \tag{8}$$

The higher-order structure is very flexible because it can reduce the number of model parameters and can provide higher-order abilities and more accurate attribute structures. Because the attributes in a test are often correlated, the higher-order structure (de la Torre and Douglas, 2004; Zhan et al., 2018) for the attributes is expressed as,

$$logit(P(\alpha_{ik} = 1)) = \theta_i{}^h\gamma_k - \lambda_k, \tag{9}$$

where $P(\alpha_{ik} = 1)$ is the probability that the ith examinee masters the kth attribute, θ_i^h is the higher-order ability of examinee i, and γ_k and λ_k are the slope and intercept parameters of attribute k, respectively. The slope parameter γ_k is positive because the knowledge attribute is mastered better with the increased ability θ_i^h.

Missing Mechanism Models

If the observation probability $p(y_{ij}|d_{ij}, \beta_j, \delta_j, \alpha_{ik})$ does not depend on $\boldsymbol{\theta}^\mathbf{d}$, when $\boldsymbol{\theta}^\mathbf{h}$ and $\boldsymbol{\theta}^\mathbf{d}$ are independent, then the missing data are ignorable. In this situation, this model is treated as a MAR model. Let $p(y_{ij}|d_{ij}, \beta_j, \delta_j, \alpha_{ik})$ be the measurement model for the observed data. In addition, let $p(d_{ij}|\theta_i^d, \eta_0, \eta_1)$ be the measurement model for the missing data indicators, and $p\left(\boldsymbol{\theta}^\mathbf{h}\right)$ and $p(\boldsymbol{\theta}^\mathbf{d})$ are densities of $\boldsymbol{\theta}^\mathbf{h}$ and $\boldsymbol{\theta}^\mathbf{d}$, respectively. To model non-ignorable missing data, it is assumed that θ_i^h and θ_i^d follow a bivariate normal distribution $N(\boldsymbol{\mu}_\mathrm{P}, \boldsymbol{\Sigma}_\mathrm{P})$; thus, the two models describe the

two missing mechanisms (i.e., MAR and NMAR). Next, we introduce the two missing data models for the not-reached items.

Missing at Random Model

The expression of the MAR model is as follows, and the likelihood function form of the MAR model can be written as,

$$\prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{k=1}^{K} p(\alpha_{ik}|\theta_i^h, \gamma_k, \lambda_k)p(d_{ij}|\theta_i^d, \eta_0, \eta_1)p(\theta_i^h)p(\theta_i^d), \quad (10)$$

where the MAR model is regarded as a model that ignores the missing data process. In fact, the latent variables θ_i^h and θ_i^d are independent in the MAR model. In other words, the model for the missing data process $p(d_{ij}|\theta_i^d, \eta_0, \eta_1)$ can be ignored in estimating the item response model.

Not Missing at Random Model

The NMAR model is often called the non-ignorable model, and in this case, θ_i^h and θ_i^d are correlated. A covariance matrix is used to describe the relationship between the latent higher-order ability parameters and the missing propensity parameters in this model. Thus, the likelihood function of the NMAR model can be written as,

$$\prod_{i=1}^{N}\prod_{j=1}^{J}\prod_{k=1}^{K} p(\alpha_{ik}|\theta_i^h, \gamma_k, \lambda_k)p(d_{ij}|\theta_i^d, \eta_0, \eta_1)p(\theta_i^h, \theta_i^d|\boldsymbol{\mu}_{\mathrm{P}}, \boldsymbol{\Sigma}_{\mathrm{P}}), \quad (11)$$

where the person parameters are assumed to follow a bivariate normal distribution, with mean vector $\boldsymbol{\mu}_{\mathrm{P}} = (\mu_{\theta^h}, \mu_{\theta^d})'$ and covariance matrix:

$$\boldsymbol{\Sigma}_{\mathrm{P}} = \begin{pmatrix} \sigma^2_{\theta^h} & \sigma_{\theta^h\theta^d} \\ \sigma_{\theta^h\theta^d} & \sigma^2_{\theta^d} \end{pmatrix}. \quad (12)$$

Model Identifications

In Equations (2) and (9), the linear parts of 1PLM and the HO-DINA model can be written as follows:

$$\theta_i^d - \beta_j^d \; and \; \theta_i^h\gamma_k - \lambda_k. \quad (13)$$

To eliminate the trade-offs between ability θ_i^d and dropping-out threshold parameter β_j^d and between the higher-order ability person parameter θ^h and the attribute intercept λ_k, the mean population level of person parameters is set to zero, that is, $\mu_{\theta^h} = 0 \; and \; \mu_{\theta^d} = 0$. $\sigma_{\theta^h} = 1$ is fixed to eliminate the scale trade-off between θ_i^h and γ_k (Lord and Novick, 1968; Fox, 2010). In addition to the identifications, two local independence assumptions are made, that is, the α_{ik} values are conditionally independent given θ_i^h, and the Y_{ij} values are conditionally independent given α_i.

Bayesian Model Assessment

In the Bayesian framework, two common Bayesian model evaluation criteria, the deviance information criteria (DIC; Spiegelhalter et al., 2002) and the logarithm of the pseudo-marginal likelihood (LPML, Geisser and Eddy, 1979; Ibrahim et al., 2001) are used to compare the differences in the missing mechanism models according to the results of MCMC sampling. Let,

$$\boldsymbol{\Omega} = \left\{\theta_i^h, \theta_i^d, \eta_0, \eta_1, \alpha_{ik}, \beta_j, \delta_j, \gamma_k, \lambda_k, \mu_\beta, \mu_\delta, \boldsymbol{\Sigma}_{\mathbf{I}}, \sigma_{\theta^h\theta^d}, \sigma^2_{\theta^d}\right\}.$$

The DIC is given by,

$$\begin{aligned} \mathrm{Dev}_{(\mathbf{Y},\mathbf{D}|\boldsymbol{\Omega})} &= -2logL(\mathbf{Y}, \mathbf{D}, \boldsymbol{\Omega}) \\ &= -2\sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K}[(Y_{ij} = d)log(P(Y_{ij} = d)) \\ &+ (Y_{ij} = 1)log((1 - P(Y_{ij} = d))P(Y_{ij} = 1)) \\ &+ (Y_{ij} = 0)log((1 - P(Y_{ij} = d))P(Y_{ij} = 0))]. \end{aligned} \quad (14)$$

On the basis of the posterior distribution of Dev $(\mathbf{Y}, \mathbf{D}, \boldsymbol{\Omega})$, the DIC was defined as,

$$\mathrm{DIC} = \overline{\mathrm{Dev}} + p_D = \overline{\mathrm{Dev}} + (\overline{\mathrm{Dev}} - \widehat{\mathrm{Dev}}), \quad (15)$$

where $\overline{\mathrm{Dev}} = E(\mathrm{Dev}(\mathbf{Y}, \mathbf{D}, \boldsymbol{\Omega})|\mathbf{Y}, \mathbf{D}) \cong \frac{1}{R}\sum_{r=1}^{R} \mathrm{Dev}(\mathbf{Y}, \mathbf{D}, \boldsymbol{\Omega}^r)$, which is the posterior mean deviance and is a Bayesian measure of fit, $r = 1, \ldots, R$ denotes the rth iteration of the algorithm, and $\widehat{\mathrm{Dev}} = \mathrm{Dev}\left(\mathbf{Y}, \mathbf{D}, \overline{\boldsymbol{\Omega}}\right)$, which is the effective number of parameters, is a Bayesian measure of complexity, with $\overline{\boldsymbol{\Omega}} = E(\boldsymbol{\Omega}|\mathbf{Y}, \mathbf{D}) \cong \frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\Omega}^r$. A smaller DIC indicates a better model fit.

The conditional predictive ordinate (CPO) index of the two models was computed. Let $Q_{ij,max} = max_{1 \le r \le R}\left\{-\log f(Y_{ij}, D_{ij}|\boldsymbol{\Omega}^r)\right\}$. Thus,

$$\log\left(\widehat{\mathrm{CPO}_{ij}}\right) = -Q_{ij,max} - log\left[\frac{1}{R}\sum_{r=1}^{r} exp\left\{-logf\left(Y_{ij}, D_{ij}|\boldsymbol{\Omega}^r\right) - Q_{ij,max}\right\}\right]. \quad (16)$$

The summary statistic for $\log\left(\widehat{\mathrm{CPO}_{ij}}\right)$ is the sum of their logarithms, which is termed the LPML and is given by,

$$\mathrm{LPML} = \sum_{i=1}^{N}\sum_{j=1}^{J}\log(\widehat{\mathrm{CPO}}_{ij}), \quad (17)$$

where the model with a larger LPML indicates a better fit to the data.

SIMULATION STUDIES

Three simulation studies were conducted to evaluate different aspects of the proposed model. Simulation study I was conducted to assess whether the MCMC algorithm could successfully recover parameters of the proposed model under different numbers of examinees and items. Simulation study II was

conducted to investigate the parameter recovery of different numbers of attributes for the same examinees and items. Simulation study III intended to show the differences in model parameter estimates between the NMAR and MAR models for different dropping-out proportions and correlations among person parameters.

Data Generation

In the three simulation studies, the item parameters were sampled from the following distributions: $\begin{pmatrix}\beta_j\\ \delta_j\end{pmatrix} \sim MVN\left(\begin{pmatrix}\mu_\beta\\ \mu_\delta\end{pmatrix}, \boldsymbol{\Sigma_I}\right)$, $\mu_\beta = -2.197$, $\mu_\delta = 4.394$, $\boldsymbol{\Sigma_I} = \begin{pmatrix}1 & -0.8\\ -0.8 & 1\end{pmatrix}$. These values were used in Shan and Wang (2020) study. The dropping-out proportions across three levels (i.e., low, medium, and high) were varied by setting different combinations of η_0 and η_1. That is, the dropping-out proportion was 3.8 (low) when $\eta_0 = 1, \eta_1 = -0.7$; the dropping-out proportion was 12 (medium) when $\eta_0 = 1, \eta_1 = -0.32$; and the dropping-out proportion was 25% (high) when $\eta_0 = 1, \eta_1 = -0.18$.

The attribute intercept parameters were $\lambda = (-1, -0.5, 0, 0.5, 1)$, and the attribute slope parameters were$\gamma_k = 1.5$ for all attributes, which were consistent with those in the study by Shan and Wang (2020). Three Q matrices with different numbers of attributes (**Figure 1**) were considered, and the three Q matrices were taken from Xu and Shang (2018) study and Shan and Wang (2020) study.

The person parameters θ_i^h and θ_i^d were simulated from the bivariate normal distribution $\begin{pmatrix}\theta^h\\ \theta^d\end{pmatrix} \sim MVN\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}1 & \sigma_{\theta^h\theta^d}\\ \sigma_{\theta^h\theta^d} & \sigma^2_{\theta^d}\end{pmatrix}\right)$, where $\sigma^2_{\theta^d} = 0.25$. Three levels of correlation between θ_i^h and θ_i^d were considered for $\rho_{\theta_i^h\theta_i^d}$: 0 (uncorrelated), −0.5 (medium), and −0.8 (high). The missing data due to dropping-out items were simulated in the following manner. The three levels of dropping-out proportions were 3.8% (low), 12% (medium), and 25% (high).

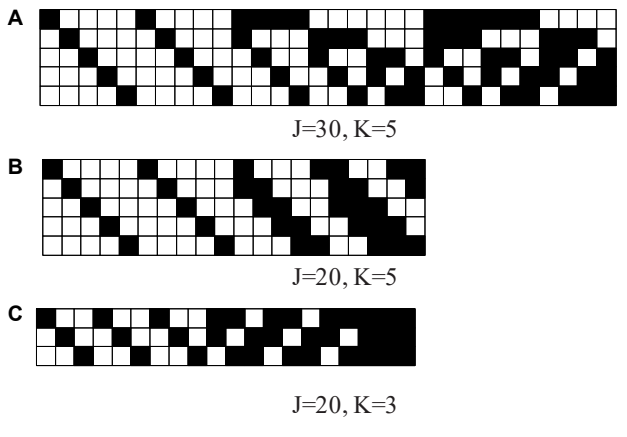


FIGURE 1 | K-by-J Q matrices in simulation studies, where black means "1" and white means "0." K is the number of attributes and J is the number of items.

Model Calibration

The priors of η_0 and η_1 were $\eta_0 \sim N(0, 2)$ and $\eta_1 \sim N(0, 2)$, respectively. The priors of the item parameters β_j and δ_j were assumed to have a bivariate normal distribution: $\begin{pmatrix}\beta_j\\ \delta_j\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_\beta\\ \mu_\delta\end{pmatrix}, \boldsymbol{\Sigma_I}\right)$. The priors of the person parameters were assumed to follow a bivariate normal distribution: $\begin{pmatrix}\theta^h\\ \theta^d\end{pmatrix} \sim N\left(\begin{pmatrix}0\\ 0\end{pmatrix}, \boldsymbol{\Sigma_P}\right)$. The priors of the higher-order structure parameters were expressed as $\lambda_k \sim N(0, 4)$ and $\gamma_k \sim N(0, 4)I(\gamma_k > 0)$, the priors of the covariance matrix of the person were expressed as $\sigma_{\theta^h\theta^d} \sim U(-1, 1)$ and $\sigma^2_{\theta^d} \sim$Inv-(2, 2), the priors of the covariance matrix of the item parameters were expressed as$\boldsymbol{\Sigma_I}$ Inv-Wishart $\left(\boldsymbol{\Sigma}_{I0}^{-1}, \nu_{I0}\right)$, and the hyperpriors were specified as $\boldsymbol{\Sigma}_{I0} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$, $\nu_{I0} = 2, k_{I0} = 1$, $\mu_\beta \sim N(-2.197, 2)$, and $\mu_\delta \sim N(4.394, 2)I(\mu_\delta > 0)$. The hyperpriors specified above were on a logit scale for $\boldsymbol{\beta}$ and$\boldsymbol{\delta}$ and were consistent with those reported by Zhan et al. (2018). The mean guessing effect was set at 0.1, which was roughly equal to a logit value −2.197 for μ_β. A standard deviation of $\sqrt{2}$ on the logit scale for μ_β indicated that the simulated mean guessing effect changed from 0.026 to 0.314. In addition, the mean slipping effect was also set at 0.1, which indicated that μ_δ was approximately 4.394 on the logit scale. The simulated mean slipping effect changed from 0.007 to 0.653 under a standard deviation of $\sqrt{2}$ on the logit scale for $\boldsymbol{\delta}$.

The initial values of the model parameters were as follows: $\beta_j = 0$, $\delta_j = 0$ for $j = 1, \ldots, J$, $\theta_i^h = 0$, $\theta_i^d = 0$ for $i = 1, \ldots, N$, $\sigma_{\theta^h\theta^d} = 0$, $\sigma^2_{\theta^d} = 1$, $\eta_0 = 0, \eta_1 = 0, \mu_\beta = 0, \mu_\delta = 0$, $\boldsymbol{\Sigma_P} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$, $\boldsymbol{\mu_P} = \begin{pmatrix}0\\ 0\end{pmatrix}$, and $\sigma^2_{\theta^h} = 1$. In addition, $\lambda_k = 0$, $\gamma_k = 1$ for $k = 1, , K$, and $\boldsymbol{\alpha} = \begin{pmatrix}\alpha_{11} & \cdots & \alpha_{1K}\\ \vdots & \ddots & \vdots\\ \alpha_{N1} & \cdots & \alpha_{NK}\end{pmatrix}$, where α_{ik} $(i = 1, \ldots, N, k = 1, \ldots, K)$ were sampled from 0 to 1 randomly. The proposal variances were chosen to give Metropolis acceptance rates between 25% and 40%. The Markov chain length was set at 10,000 so that the potential scale reduction factor (PSRF; Brooks and Gelman, 1998) was less than 1.1 for all parameters, which implied proper chain convergence. Five thousand iterations were treated as burn-in. The final parameter estimates were obtained as the average of the post-burn-in iterations.

In terms of evaluation criteria, the bias and root mean squared error (RMSE) are used to assess the accuracy of the parameter estimates. In particular, the bias for parameter η was,

$$\text{bias}(\eta) = \frac{1}{R}\sum_{r=1}^{R}\left(\widehat{\eta}^{(r)} - \eta\right), \tag{18}$$

and the RMSE for parameter η is defined as,

$$\text{RMSE}(\eta) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\widehat{\eta}^{(r)} - \eta\right)^2}, \quad (19)$$

where η is the true value of the parameter, and $\widehat{\eta}^{(r)}$ is the estimate for the rth replication. There were $R = 30$ replications for each simulation condition. The recoveries of attributes are evaluated using the attribute correct classification rate (ACCR) and the pattern correct classification rate (PCCR):

$$\text{ACCR} = \frac{\sum_{i=1}^{N} I\left(\widehat{\alpha_{ik}} = \alpha_{ik}\right)}{N}, \quad (20)$$

$$\text{PCCR} = \frac{\sum_{i=1}^{N}\left[\prod_{k=1}^{K} I\left(\widehat{\alpha_{ik}} = \alpha_{ik}\right)\right]}{N}, \quad (21)$$

where $I\left(\widehat{\alpha_{ik}} = \alpha_{ik}\right)$ is the indicator function that is, $I\left(\widehat{\alpha_{ik}} = \alpha_{ik}\right) = 1$ if $\widehat{\alpha_{ik}} = \alpha_{ik}$, otherwise $I\left(\widehat{\alpha_{ik}} = \alpha_{ik}\right) = 0$.

Simulation Study I

In simulation study I, the different numbers of examinees and items were considered to estimate the model parameters under a fixed number of five attributes. Three conditions were considered in this simulation: (a) 500 examinees and 30 items, (b) 1,000 examinees and 30 items, and (c) 500 examinees and 20 items. The correlation between θ_i^h and θ_i^d was −0.3, and the dropping-out proportion was medium.

Table 1 presents the bias and RMSE of the ability parameters and item parameters, as well as the attribute parameter estimates. For the 30 items and the 5 attributes (please see the first four columns of **Table 1**), the item parameter estimates improve when the number of examinees increases from 500 to 1,000, the bias and RMSE of $\boldsymbol{\delta}$ and μ_{β} decrease, and the RMSE of $\boldsymbol{\beta}$, μ_{δ}, and item covariance matrix elements reduce. For the 500 examinees and the 5 attributes (please see the middle four columns of **Table 1**), the person parameter estimates improve when the number of items increases from 20 to 30, and $\boldsymbol{\theta^h}$ and $\sigma^2_{\theta^d}$ are more accurate. The ACCRs and PCCRs are presented in **Table 2**. The ACCRs and PCCRs could be recovered satisfactorily with a larger sample and longer test length. The ACCRs and PCCRs decrease when the number of examinees or test length decreases (please see the first three columns in **Table 2**), and the changes are particularly marked when the test length is reduced. **Figure 2** shows the PSRF of several items and attribute parameters under 500 examinees and 30 items. It is observed that the item intercept parameter $\boldsymbol{\beta}$, the interaction parameter $\boldsymbol{\delta}$, the attribute slope parameter $\boldsymbol{\gamma}$, and the attribute intercept parameter $\boldsymbol{\lambda}$ converge at 5,000 iterations, and the convergence of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are significantly faster than that of $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$.

TABLE 1 | Bias and RMSE of the parameter estimates in simulation studies I and II.

	N **= 1000**		***N*** **= 500**		***N*** **= 500**		***N*** **= 500**	
	J **= 30**		***J*** **= 30**		***J*** **= 20**		***J*** **= 20**	
	K **= 5**		***K*** **= 5**		***K*** **= 5**		***K*** **= 3**	
Parameter	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
β	0.009	**0.167**	−0.002	**0.198**	−0.134	0.272	−0.020	0.282
δ	**−0.001**	**0.274**	**−0.051**	**0.339**	0.072	0.345	0.017	0.351
μ_{β}	**−0.111**	**0.203**	**−0.120**	**0.215**	−0.296	0.374	−0.192	0.268
μ_{δ}	0.035	**0.181**	−0.017	**0.196**	0.236	0.356	0.191	0.313
λ_1	0.078	0.137	0.063	0.179	0.066	0.191	−0.109	0.181
λ_2	0.029	0.100	−0.133	0.193	−0.149	0.199	−0.030	0.130
λ_3	0.052	0.104	−0.058	0.143	−0.127	0.202	−0.204	0.245
λ_4	0.040	0.106	−0.069	0.145	−0.121	0.178	–	–
λ_5	0.201	0.239	−0.089	0.188	−0.181	0.246	–	–
γ_1	0.129	0.249	0.296	0.457	0.222	0.403	−0.179	0.451
γ_2	0.034	0.189	0.065	0.268	−0.288	0.360	−0.156	0.545
γ_3	−0.063	0.182	−0.002	0.252	0.359	0.527	−0.301	0.626
γ_4	−0.027	0.180	−0.202	0.298	−0.139	0.276	–	–
γ_5	0.039	0.206	0.153	0.326	−0.083	0.282	–	–
σ^2_{β}	−0.152	**0.281**	−0.051	**0.282**	−0.035	0.374	−0.353	0.429
$\sigma_{\beta\delta}$	0.093	**0.244**	−0.027	**0.280**	−0.118	0.415	0.131	0.318
σ^2_{δ}	−0.103	**0.282**	0.066	**0.340**	0.315	0.611	0.132	0.457
η_0	−0.051	**0.086**	**−0.014**	**0.097**	0.053	0.112	−0.130	0.161
η_1	**−0.004**	**0.013**	**0.005**	**0.017**	0.008	0.019	−0.013	0.022
$\sigma_{\theta^h\theta^d}$	−0.056	0.077	−0.046	0.091	0.001	0.083	0.057	0.105
$\sigma^2_{\theta^d}$	−0.001	0.081	**0.008**	**0.094**	**0.018**	**0.101**	−0.029	0.075
θ^h	0.071	0.625	**−0.043**	**0.594**	**−0.044**	**0.612**	−0.044	0.701
θ^d	−0.039	0.480	0.006	0.475	0.006	0.468	0.006	0.479

The boldfaced values indicate that much smaller Bias and RMSE are obtained from the model.

TABLE 2 | ACCRs and PCCRs in simulation studies I and II.

	***N* = 1000**	***N* = 500**	***N* = 500**	***N* = 500**
	***J* = 30**	***J* = 30**	***J* = 20**	***J* = 20**
	***K* = 5**	***K* = 5**	***K* = 5**	***K* = 3**
ACCR	0.968	0.966	0.922	0.985
	0.980	0.976	0.966	0.993
	0.984	0.985	0.960	0.982
	0.986	0.977	0.984	–
	0.986	0.981	0.954	–
PCCR	**0.910**	**0.898**	**0.811**	**0.961**

The boldfaced values indicate that much smaller Bias and RMSE are obtained from the model.

Simulation Study II

This simulation study was conducted to investigate the parameter recovery of different numbers of attributes for fixed 500 examinees and 20 items. The correlation between θ_i^h and θ_i^d was set at −0.3, and the dropping-out proportion was medium.

The last four columns of **Table 1** show the results of simulation study II. The RMSE of the estimates of item and person parameters with attribute $K = 5$ are smaller than those with attribute $K = 3$. The RMSE of the attribute slope parameters and intercept parameters recover more satisfactorily with attribute $K = 3$ than with attribute $K = 5$. The last two columns of **Table 2** show the ACCRs and PCCRs for simulation study II. The ACCRs with attribute $K = 3$ are higher than those with attribute $K = 5$ and improve from 0.957 to 0.987 on average. Moreover, the PCCRs are significantly higher when the number of attributes decreases. That is, the PCCR with attribute $K = 5$ is 0.811, and the PCCR with attribute $K = 3$ is 0.961.

Simulation Study III

The purpose of this simulation study was to investigate the parameter recovery with the NMAR model, MAR model, and HO-DINA model that ignores the not-reached items under different simulation conditions. The data were generated using the proposed model with the NMAR mechanism. A total of 500 examinees answered 30 items, and each item had 5 attributes.

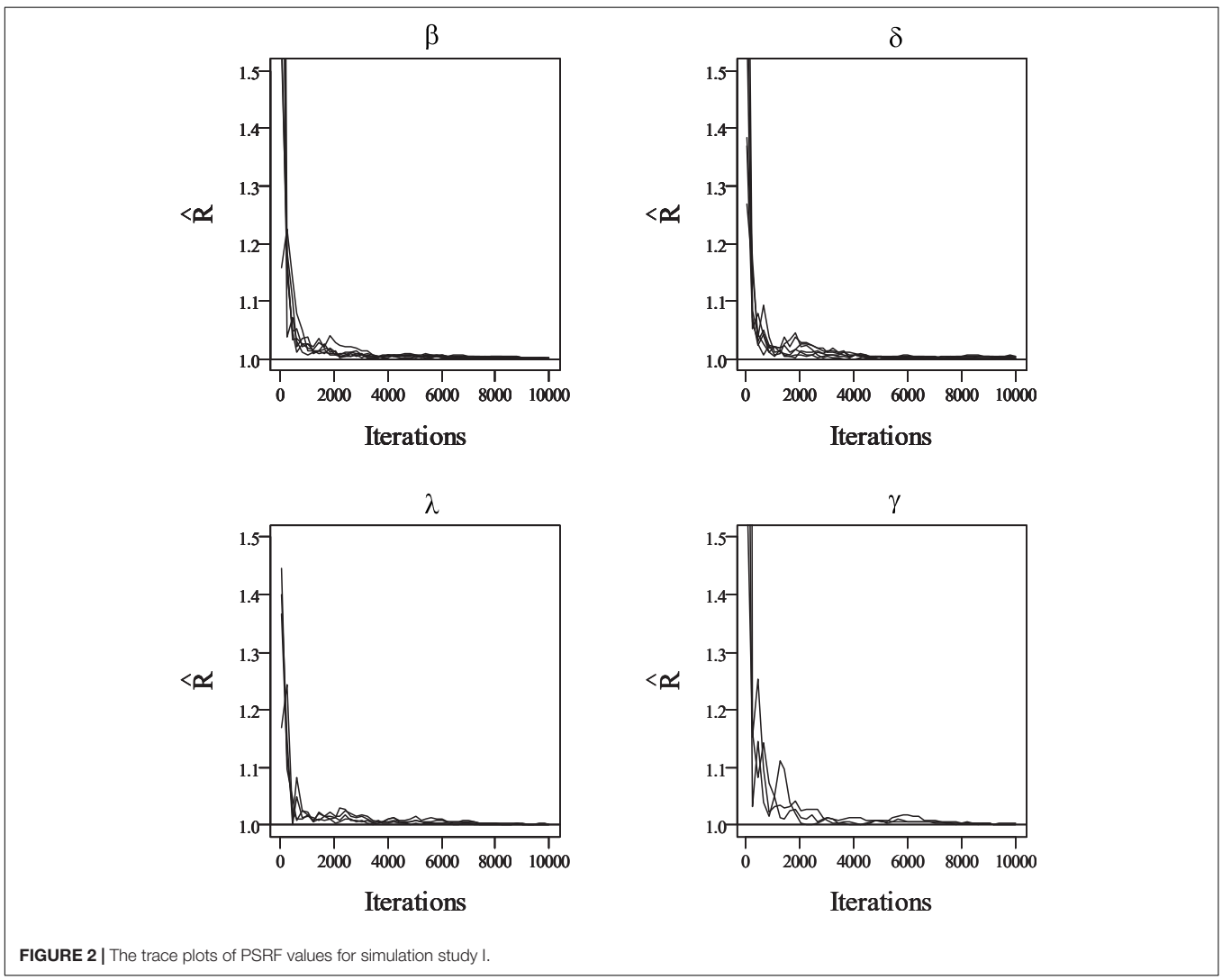


FIGURE 2 | The trace plots of PSRF values for simulation study I.

TABLE 3 | Bias and RMSE of parameter estimates of three models with low dropping-out proportion under different correlations between θ_i^h and θ_i^d in simulation study III.

		$\rho = 0$			$\rho = -0.5$			$\rho = -0.8$		
Parameter	**Statistics**	**NMAR**	**MAR**	**HO-DINA**	**NMAR**	**MAR**	**HO-DINA**	**NMAR**	**MAR**	**HO-DINA**
η_0	Bias	0.003	0.001	–	0.036	−0.001	–	**0.015**	**−0.019**	–
	RMSE	0.123	0.125	–	**0.155**	**0.174**	–	**0.134**	**0.162**	–
η_1	Bias	0.005	0.004	–	**−0.004**	**−0.109**	–	**−0.003**	**−0.107**	–
	RMSE	0.055	0.055	–	**0.065**	**0.137**	–	**0.059**	**0.131**	–
β	Bias	−0.018	−0.016	−0.015	**−0.003**	**0.124**	0.121	**−0.029**	**0.093**	0.093
	RMSE	0.234	0.233	0.234	**0.239**	**0.299**	0.297	**0.237**	**0.285**	0.286
δ	Bias	0.039	0.047	0.045	0.022	−0.017	−0.015	0.063	0.021	0.021
	RMSE	0.336	0.345	0.346	**0.341**	**0.369**	0.369	**0.346**	**0.369**	0.369
μ_β	Bias	−0.136	−0.117	−0.115	−0.120	0.006	0.004	−0.146	−0.022	−0.022
	RMSE	0.228	0.217	0.218	0.218	0.201	0.201	0.235	0.204	0.204
μ_δ	Bias	0.073	0.067	0.064	0.054	0.016	0.017	0.095	0.052	0.052
	RMSE	0.216	0.228	0.229	**0.205**	**0.255**	0.255	**0.223**	**0.259**	0.263
σ_β^2	Bias	−0.052	−0.053	−0.056	**−0.067**	**0.074**	0.075	**−0.055**	**0.096**	0.096
	RMSE	0.290	0.290	0.289	**0.291**	**0.322**	0.322	**0.291**	**0.331**	0.332
$\sigma_{\beta\delta}$	Bias	0.008	−0.005	−0.003	**0.051**	**−0.275**	−0.276	**0.028**	**−0.316**	−0.314
	RMSE	0.286	0.299	0.296	**0.281**	**0.446**	0.445	**0.289**	**0.479**	0.478
σ_δ^2	Bias	0.054	0.225	0.222	**−0.021**	**0.656**	0.657	**0.004**	**0.703**	0.700
	RMSE	0.358	0.447	0.443	**0.333**	**0.812**	0.811	**0.355**	**0.856**	0.855
λ_1	Bias	0.039	0.017	0.017	**0.098**	**0.298**	0.285	**0.056**	**0.220**	0.224
	RMSE	0.168	0.172	0.173	**0.193**	**0.370**	0.363	**0.181**	**0.331**	0.330
λ_2	Bias	−0.096	−0.111	−0.108	−0.103	−0.051	−0.055	−0.096	−0.049	−0.048
	RMSE	0.168	0.180	0.178	0.168	0.160	0.163	0.166	0.159	0.159
λ_3	Bias	−0.051	−0.053	−0.052	−0.127	−0.003	−0.011	−0.091	0.030	0.033
	RMSE	0.147	0.149	0.150	0.188	0.163	0.162	**0.167**	**0.169**	0.168
λ_4	Bias	−0.089	−0.084	−0.083	**−0.068**	**0.023**	0.018	−0.080	0.002	0.003
	RMSE	0.162	0.161	0.161	0.152	0.149	0.150	0.153	0.141	0.141
λ_5	Bias	−0.102	−0.076	−0.081	−0.142	0.019	0.017	−0.135	0.006	0.007
	RMSE	0.194	0.186	0.190	0.214	0.185	0.187	0.210	0.181	0.180
γ_1	Bias	0.122	0.173	0.179	**0.178**	**0.294**	0.263	**0.277**	**0.501**	0.520
	RMSE	0.346	0.371	0.374	**0.387**	**0.472**	0.433	**0.451**	**0.698**	0.710
γ_2	Bias	−0.004	0.044	0.035	**−0.117**	**0.246**	0.245	**−0.084**	**0.246**	0.247
	RMSE	0.276	0.284	0.275	**0.281**	**0.380**	0.381	**0.271**	**0.372**	0.377
γ_3	Bias	0.080	0.104	0.111	**0.126**	**0.474**	0.477	**0.141**	**0.485**	0.494
	RMSE	0.301	0.312	0.313	**0.323**	**0.577**	0.583	**0.332**	**0.594**	0.603
γ_4	Bias	−0.103	−0.077	−0.078	−0.114	0.025	0.021	−0.178	−0.037	−0.035
	RMSE	0.267	0.263	0.264	0.274	0.252	0.256	0.287	0.235	0.233
γ_5	Bias	−0.052	0.005	−0.006	**−0.075**	**0.114**	0.115	**−0.039**	**0.137**	0.132
	RMSE	0.289	0.286	0.290	**0.284**	**0.307**	0.310	**0.280**	**0.313**	0.309
θ^d	Bias	−0.002	−0.002	–	**0.011**	**0.011**	–	**0.017**	**0.018**	–
	RMSE	0.499	0.492	–	**0.454**	**0.667**	–	**0.377**	**0.668**	–
θ^h	Bias	−0.044	−0.046	−0.046	**−0.044**	**−0.044**	−0.047	**−0.044**	**−0.046**	−0.045
	RMSE	0.581	0.581	0.580	**0.582**	**0.591**	0.592	**0.578**	**0.591**	0.591
$\sigma_{\theta^d}^2$	Bias	−0.002	0.007	–	**0.013**	**1.022**	–	**0.015**	**1.023**	–
	RMSE	0.089	0.088	–	**0.095**	**1.160**	–	**0.081**	**1.097**	–
$\sigma_{\theta^h\theta^d}$	Bias	0.011	–	–	0.015	–	–	0.010	–	–
	RMSE	0.131	–	–	0.113	–	–	0.082	–	–

NMAR means not missing at random model, MAR means missing at random model, HO-DINA means higher-order DINA model. The boldfaced values indicate that much smaller Bias and RMSE are obtained from the model.

Three dropping-out proportions (i.e., 3.8% [low], 12% [medium], and 25% [high]) and three correlations between θ_i^h and θ_i^d (i.e., 0 [uncorrelated], −0.5 [medium], and −0.8 [high]) were manipulated. Thus, there were 3 × 3 simulation conditions.

Table 3 shows the bias and RMSE of the parameters of three models with low dropping-out proportions under different correlations between θ_i^h and θ_i^d. Results show that the parameter estimates from the three models are similar when the correlation

TABLE 4 | Bias and RMSE of parameter estimates of three models with medium dropping-out proportion under different correlations between θ_i^h and θ_i^d in simulation study III.

		$\rho = 0$			$\rho = -0.5$			$\rho = -0.8$		
Parameter	**Statistics**	**NMAR**	**MAR**	**HO-DINA**	**NMAR**	**MAR**	**HO-DINA**	**NMAR**	**MAR**	**HO-DINA**
η_0	Bias	0.014	0.009	–	**−0.006**	**−0.159**	–	**−0.033**	**−0.181**	–
	RMSE	0.133	0.131	–	**0.131**	**0.216**	–	**0.123**	**0.226**	–
η_1	Bias	0.001	0.001	–	**−0.001**	**−0.039**	–	**−0.011**	**−0.048**	–
	RMSE	−0.002	−0.003	–	**−0.001**	**−0.024**	–	**−0.001**	**−0.019**	–
β	Bias	−0.022	−0.019	−0.019	**−0.028**	**0.114**	0.113	−0.021	0.119	0.118
	RMSE	0.249	0.248	0.249	**0.265**	**0.323**	0.322	**0.249**	**0.309**	0.309
δ	Bias	0.071	0.082	0.081	0.042	−0.002	−0.001	0.052	0.005	0.007
	RMSE	0.365	0.378	0.377	**0.360**	**0.401**	0.400	**0.357**	**0.389**	0.391
μ_β	Bias	−0.137	−0.121	−0.120	−0.146	−0.001	0.001	−0.134	0.008	0.004
	RMSE	0.229	0.226	0.223	0.238	0.206	0.204	0.229	0.207	0.202
μ_δ	Bias	0.102	0.103	0.102	0.077	0.029	0.026	0.080	0.032	0.037
	RMSE	0.232	0.250	0.247	**0.224**	**0.266**	0.264	**0.226**	**0.269**	0.268
σ_β^2	Bias	−0.031	−0.031	−0.033	**−0.032**	**0.105**	0.108	**−0.046**	**0.095**	0.095
	RMSE	0.308	0.307	0.306	**0.299**	**0.341**	0.342	**0.299**	**0.338**	0.338
$\sigma_{\beta\delta}$	Bias	−0.015	−0.024	−0.023	**−0.015**	**−0.344**	−0.346	**0.029**	**−0.304**	−0.304
	RMSE	0.310	0.319	0.319	**0.296**	**0.504**	0.505	**0.286**	**0.471**	0.471
σ_δ^2	Bias	0.107	0.277	0.274	**0.075**	**0.764**	0.765	**0.016**	**0.710**	0.712
	RMSE	0.393	0.490	0.488	**0.361**	**0.919**	0.919	**0.340**	**0.864**	0.866
λ_1	Bias	0.047	0.026	0.028	**0.109**	**0.349**	0.344	**0.070**	**0.267**	0.268
	RMSE	0.170	0.173	0.172	**0.195**	**0.414**	0.410	**0.187**	**0.375**	0.372
λ_2	Bias	−0.104	−0.116	−0.114	−0.106	−0.055	−0.052	−0.110	−0.051	−0.048
	RMSE	0.174	0.184	0.182	0.171	0.163	0.163	0.173	0.158	0.157
λ_3	Bias	−0.044	−0.047	−0.044	−0.112	0.025	0.032	−0.097	0.027	0.026
	RMSE	0.146	0.148	0.150	0.180	0.171	0.180	0.168	0.165	0.161
λ_4	Bias	−0.091	−0.086	−0.083	−0.064	0.034	0.034	−0.081	0.011	0.009
	RMSE	0.165	0.162	0.162	**0.152**	**0.154**	0.155	0.156	0.144	0.144
λ_5	Bias	−0.107	−0.083	−0.082	−0.153	0.003	0.005	−0.138	0.033	0.037
	RMSE	0.197	0.194	0.192	0.221	0.182	0.181	0.214	0.190	0.191
γ_1	Bias	0.119	0.183	0.168	**0.113**	**0.301**	0.285	**0.236**	**0.723**	0.712
	RMSE	0.170	0.173	0.172	**0.195**	**0.414**	0.410	**0.187**	**0.375**	0.372
γ_2	Bias	−0.006	0.032	0.029	**−0.110**	**0.267**	0.269	**−0.098**	**0.233**	0.232
	RMSE	0.268	0.277	0.271	**0.280**	**0.393**	0.398	**0.274**	**0.365**	0.367
γ_3	Bias	0.096	0.104	0.124	**0.127**	**0.504**	0.516	**0.127**	**0.473**	0.472
	RMSE	0.313	0.307	0.322	**0.332**	**0.611**	0.632	**0.323**	**0.580**	0.578
γ_4	Bias	−0.122	−0.093	−0.089	−0.091	0.056	0.046	−0.176	−0.046	−0.054
	RMSE	0.277	0.269	0.264	0.267	0.265	0.263	0.295	0.240	0.237
γ_5	Bias	−0.059	−0.006	−0.011	**−0.079**	**0.087**	0.084	**−0.045**	**0.152**	0.161
	RMSE	0.284	0.298	0.285	**0.285**	**0.293**	0.289	**0.284**	**0.325**	0.333
θ^d	Bias	−0.002	−0.002	–	**0.011**	**0.013**	–	**0.017**	**0.019**	–
	RMSE	0.484	0.483	–	**0.443**	**0.577**	–	**0.379**	**0.586**	–
θ^h	Bias	−0.044	−0.045	−0.045	**−0.044**	**−0.047**	−0.045	**−0.044**	**−0.045**	−0.045
	RMSE	0.585	0.583	0.583	**0.581**	**0.593**	0.592	**0.574**	**0.592**	0.593
$\sigma_{\theta^d}^2$	Bias	0.008	0.011	–	**0.013**	**0.598**	–	**0.051**	**0.648**	–
	RMSE	0.001	0.001	–	**0.017**	**0.494**	–	**0.029**	**0.411**	–
$\sigma_{\theta^h\theta^d}$	Bias	−0.003	–	–	0.008	–	–	0.001	–	–
	RMSE	0.023	–	–	0.005	–	–	0.008	–	–

The boldfaced values indicate that much smaller Bias and RMSE are obtained from the model.

between θ_i^h and θ_i^d is 0. When the correlation between θ_i^h and θ_i^d increases, the bias and RMSE of η_1, $\boldsymbol{\beta}$, $\boldsymbol{\Sigma_I}$, and $\boldsymbol{\gamma}$ in the NMAR model are much smaller than those in the MAR and HO-DINA models. Moreover, for low dropping-out proportions, when the correlation between θ_i^h and θ_i^d increases, the bias of the person parameters of the three models changes very little,

TABLE 5 | Bias and RMSE of parameter estimates of three models with high dropping-out proportion under different correlations between θ_i^h and θ_i^d in simulation study III.

		$\rho = 0$			$\rho = -0.5$			$\rho = -0.8$		
Parameter	**Statistics**	**NMAR**	**MAR**	**HO-DINA**	**NMAR**	**MAR**	**HO-DINA**	**NMAR**	**MAR**	**HO-DINA**
η_0	Bias	−0.013	−0.019	–	**0.016**	**−0.221**	–	**0.014**	**−0.174**	–
	RMSE	0.134	0.132	–	**0.146**	**0.271**	–	**0.130**	**0.237**	–
η_1	Bias	−0.002	−0.003	–	**−0.001**	**−0.024**	–	**−0.001**	**−0.019**	–
	RMSE	0.012	0.011	–	**0.013**	**0.028**	–	**0.011**	**0.024**	–
β	Bias	−0.025	−0.021	−0.021	**−0.027**	**0.175**	0.177	**−0.010**	**0.187**	0.185
	RMSE	0.275	0.274	0.273	**0.284**	**0.383**	0.384	**0.267**	**0.373**	0.371
δ	Bias	0.058	0.071	0.069	0.060	−0.001	−0.003	0.043	−0.021	−0.019
	RMSE	0.392	0.405	0.404	**0.392**	**0.441**	0.443	**0.378**	**0.427**	0.425
μ_β	Bias	−0.142	−0.120	−0.124	−0.144	0.056	0.061	−0.126	0.071	0.067
	RMSE	0.235	0.225	0.228	0.240	0.217	0.218	0.227	0.216	0.215
μ_δ	Bias	0.091	0.089	0.092	0.093	0.035	0.028	0.075	0.011	0.015
	RMSE	0.234	0.252	0.253	**0.236**	**0.271**	0.272	**0.219**	**0.263**	0.216
σ_β^2	Bias	−0.032	−0.033	−0.032	**−0.012**	**0.084**	0.086	**−0.025**	**0.084**	0.085
	RMSE	0.302	0.302	0.302	**0.313**	**0.339**	0.339	**0.304**	**0.332**	0.334
$\sigma_{\beta\delta}$	Bias	−0.004	−0.013	−0.017	**−0.047**	**−0.336**	−0.339	**−0.004**	**−0.313**	−0.314
	RMSE	0.302	0.316	0.314	**0.322**	**0.505**	0.507	**0.309**	**0.485**	0.486
σ_δ^2	Bias	0.083	0.271	0.271	**0.118**	**0.802**	0.806	**0.037**	**0.738**	0.740
	RMSE	0.383	0.491	0.489	**0.405**	**0.960**	0.965	**0.378**	**0.898**	0.901
λ_1	Bias	0.047	0.027	0.021	**0.110**	**0.490**	0.502	**0.089**	**0.474**	0.467
	RMSE	0.182	0.181	0.184	**0.201**	**0.553**	0.566	**0.198**	**0.552**	0.544
λ_2	Bias	−0.110	−0.120	−0.122	−0.099	0.013	0.015	−0.102	0.006	0.003
	RMSE	0.182	0.190	0.191	**0.170**	**0.173**	0.174	0.174	0.164	0.163
λ_3	Bias	−0.055	−0.055	−0.055	−0.116	0.102	0.104	**−0.091**	**0.152**	0.144
	RMSE	0.156	0.158	0.156	**0.186**	**0.206**	0.207	**0.171**	**0.237**	0.232
λ_4	Bias	−0.096	−0.089	−0.092	**−0.074**	**0.098**	0.098	**−0.076**	**0.085**	0.084
	RMSE	0.170	0.167	0.168	**0.160**	**0.188**	0.188	**0.159**	**0.178**	0.177
λ_5	Bias	−0.077	−0.045	−0.051	−0.147	0.141	0.147	**−0.140**	**0.171**	0.164
	RMSE	0.196	0.197	0.195	**0.223**	**0.244**	0.247	**0.225**	**0.269**	0.263
γ_1	Bias	−0.133	0.174	0.186	**0.147**	**0.672**	0.720	**0.251**	**1.029**	0.995
	RMSE	0.374	0.390	0.400	**0.375**	**0.892**	0.952	**0.439**	**1.294**	1.249
γ_2	Bias	−0.020	0.059	0.058	**−0.102**	**0.334**	0.340	**−0.058**	**0.341**	0.333
	RMSE	0.287	0.302	0.294	**0.285**	**0.458**	0.461	**0.271**	**0.452**	0.444
γ_3	Bias	−0.091	0.117	0.117	**0.111**	**0.537**	0.532	**0.143**	**0.591**	0.584
	RMSE	0.328	0.330	0.328	**0.332**	**0.648**	0.644	**0.341**	**0.700**	0.693
γ_4	Bias	−0.130	−0.104	−0.102	−0.122	0.055	0.053	−0.146	0.026	0.026
	RMSE	0.289	0.280	0.277	0.290	0.271	0.270	0.294	0.265	0.261
γ_5	Bias	−0.014	0.046	0.039	**−0.078**	**0.147**	0.152	**−0.050**	**0.226**	0.213
	RMSE	0.304	0.321	0.314	**0.289**	**0.330**	0.334	**0.296**	**0.386**	0.376
θ^d	Bias	−0.002	−0.002	–	**0.011**	**0.016**	–	**0.018**	**0.021**	–
	RMSE	0.479	0.475	–	**0.442**	**0.549**	–	**0.374**	**0.526**	–
θ^h	Bias	−0.044	−0.045	−0.046	**−0.043**	**−0.046**	−0.045	**−0.044**	**−0.045**	−0.046
	RMSE	0.595	0.594	0.594	**0.590**	**0.607**	0.608	**0.584**	**0.607**	0.607
$\sigma_{\theta^d}^2$	Bias	0.001	0.001	–	**0.017**	**0.494**	–	**0.029**	**0.411**	–
	RMSE	0.088	0.085	–	**0.102**	**0.555**	–	**0.097**	**0.463**	–
$\sigma_{\theta^h\theta^d}$	Bias	0.023	–	–	0.005	–	–	0.008	–	–
	RMSE	0.102	–	–	0.092	–	–	0.085	–	–

The boldfaced values indicate that much smaller Bias and RMSE are obtained from the model.

whereas the RMSE of the person parameters in the MAR and HO-DINA models increases significantly. As expected, the NMAR model has higher accuracy of parameters than that of the other two models. Furthermore, the parameter estimates of the MAR and HO-DINA models are similar for all simulation conditions because θ_i^h and θ_i^d are uncorrelated in both the MAR and HO-DINA models, which ignore the not-reached items. **Table 4** shows the bias and RMSE of the parameters of the three models with

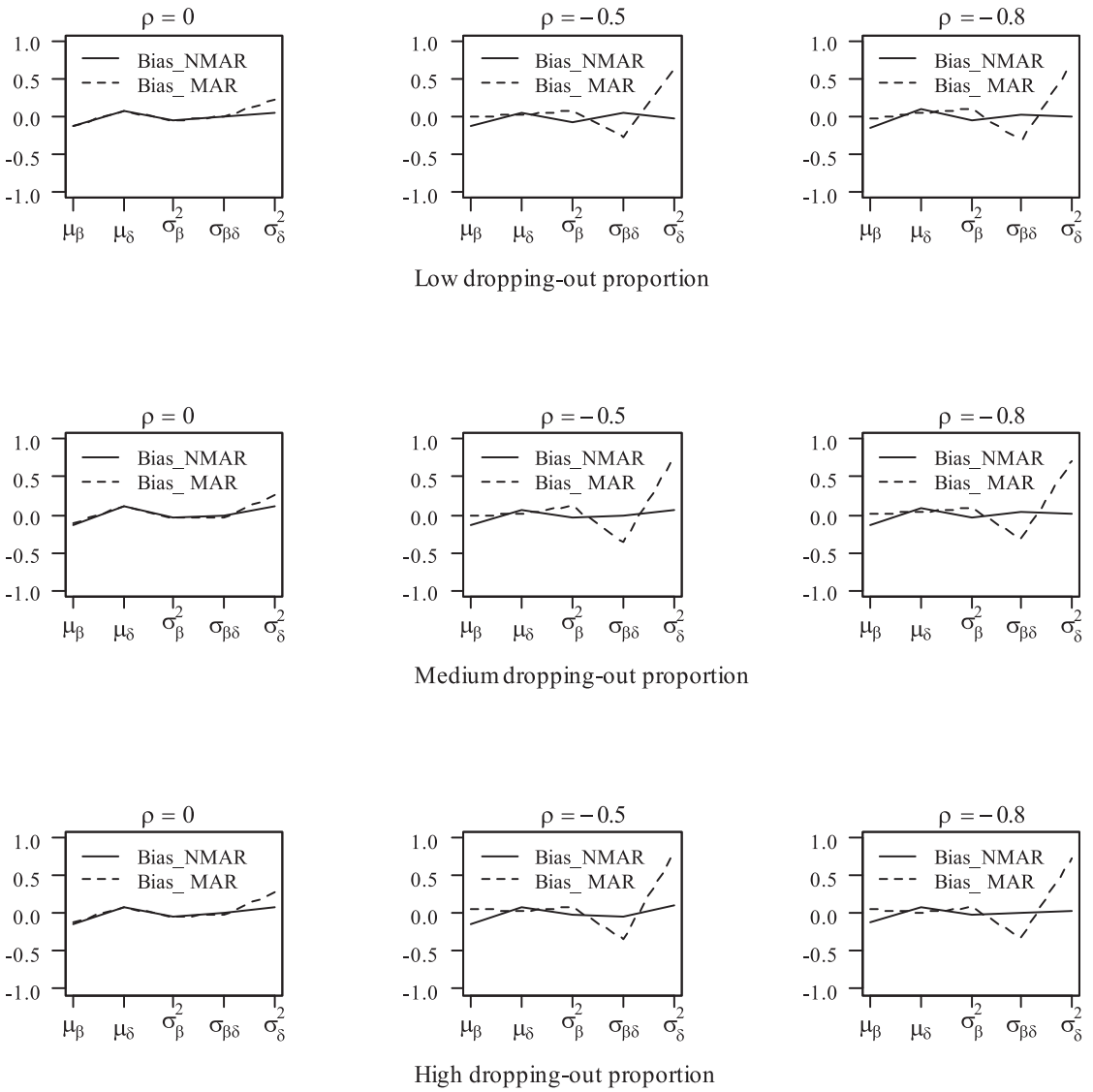


FIGURE 3 | Bias of parameter estimates in the mean item vector and the item covariance matrix elements under different dropping-out proportions and correlations between θ_i^h and θ_i^d in simulation study III. Note that the Bias_NMAR is the bias of parameter estimates in NMAR model, and Bias_MAR is the bias of parameter estimates in the MAR model.

medium dropping-out proportions under different correlations between θ_i^h and θ_i^d. Similar parameter estimates are obtained from the three models when the correlation between θ_i^h and θ_i^d is 0. When the correlation between θ_i^h and θ_i^d increases, not only the bias but also the RMSE of the person parameters are lower in the NMAR model than those in the MAR and HO-DINA models, and the other results are similar to those with low dropping-out proportions. **Table 5** shows the bias and RMSE of the parameters of the three models with high dropping-out proportions under different correlations between θ_i^h and θ_i^d. We find that the parameter estimates improve significantly with high dropping-out proportions. **Figure 3** shows the bias of the estimates of item mean vector and the item covariance matrix elements in the NMAR and MAR models under different dropping-out proportions and correlations between θ_i^h and θ_i^d. The results show that the estimates of the parameters are more accurate in the NMAR model than those in the MAR model when the correlation is increased. Moreover, it is observed that the bias of the parameters of the NMAR model is close to 0 as the correlation between θ_i^h and θ_i^d increases. In contrast, the bias of the parameters of the MAR model is significantly larger than that of the NMAR model. **Figure 4** shows the RMSE of the estimates of the item mean vector and the item covariance matrix elements in the NMAR and MAR models under different dropping-out proportions and correlations between θ_i^h and θ_i^d. The results show that the RMSE of the item mean vector in the NMAR model improves slightly than that in the MAR model. Moreover, the RMSE of the item covariance matrix elements shows significant

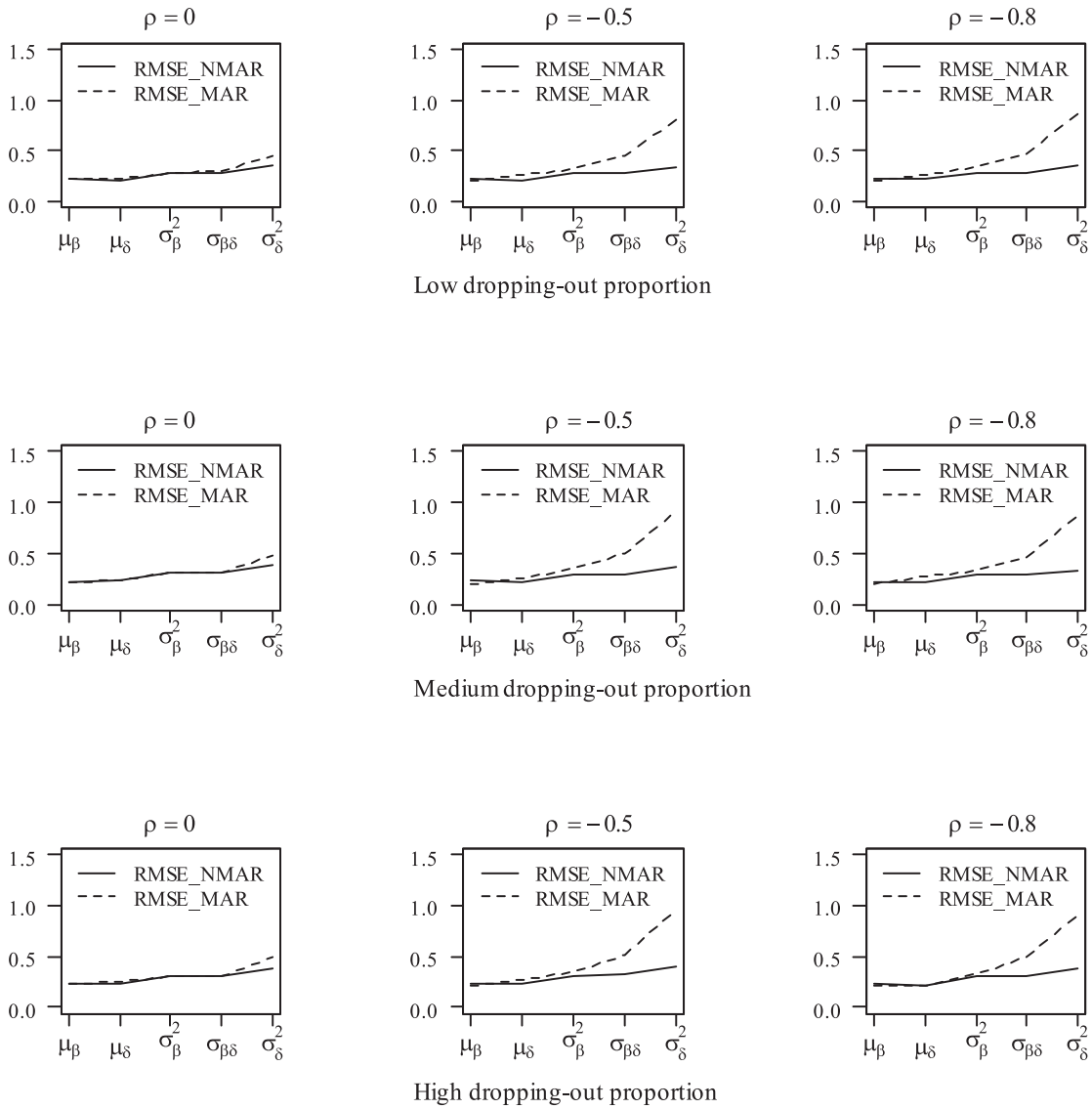


FIGURE 4 | RMSE of parameter estimates in the mean item vector and the item covariance matrix elements under different dropping-out proportions and correlations between θ_i^h and θ_i^d in simulation study III. Note that the Bias_NMAR is the bias of parameter estimates in the NMAR model, and Bias_MAR is the bias of parameter estimates in the MAR model.

improvements, and the estimates of the item covariance matrix elements are precise when the correlation is high. **Figure 5** shows the ACCRs and PCCRs under nine simulation conditions. Detailed results are provided in **Supplementary Table 1**. It is found that ACCRs and PCCRs in the NMAR model are improved significantly when the missing proportion or the correlation between θ_i^h and θ_i^d is high. This indicates that the MAR model could not recover the attribute pattern effectively when the missing data mechanism is indeed non-ignorable. **Table 6** shows the model selection results. The differences in DIC and LPML are not obvious when the correlation between θ_i^h and θ_i^d is 0. The DICs of the NMAR model are smaller than those of the MAR model under nine simulation conditions. Moreover, the LPMLs of the NMAR model are higher than those of the MAR model. Thus, the DIC and LPML indices are able to select the true model accurately.

REAL DATA ANALYSIS

This study analyzed one dataset from the computer-based PISA 2018 (OECD, 2021) mathematics cognitive test with nine items in Albania, which was also used in the study by Shan and Wang (2020). According to the PISA 2018 (OECD, 2021) mathematics assessment framework, four attributes belonging to the mathematical content knowledge were assessed: change and relationship (α_1), quantity (α_2), space and shape (α_3), and uncertainty and data (α_4). Item responses were coded 0

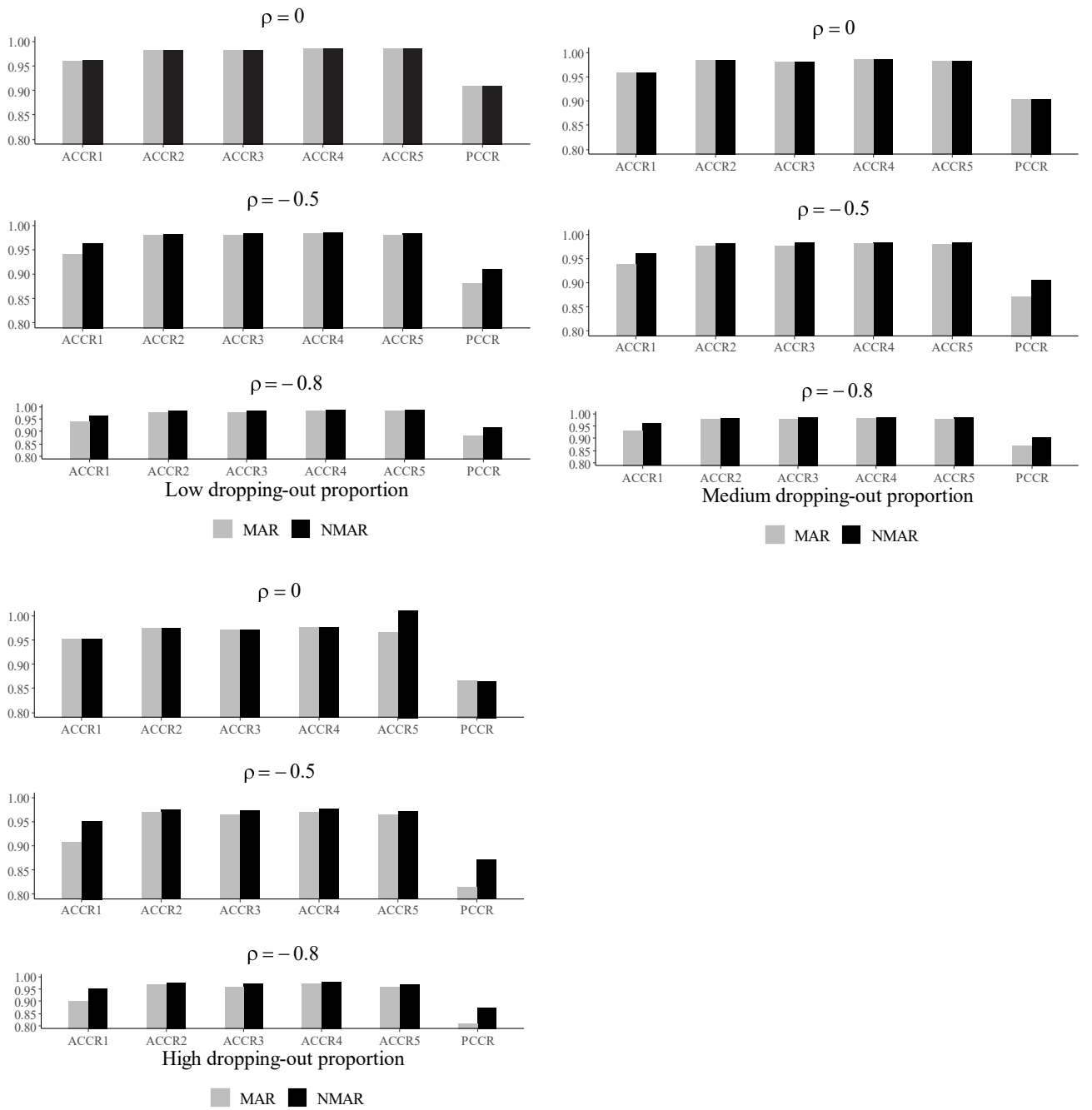


FIGURE 5 | The ACCRs and PCCRs of NMAR and MAR models under different correlations between θ_i^h and θ_i^d and different dropping-out proportions in simulation study III.

(no credit), 1 (full credit), 6 (not reached), 7 (not applicable), 8 (invalid), and 9 (nonresponse). There were 798 examinees after removing examinees with codes 7 (not applicable) and 8 (invalid). In addition, 224 examinees with code 9 were also removed from this study because this study mainly focused on dropping-out missingness. Thus, the final sample was 574. The overall not-reached proportion was about 2%, and the not-reached proportions at the item level were from 0.7% to 3.3%. The item IDs and Q matrices are presented in **Table 7**.

The DIC and LPML of the NMAR model in the real data were 5,760.28 and −3,040.03, respectively, and the DIC and LPML of the MAR model were 6,521.21 and −3,213.94, respectively. These two model fit indices indicated that the NMAR model fits the real data better than the MAR model. Thus, the NMAR model was adopted to fit this real dataset.

TABLE 6 | DICs and LPMLs of NMAR and MAR models under different correlations between θ_i^h and θ_i^d and different dropping-out proportions in simulation study III.

		Low dropping-out proportion		Medium dropping-out proportion		High dropping-out proportion	
		NMAR	**MAR**	**NMAR**	**MAR**	**NMAR**	**MAR**
$\rho = 0$	DIC	12139.3	12146.3	12283.9	12290.6	12084.8	12090.3
	LPML	−6348.4	−6352.7	−6465.8	−6468.3	−6532.1	−6539.9
$\rho = -0.5$	DIC	12152.6	12541.4	12225.5	12653.3	12113.8	12570.5
	LPML	−6354.7	−6592.1	−6431.9	−6660.7	−6539.8	−6747.6
$\rho = -0.8$	DIC	12132.3	12517.4	12215.6	12672.1	12029.8	12461.9
	LPML	−6333.8	−6579.2	−6412.4	−6663.1	−6476.2	−6681.6

TABLE 7 | The Q matrix in the real data.

Attribute	**CM033Q01**	**CM474Q01**	**CM155Q01**	**CM155Q04**	**CM411Q01**	**CM411Q02**	**CM803Q01**	**CM442Q02**	**CM034Q01**
α_1	0	0	1	1	0	0	0	0	0
α_2	1	0	0	0	0	0	0	0	1
α_3	0	1	0	0	1	0	0	1	0
α_4	0	0	0	0	0	1	1	0	0

TABLE 8 | Estimates and standard errors of the parameters for the real data.

Statistics	$\sigma_{\theta^h\theta^d}$	$\sigma^2_{\theta^d}$	μ_β	μ_δ	σ^2_β	$\sigma_{\beta\delta}$	σ^2_δ	λ_1	λ_2	λ_3	λ_4	γ_1	γ_2	γ_3	γ_4
Est.	−0.224	0.159	−1.749	2.380	3.058	−0.887	1.257	1.505	2.081	1.851	2.184	3.957	3.645	3.921	3.585
SD	0.149	0.040	0.379	0.292	2.108	1.241	0.979	0.399	0.427	0.443	0.382	0.441	0.432	0.446	0.482

Est. is the estimated value, SD is the standard deviation.

TABLE 9 | Estimates and standard errors of the item parameters for the real data.

Parameter	**Statistics**	**033Q01**	**474Q01**	**155Q01**	**155Q04**	**411Q01**	**411Q02**	**803Q01**	**442Q02**	**034Q01**
β_j	Est.	0.350	−0.251	−0.239	−1.213	−1.522	−1.296	−4.061	−4.325	−2.424
	SD	0.132	0.125	0.152	0.167	0.223	0.151	0.687	0.776	0.250
δ_j	Est.	2.433	1.418	3.265	1.559	2.541	0.781	3.485	3.218	2.326
	SD	0.520	0.225	0.561	0.280	0.396	0.323	0.755	0.801	0.371

Est. is the estimated value, SD is the standard deviation.

Tables 8, **9** show the estimated values and standard deviations of the item, person, and attribute parameters. Results show that the correlation coefficient of the person parameters is negative (i.e., −0.516), which indicates that the examinees with the higher abilities are less likely to drop out of the test. The estimated attribute slope parameters are positive, which implies that the knowledge attribute is better mastered with the increased ability θ_i^h. The item mean parameter μ_β is estimated to be −1.749, which shows that the mean guessing probability is approximately 0.15. In addition, for the estimation of item parameters, only β_j for CM033Q01 is positive, while the β_j values for other items are negative, which implies that the guessing probability of item CM033Q01 is higher than 0.5 and the guessing probability of all other items is lower than 0.5. All δ_j are positive, which satisfies $g_j < 1 - s_j$, as expected. **Supplementary Figure 1** shows the proportions of attribute patterns for examinees with not-reached items, which illustrate that the most prevalent attribute pattern for examinees with not-reached items is (0000), which is unsurprising.

CONCLUSION

Not-reached items occurred frequently in cognitive diagnosis assessments. Missing data could help researchers understand examinees' attributes, skills, or knowledge structures. Studies dealing with item nonresponses have used imputation approaches in cognitive diagnosis models, which may lead to biased parameter estimations. Shan and Wang (2020) introduced latent missing propensities of examinees for a cognitive diagnosis model that was governed by the potential category variables. However, their model did not distinguish the type of item nonresponses, which could result in inaccurate inferences regarding cognitive attributes and patterns.

In this study, a missing data model for not-reached items in cognitive diagnosis assessments was proposed. A DINA model was used as the response model, and a 1PLM was used as the missing indicator model. The two models were connected by two bivariate normal distributions for person parameters and item parameters. This new model was able to obtain more

fine-grained attributes or knowledge structure as diagnostic feedback for examinees.

Simulation studies were conducted to evaluate the performance of the MCMC algorithm using the proposed model. The results showed that not-reached items provide useful information for further understanding the knowledge structure of examinees. Additionally, the HO-DINA model for the cognitive diagnosis assessments explained examinees' cognitive processes, thus precise estimations of parameters were obtained from the proposed NMAR model. We compared the recovery of parameters under the two missing mechanisms, which revealed that the bias and RMSE of person parameters decreased significantly when using the proposed NMAR model when the missing proportion and the correlation of ability parameters were high. Moreover, considerable differences in the ACCRs and PCCRs between the NMAR and MAR models were found. With regard to model selection, the proposed NMAR model fitted the data better than the MAR model when the missing data mechanism was non-ignorable. The proposed NMAR model was successfully applied to the 2018 computer-based PISA mathematics data.

Several limitations of the study warrant mentioning, alongside future research avenues. First, this study only modeled not-reached items; however, examinees may skip the items in a cognitive test, which is another type of missing data that needs to be explored further. Second, missing data mechanisms in cognitive assessments may depend on individual factors, such as sex, culture, and race. In addition, different training and problem-solving strategies of examinees, and different school locations may also affect the pattern of nonresponses. Future studies can extend our model to account for the above-mentioned factors. Third, future studies could also incorporate the additional sources of process data, such as the response times, to explore the missing data mechanisms.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.oecd.org/PISA/.

AUTHOR CONTRIBUTIONS

LL completed the writing of the article. JL provided the original thoughts. LL and JL provided key technical support. JZ, JL, and NS completed the article revisions. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. 12001091), China Postdoctoral Science Foundations (Grant Nos. 2021M690587 and 2021T140108), and the Fundamental Research Funds for the Central Universities of China (Grant No. 2412020QD025).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.889673/full#supplementary-material

REFERENCES

Anselmi, P., Robusto, E., Stefanutti, L., and de Chiusole, D. (2016). An upgrading procedure for adaptive assessment of knowledge. *Psychometrika* 81, 461–482. doi: 10.1007/s11336-016-9498-9

Brooks, S. P., and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* 7, 434–455. doi: 10.1080/10618600.1998.10474787

Chen, M. H., Shao, Q. M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York, NY: Springer. doi: 10.1007/978-1-4612-1276-8

de Chiusole, D., Stefanutti, L., Anselmi, P., and Robusto, E. (2015). Modeling missing data in knowledge space theory. *Psychol. Methods* 20, 506–522. doi: 10.1037/met0000050

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *J. Educ. Meas.* 45, 343–362. doi: 10.1111/j.1745-3984.2008.00069.x

de la Torre, J. (2009). DINA model and parameter estimation: a didactic. *J. Educ. Behav. Stat.* 34, 115–130. doi: 10.3102/1076998607309474

de la Torre, J. (2011). The generalized DINA model frame work. *Psychometrika* 76, 179–199. doi: 10.1007/s11336-011-9207-7

de la Torre, J., and Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika* 69, 333–353. doi: 10.1007/BF02295640

Debeer, D., Janssen, R., and De Boeck, P. (2017). Modeling skipped and not-reached items using irtrees. *J. Educ. Meas.* 54, 333–363. doi: 10.1111/jedm.12147

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: the DINA model, classification, latent class sizes, and the Q-matrix. *Appl. Psychol. Meas.* 35, 8–26. doi: 10.1177/0146621610377081

Doignon, J. P., and Falmagne, J. C. (1999). *Knowledge Spaces*. New York:NY: Springer. doi: 10.1007/978-3-642-58625-5

Falmagne, J. C., and Doignon, J. P. (2011). *Learning Spaces: Interdisciplinary Applied Mathematics*. New York:NY: Springer. doi: 10.1007/978-3-642-01039-2

Finch, H. (2008). Estimation of item response theory parameters in the presence of missing data. *J. Educ. Meas.* 45, 225–245. doi: 10.1111/j.1745-3984.2008.00062.x

Fox, J. P. (2010). *Bayesian Item Response Modeling: Theory and Applications*. New York:NY: Springer. doi: 10.1007/978-1-4419-0742-4

Geisser, S., and Eddy, W. F. (1979). A predictive approach to model selection. *J. Am. Stat. Assoc.* 74, 153–160. doi: 10.1080/01621459.1979.10481632

Glas, C. A. W., and Pimentel, J. L. (2008). Modeling nonignorable missing data in speeded tests. *Educ. Psychol. Meas.* 68, 907–922. doi: 10.1177/0013164408315262

Heller, J., Stefanutti, L., Anselmi, P., and Robusto, E. (2015). On the link between cognitive diagnostic models and knowledge space theory. *Psychometrika* 80, 995–1019. doi: 10.1007/s11336-015-9457-x

Henson, R. A., Templin, J. L., and Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika* 74, 191–210. doi: 10.1007/s11336-008-9089-5

Holman, R., and Glas, C. A. W. (2005). Modelling non-ignorable missing-data mechanisms with item response theory models. *Br. J. Math. Stat. Psychol.* 58, 1–17. doi: 10.1111/j.2044-8317.2005.tb00312.x

Huisman, M. (2000). Imputation of missing item responses: Some simple techniques. *Q. Q.* 34, 331–351. doi: 10.1023/A:1004782230065

Ibrahim, J. G., Chen, M. H., and Sinha, D. (2001). *Bayesian Survival Analysis*. New York:NY: Springer. doi: 10.1007/978-1-4757-3447-8

Little, R. J. A., and Rubin, D. B. (2002). *Statistical Analysis With Missing Data*, 2nd Edn. New York:NY: Springer. doi: 10.1002/9781119013563

Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika* 39, 247–264. doi: 10.1007/BF02291471

Lord, F. M. (1983). Maximum likelihood estimation of item response parameters when some responses are omitted. *Psychometrika* 48, 477–482. doi: 10.1007/BF02293689

Lord, F. M., and Novick, M. R. (1968). *Statistical Theories Of Mental Test Scores*. Berlin: Addison-Wesley.

Lu, J., and Wang, C. (2020). A response time process model for not-reached and omitted items. *J. Educ. Meas.* 57, 584–620. doi: 10.1111/jedm.12270

Ludlow, L. H., and O'Leary, M. (1999). Scoring omitted and not-reached items: practical data analysis implications. *Educ. Psychol. Meas.* 59, 615–630. doi: 10.1177/0013164499594004

Ma, W. (2021). *A Higher-Order Cognitive Diagnosis Model With Ordinal Attributes For Dichotomous Response Data. Multivariate Behavioral Research*. Milton Park: Taylor & Francis. doi: 10.1080/00273171.2020.1860731

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika* 64, 187–212. doi: 10.1007/BF02294535

OECD (2009). *PISA 2006 Technical Report*. Paris: OECD Publishing.

OECD (2018). *PISA 2015 Technical Report*. Paris: OECD Publishing.

OECD (2021). *PISA 2018 Technical Report*. Paris: OECD Publishing.

Ömür Sünbül, S. (2018). The impact of different missing data handling methods on DINA model. *Int. J. Eval. Res. Educ.* 7, 77–86. doi: 10.11591/ijere.v1i1.11682

Patz, R. J., and Junker, B. W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics* 24, 146–178. doi: 10.3102/10769986024002146

Pohl, S., Ulitzsch, E., and von Davier, M. (2019). Using response times to model not-reached items due to time limits. *Psychometrika* 84, 892–920. doi: 10.1007/s11336-019-09669-2

Rasch, G. (1960). *Probabilistic Models For Some Intelligence And Attainment Tests*. Copenhagen Denmark: Danish Institute for Educational Research.

Rose, N., von Davier, M., and Nagengast, B. (2017). Modeling omitted and not-reached items in IRT models. *Psychometrika* 82, 795–819. doi: 10.1007/s11336-016-9544-7

Rose, N., von Davier, M., and Xu, X. (2010). *Modeling nonignorable missing data with IRT. Research Report No. RR-10-11*. Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333-8504.2010.tb02218.x

Rubin, D. B. (1976). Inference and missing data. *Biometrika* 63, 581–592. doi: 10.1093/biomet/63.3.581

Shan, N., and Wang, X. (2020). Cognitive diagnosis modeling incorporating item-level missing data mechanism. *Front. Psychol.* 11:564707. doi: 10.3389/fpsyg.2020.564707

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. Royal Stat. Soci. Series B* 64, 583–639. doi: 10.1111/1467-9868.00353

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *J. Educ. Meas.* 20, 345–354. doi: 10.1111/j.1745-3984.1983.tb00212.x

Templin, J. L., and Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychol. Methods* 11, 287–305. doi: 10.1037/1082-989X.11.3.287

von Davier, M. (2008). A general diagnostic model applied to language testing data. *Br. J. Math. Stat. Psychol.* 61, 287–307. doi: 10.1348/000711007X193957

von Davier, M. (2014). *The Log-Linear Cognitive Diagnostic Model As A Special Case Of The General Diagnostic Model. Research Report No. RR-14-40*. Princeton, NJ: Educational Testing Service. doi: 10.1002/ets2.12043

von Davier, M. (2018). Diagnosing diagnostic models: FromVon Neumann's elephant to model equivalencies and network psychometrics. *Meas. Int. Res. Pers.* 16, 59–70. doi: 10.1080/15366367.2018.1436827

Xu, G., and Shang, Z. (2018). Identifying latent structures in restricted latent class models. *J. Am. Stat. Assoc.* 113, 1284–1295. doi: 10.1080/01621459.2017.1340889

Xu, G., and Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika* 81, 625–649. doi: 10.1007/s11336-015-9471-z

Zhan, P., Jiao, H., and Liao, D. (2018). Cognitive diagnosis modelling incorporating item response times. *Br. J. Math. Stat. Psychol.* 71, 262–286. doi: 10.1111/bmsp.12114

Zhang, Z., Zhang, J., Lu, J., and Tao, J. (2020). Bayesian Estimation of the DINA Model with Pólya-Gamma Gibbs Sampling. *Front. Psychol.* 11:384. doi: 10.3389/fpsyg.2020.00384

Conflict of Interest: 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 | Frontiers in Psychology

TYPE Original Research
PUBLISHED 04 August 2022
DOI 10.3389/fpsyg.2022.946273

OPEN ACCESS

EDITED BY
Rosa Angela Fabio,
University of Messina, Italy

REVIEWED BY
Tindara Caprì,
Institute for Biomedical Research and Innovation (IRIB) (CNR), Italy
Fabrizio Stasolla,
Giustino Fortunato University, Italy

*CORRESPONDENCE
José Antonio Lozano-Lozano
jose.lozano@uautonoma.cl
Diana Martella
dmartella@uloyola.es

SPECIALTY SECTION
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

RECEIVED 17 May 2022
ACCEPTED 18 July 2022
PUBLISHED 04 August 2022

CITATION
Fioravante I, Lozano-Lozano JA and Martella D (2022) Attention deficit hyperactivity disorder: A pilot study for symptom assessment and diagnosis in children in Chile. *Front. Psychol.* 13:946273. doi: 10.3389/fpsyg.2022.946273

Attention deficit hyperactivity disorder: A pilot study for symptom assessment and diagnosis in children in Chile

Isabella Fioravante[1], José Antonio Lozano-Lozano[2]* and Diana Martella[3]*

[1]Escuela de Psicología, Pontificia Universidad Católica de Chile, Santiago, Chile, [2]Instituto de Ciencias Biomédicas, Instituto Iberoamericano Desarrollo Sostenible, Claustro Académico del Doctorado en Ciencias Sociales, Universidad Autónoma de Chile, Santiago, Chile, [3]Departamento de Psicología, Universidad Loyola de Andalucía, Sevilla, Spain

Background: Attention Deficit Hyperactivity Disorder (ADHD) is one of the most prevalent psychiatric disorders among school-age children and is characterized by varying degrees of inattention, hyperactivity, and impulsivity. Diagnosis, which currently relies on the DSM-V criteria, is complex. This research proposes an integrated procedure for ADHD diagnosis in children, improving the diagnostic process and scientific research on etiopathology.

Materials and methods: We conducted a clinical report on ADHD diagnosis in children (n = 92) between the ages of 8 and 13, based on the results of the application of different scales to parents of school-age children in Chile. The children were divided into two groups, those with an ADHD diagnosis (n = 44) and those without (n = 48) (24% females).

Results: The results revealed statistically significant differences between groups for scales EDAH y SDQ-Cas, Conners Comprehensive Behavior Scale, Conners Parent Scale and the criteria according to the DSM-V and its dimensions, with the exception of inattention.

Conclusion: The findings indicate the importance of appropriate criteria and procedures to establish a diagnosis and implement effective interventions in ADHD.

KEYWORDS

attention deficit hyperactivity disorder (ADHD), children diagnosis, symptomatologic assessment, ADHD in Chile, DSM-V, TDAH in Chile

Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a neurodevelopmental disorder characterized by symptoms of inattention, impulsivity, and hyperactivity (with subtypes hyperactive-impulsive, inattentive, or combined), which interfere with development and impact the individual's functional, personal, and social spheres (American Psychiatric Association, 2013), which could not be attributed to another neurological, sensory, language, or motor disorder. This is the main health problem that affects children, according to the most recent epidemiological data (Polanczyk et al., 2007, 2014; Thomas et al., 2015), and is the most frequent diagnosis in scholar children and adolescents, affecting their adaptation to the school environment. Children and young people with ADHD have a higher risk of school failures or delays, family conflicts, risk behaviors, substance abuse, among others (Leahy, 2017). Due to its clinical heterogeneity and the absence of a biological marker, the diagnosis of ADHD is currently complex (Drechsler et al., 2020; Martella et al., 2020; Sutcubasi et al., 2020).

According to the last review by Polanczyk et al. (2014), it is estimated that the worldwide prevalence rate of ADHD is between 6 and 7% among the population under age 18, slightly higher than the 5.3% estimated in another study from 2007 (Polanczyk et al., 2007). Although ADHD is configured as the most frequent neurodevelopmental disorder, few studies offer prevalence estimates regarding mental health pathologies such as ADHD in Chile. Thus, the benchmark study for Chile continues to be De la Barra et al. (2013) and was noted in the country's most recent National Plan for Mental Health (2017–2025) as the source of prevalence data, but it focuses only on the epidemiological aspect of the disorder, especially yielding information to children and adolescent mental health programmes in Chile, while a variety of strategies to detect, diagnose and treat pathologies in health and education areas have emerged (Chile Ministerio de Salud, 2008; Chile Ministerio de Educación, 2015; Reyes et al., 2019).

Rendering to that study, the prevalence of ADHD in Chile is 10.3% among children ages 4–18, with the highest prevalence in ages four to eleven (15.5% nationally and 18.7% in Santiago), representing one of the highest prevalence in the world (Uribe et al., 2019). The most prevalent subtype is the hyperactive-impulsive, showing no gender differences, and the most prevalent comorbidities are anxiety disorders and oppositional disorder. Some correlations are relevant to point: the perception of a good and functional family have a negative association with ADHD diagnosis, while maltreatment has a positive correlation (De la Barra et al., 2013). Importantly, the reported prevalence is significantly higher than the overall prevalence in Chile, currently estimated at 7% (Thomas et al., 2015), and is also important to note that children and adolescents diagnosed with ADHD have high rates of consultation of mental health services (50.9%), compared to those with other disruptive disorders (27.6%) and non-disruptive disorders (36.8%) (De la Barra et al., 2013).

An ADHD diagnosis mainly relies on the criteria established in the Diagnostic and Statistical Manual of the American Academy of Psychiatry (DSM) (American Psychiatric Association, 2013). Both the DSM and other diagnostic classification manuals primarily group symptoms based on criteria to help the professional group different disorders. Therefore, these manuals constitute non-dimensional, consensus descriptive classifications and are currently multiaxial, as is the case of DSM-V. In other words, they are organized around different diagnostic axes that allow additional information relevant to the principal diagnosis to be included. However, all criteria continue to be clinical and only descriptive, despite ongoing efforts to standardize them and increase their objectivity. In keeping with the fifth version of the DSM, symptoms are given greater emphasis than the dysfunctionality; that is, the importance of symptoms corresponding to the dysfunction is diminished (Rojas et al., 2018).

This shows that the thresholds for classifying a child's behavior as disproportionate will always be arbitrary to some degree; here factors such as the cultural norms of each context, an adequate knowledge of typical childhood development, and the expectations of parents and teachers all come into play. These differences led to discrepancies, which some authors have argued that the prevalence varies according to age and the number of symptoms observed by informants (McKeown et al., 2015). Moreover, importantly, this diagnostic procedure has been criticized for not allowing sufficient reliability or validity (Faraone et al., 2014). In this framework, there are currently studies that suggest differences in cognitive functioning depending on the ADHD subtype, which would mean that the associated clinical deterioration is heterogeneous (Rivera, 2016). For example, regarding attention processes, in the inattentive subtype of ADHD, the alteration of selective attention would be seen more frequently; while in the combined subtype sustained attention would be more affected. Due to the above, the gigantic methodological differences in the experimental studies make the revision of this matter an arduous task.

There is no standardized approach to integrating the multiple sources of information into an ADHD diagnosis. Nor does the recent fifth edition of the DSM offer this possibility, due to a lack of empirical data that would allow for the integration of all mechanisms that figure into a diagnosis (Martel et al., 2015). In fact, the increasing diagnosis of ADHD in children around the world has started many debates about the validity of the diagnosis process, including in the social sciences field, with debates especially concerning the diagnosis and treatment of ADHD from a children's behavior "medicalization" perspective (Rafalovich, 2008; Hinshaw and Scheffler, 2014; Reyes et al., 2019).

Various studies based on neuroimaging and electrophysiological measurements have supported the hypothesis of ADHD's neurobiological origin, although its exact etiology cannot yet be confirmed. It is also essential to consider that ADHD is a pathology with a high heritability rate, estimated at up to 80% (Faraone et al., 2014). Other studies using electroencephalography (EEG) technique have found controversial results, with no consensus on analysis of EEG frequency bands in ADHD subjects, and the likely reason for this lack of consistent results is the heterogeneity of ADHD subtypes and of tasks (Fabio et al., 2018). In Chile, some studies have presented evidence from the neurobiological aspects of ADHD (Aboitiz and Schröter, 2005; Aboitiz et al., 2012), and the more promising founds are related to deficit in the functioning of neurotransmitters, cerebral dysfunction in frontal structures and deficit in executive functions. Especially considering the clinical heterogeneity of ADHD children, they will probably exhibit a heterogeneous neuropsychological profile too (Fabio et al., 2018).

Many authors (e.g., Abad-Mas et al., 2017; Santana-Vidal et al., 2020) have proposed, even previous the DSM update from fourth to fifth edition (Barkley, 2009) that there are problems with the clinical application of the ADHD criteria. They refer especially to the extent of symptoms list and their operational definitions, calling for the need to review them, and most important, to integrate other measures specifically regarding the executive functions. Some authors have been interested in the hypothesis that children with ADHD have an underlying executive dysfunction, maybe due to an impairment of the automatic processing of basic skills (Martino et al., 2017), proposing that in addition to attention difficulties, there are other impairments that affects children with ADHD, such as memory, inhibition, and planning difficulties (Fabio, 2017).

Additionally, due to the heterogeneity of the disorder's clinical presentation and the absence of a biomarker, professionals often resort to diagnosis by exclusion after assessing for other comorbid pathologies that present similar behavioral manifestations, as reported in the study by De la Barra et al. (2013), mainly anxiety disorder and oppositional-defiant disorder. There is rarely a "pure" presentation of ADHD, yielding a high rate of comorbidity with other conditions that can hinder an initial diagnosis of ADHD (Fenollar-Cortés and Fuentes, 2016). Between 40 and 80% of those with ADHD present some type of comorbid association (Aìlvarez et al., 2013). Therefore, it is typical for presentation of ADHD to occur in conjunction with another disorder (Orjales, 2012; Roessner et al., 2016; Liu et al., 2021).

Following that the state of the art about ADHD limit it diagnosis to the solely clinical analysis, in Chile the disorder is diagnosed by qualified professionals (which could be a neurologist, psychiatrist, pediatrician, general physician, psychologist, teachers' differential behavior, or educational psychologist) in accordance with guidelines by the Ministry of Education and Ministry of Health (Decree number 170/2010). In short, it means that the diagnosis is made through the check of the criteria that encompasses the disorder.

The Chilean decree guidelines indicate that the diagnosis process follows a three steps protocol, which includes (1) the classification according to the most recent edition of diagnostic classification manuals, a (2) detection and assessing process based on criteria such Conners Test, and a (3) comprehensive diagnostic process including a diagnosis by exclusion review. To this point, it is important to note that the behavioral observation is made based on the Conners Test, which is a questionnaire that is widely disseminated on the internet and does not fulfill the international diagnostic recommendations (Santana-Vidal et al., 2020).

More specifically, the increase of ADHD diagnosis in Chile has become on 2000s, and led their incorporation into children's health plan named "*Habilidades para la vida*" (Chile Ministerio de Salud, 2008), which aims are prevent this type of disorder through a joint work between school and health services, and has been accompanied by the creation of devices such as the School Integration Program (PIE) in 2015 (Chile Ministerio de Educación, 2015). This leads to another relevant issue: the fact that the diagnosis rate of "special educational need" disorders (defined by "PIE") are the basis to a state subside to the schools, and since ADHD is one of them, they diagnosis can be used as a strategy to obtain additional economic resources, which has aroused different suspicions (Reyes et al., 2019). Surprisingly or not, the sophistication of these strategies has coincided with the sustained increase of ADHD prevalence rates on child and youth population (Uribe et al., 2019). In this work, ADHD is understood as a contingent pathology and a public and clinical health problem (Pelham et al., 2020), due to its transversal impact on the different areas of childhood development and the importance of clearly establishing its structure, etiology, and expression. This work seeks to point out a problem that has receive few attentions in Chile, and that constitutes a major problem around the world: the lack for an integrative and objective methodology for ADHD diagnosis, which could derive both in over and underdiagnosis. The current empirical scientific literature on this field, especially in Chile, is scarce and reveals the relevance of making this problem visible.

As a result, the aim herein was to propose a battery of instruments for an independent procedure of symptom analysis and diagnosis of ADHD: the Conners scales, EDAH, SDQ-Cas and the criteria established in the DSM-V. Furthermore, this work seeks to foster a discussion, especially in Chile, about the need for integrated diagnostic procedures in children, supported by the belief an objective, integrated diagnostic system is the best way to approach complex disorders such as ADHD. Such a system would contribute to standardized, enhanced diagnoses and also to scientific research into the etiological mechanisms of such disorders. Specifically in Chile, this work will contribute to present empirical evidence to support the claim to improve

the current ADHD assess standard, in line with the current worldwide research.

Materials and methods

This research relies on a pre-experimental design involving two non-randomized groups (ADHD and Normotypical) and a single measure (Chacón-Moscoso et al., 2008; Chacón-Moscoso et al., 2016).

Participants

Ninety-two boys (n = seventy) and girls (n = twenty-two), between the ages of eight and sixteen (M = 11.07; SD = 1.561) participated in the study. Children were recruited in Chilean public schools catering to populations with a similar socioeconomic status, belonging to three educational establishments in the city of Talca and one educational establishment in the city of Santiago (Chile). These schools were characterized to implement the Integration School Program (PIE), where children with ADHD have a previous diagnosis from a psychiatrist. The schools were selected from a database of schools participating in other research projects. The inclusion criteria for this study were: (a) a diagnosis of ADHD and (b) no history of cognitive impairment, brain trauma, neurological disease, physical disability, comorbid mental disorders (except oppositional defiant disorder), or learning disorders. For the neurotypical group, participants were matched by gender, age, and IQ scores (in the case of the ADHD group). Any children presenting symptoms that could indicate ADHD were excluded. All children had an IQ above the 75th percentile, according to the results of Raven's Colored and Progressive Matrices (Raven, 1976). The mean age and IQ scores of the two groups were not significantly different. For each child, a Hand Preference Index was assessed by means of a standard Lateral Preference Questionnaire.

According to the allocation criteria, 44 children were assigned to the ADHD group (Group 1) and 48 children to the neurotypical group (Group 2). No significant differences were found in most of the sociodemographic variables, except in the use of medication: χ^2 (1, N = 92) = 17.502, $p < 0.05$. **Table 1** shows the sociodemographic characteristics of the sample.

TABLE 1 Students and parents' characterization (N = 92).

Research variables	ADHD-G N = 44	N-G N = 48	Test	P
Children's gender			χ^2	0.07
Female	5 (11.4)	17 (35.4)		
Male	39 (88.6)	31 (64.6)		
Age			U	0.063
Mean (SD)	10.77(1.428)	11.33(1.642)		
Median (IQR)	10(1)	12(3)		
Min-max	9-15	8-16		
Childbirth			χ^2	0.146
Term	41 (93.2)	40 (83.3)		
Premature	3 (6.8)	8(16.7)		
Birth type			χ^2	0.883
Normal	19 (43.2)	20 (41.7)		
Cesarean section	25 (56.8)	28(58.3)		
Problem during birth			F	0.511
Yes	6 (13.6)	4(8.3)		
No	38 (86.4)	44 (91.7)		
Sleep disorder			χ^2	0.211
Yes	5 (11.4)	3 (6.3)		
No	37 (84.1)	45 (93.8)		
Medication use				
Yes	27 (61.4)	9 (18.8)	χ^2	0.001
No	17 (38.6)	39 (81.3)		
Mother education			U	0.625
Elementary/Middle	14 (31.8)	15 (31.3)		
High school	24(54.5)	23 (74.9)		
College	6 (13.6)	10 (20.8)		
Father education			U	0.122
Elementary/Middle	25 (56.8)	20 (41.7)		
High school	15 (34.1)	20 (41.7)		
College	4 (9.1)	8 (16.7)		

ADHD-G, ADHD group; N-G, neurotypical group; χ^2, Chi-square; U, Mann-Whitney U; F, Fisher test.

Measures

Standard instruments validated for a Spanish-speaking population were used, except for the anamnesis form and the ADHD diagnostic checklist, which the authors developed for the specific purposes of this study.

1. Anamnesis Record: an *ad hoc* instrument composed of nine elements, which records information provided by the parent/guardian regarding the birth history, development, and health of the child, and sociodemographic characteristics of the family unit.
2. Conners Scales (Conners, 1970, 1989, 1997; Farré and Narbona, 1989): the adaptations for the Spanish population of the Comprehensive Behavior Ratings Scale (α = 0.94; extended form with 48 items) and the Parents Rating Scale (α = 0.90; abbreviated form with ten items) adapted by Farré and Narbona (1989) were used. These scales collect reported information to identify behavioral changes and symptoms of ADHD.
3. Scale for the Evaluation of Attention Deficit Hyperactivity Disorder— EDAH-, its Spanish acronym (Farreí and Narbona, 2000; Belmar et al., 2015): validated for the

Chilean population (α = 0.95) by Belmar et al. (2015), it consists of 20 items and aims to assess the main features of ADHD and any coexisting behavioral disorders.

4. The SDQ-Cas questionnaire (Goodman, 1997; Brown et al., 2014): a study of psychometric properties among the Chilean population (α = 0.79; Brown et al., 2014), it consists of 25 items that gauge behaviors, emotions, and interpersonal interactions associated with psychological problems in children and adolescents. In addition, the impact supplement (on the reverse side of the questionnaire) enables professionals to ask parents if the child shows any type of problem covered in the scales, with another series of questions regarding chronicity, distress, social impairment, and the burden to others that behavioral problems can generate.
5. Attention deficit hyperactivity disorder diagnostic checklist: the ADHD diagnostic criteria defined in the DSM-V (American Psychiatric Association, 2013) and used by the Chilean Ministry of Health were applied. The criteria were converted into a table and applied as a checklist consisting of three sections (inattention, hyperactivity, other criteria) and the total (sum of presence/absence of all criteria).

Procedures

This study is part of a FONDECYT–project (1181472) and obtained ethical approval by the National Agency for Research and Development (ANID) of Chile. The Research Ethics Committee of the Autonomous University of Chile also approved the study (approval number 012–2019).

First, an invitation was extended to each school to participate in the research. Once the school principal had provided informed consent, the project was overseen by the research team in conjunction with the directors of technical-pedagogical units, school integration programs, or other pertinent professionals. Then, the parents of the children who were potential participants were contacted to respond to the battery of instruments used to characterize the children. After this stage of evaluation, the data were screened according to inclusion and exclusion criteria, and the pertinent statistical analyses were carried out.

Data analysis

The assumptions of normality and homogeneity of variance were verified, and all variables followed a non-normal distribution except for the EDAH and the SDQ-CAS scales.

Version 26.0 of the statistical package SPSS was used for the descriptive calculations and contrasts of means. And for the estimation of statistical power and effect size, the GPower version 3.1 package was used. To compare the means between the groups, a minimum significance level of 0.05 was considered. The confidence intervals in the estimates of the parameters were 95%. Normality assumptions were verified using the Shapiro-Wilk test (normal distribution assumed $p > 0.05$); linearity was checked (met when $p < 0.05$); and error independence was verified with the Durbin-Watson test (values between $1.5 < d < 2.5$ were considered adequate). Since not all assumptions were satisfactory, Spearman's bivariate correlations (ρ) were calculated. In those cases where the chi square cannot be applied, Fisher's exact test was used. The Mann-Whitney U statistic was used in all cases except those that fulfilled the assumption of normality and homoscedasticity, in which case Student's *t* statistic was used. Cohen's d was used to calculate the effect size based on the differences.

Results

In order to assess any significant differences between the ADHD group and the neurotypical group in each of the variables, a means comparison analysis was carried out. As seen on **Tables 2**, **3**, all study variables showed significant differences, except for the inattention dimension of DSM.

Likewise, a bivariate correlation analysis was performed using Spearman's test to examine differences between the instruments (see **Table 4**). The results show a statistically significant correlation between the instruments, except between the inattention dimension of the DSM-V and the dimensions of hyperactivity, other indicators, and the SDQ-Cas test.

Discussion

Attention deficit hyperactivity disorder has been the subject of a plethora of studies and reviews that have led to changes in its diagnosis and treatment over the years. Although its symptoms have been known for centuries, it has only been recognized as a pathology in children since the 1980s and adults since 2013 (APA). Since then, ADHD has become one of the most extensively studied—yet also one of the most controversial—disorders (Wolraich, 1999). Therefore, the role of those involved in the suspicion, diagnosis, and intervention in cases of ADHD becomes exceedingly relevant since they are in positions of power that allow core practices to be instilled and/or reinforced. Increasingly rigorous research on determining whether a child may have ADHD thus becomes crucial.

The results of the application of the clinical tests in this study demonstrate the discrimination capacity of the instruments used for the evaluation and diagnosis of ADHD. One interesting finding was the results of the SDQ-Cas, consistent with prior scientific literature indicating its ability to discriminate

TABLE 2 Differences between groups for scales means (excepting EDAH and SDQ-Cas) (*N* = 92).

Measure	Group	N	Rank	z	U	*p*	1-β	d
DSM inattention	ADHD-G	44	49.01	−0.886	945.500	0.376	0.058	0.0583
	N-G	48	44.20					
DSM hyperactivity	ADHD-G	44	57.19	−3.710	585.500	0.001	0.833	0.833
	N -G	48	36.70					
DSM other criteria	ADHD-G	44	55.86	−3.496	644.000	0.001	0.913	0.718
	N-G	48	37.92					
DSM total	ADHD-G	44	55.44	−3.083	662.500	0.002	0.871	0.667
	N-G	48	38.30					
Conners scale parents/custodians	ADHD-G	44	55.07	−2.952	679.000	0.003	0.873	0.670
	N-G	48	38.65					
Conners scale home behavior	ADHD-G	44	54.61	−2.792	699.000	0.005	0.824	0.625
	N-G	48	39.06					

ADHD-G, ADHD group; N-G, neurotypical group; U, Mann-Whitney U; 1-β, statistical power; d: effect size.

TABLE 3 Differences between groups for EDAH and SDQ-Cas—scales means (*N* = 92).

	ADHD-G *N* = 44		N-G *N* = 48						
	M	DE	M	DE	gl	t	*p*	1-β	d
EDAH	28.75	12.908	20.12	11.452	90	3.397	0.001	0.957	0.707
SDQ-CAS	25.82	4.962	21.58	5.119	90	4,022	0.001	0.983	0.793

ADHD-G: ADHD group; N-G, neurotypical group; *t*: *t* student; 1-β, statistical power; d, effect size.

TABLE 4 Correlation coefficients between instruments (Spearman's Rho).

	1	2	3	4	5	6	7	8
1. DSM-inattention	–							
2. DSM-hyperactivity	0.187	–						
3. DSM-other criteria	0.188	0.293**	–					
4. DSM-total	0.697**	0.638**	0.412**	–				
5. EDAH	0.417**	0.500**	0.215*	0.615**	–			
6. Conners-P/C	0.383**	0.436**	0.294**	0.516**	0.823**	–		
7. Conners-H/B	0.371**	0.431**	0.240*	0.537**	0.760**	0.839**	–	
8. SDQ-CAS	0.189	0.381**	0.246*	0.411**	0.508**	0.487**	0.645**	–

N = 92. Spearman's correlations are shown. **$p < 0.01$; *$p < 0.05$.

and suggesting it may be helpful during a diagnosis as a supplementary indicator of ADHD. In turn, the analyses of the detailed results of the DSM-V criteria showed that the hyperactivity dimensions and other indicators were statistically significant. This was not the case for inattention, which suggests that this indicator is not determinant in a diagnosis of the disorder and should be considered jointly with other markers that confirm a hypothesis of ADHD.

Finally, the correlational analyses revealed significant direct effects for the correlations between most of the instruments. This indicates the usefulness of combining the instruments to enhance the process of diagnosing ADHD, which, in turn, ratifies the importance of defining a diagnostic protocol.

By contributing to the discussion on the evaluation and diagnosis of ADHD, this article set out to demonstrate the importance of establishing appropriate mechanisms to ensure that children receive a correct assessment and diagnosis regardless of their sociodemographic characteristics. Based on the findings herein and scientific advances in this field, there is a need for a protocol that can render professional practice more effective and standardize it for children with ADHD.

Children's families and schools are generally the first to distinguish the symptoms of ADHD, which is why most of the instruments used for a diagnosis involve an initial assessment of the child's behavior by their teachers and parents (Garcia-Rosales et al., 2020). Also, it is likely that the high expectations of school performance, which parents and teachers place on children, increase the need to find clinical explanations for school failure (Santana-Vidal et al., 2020). However, according to different studies, parent-teacher agreement on ADHD symptoms has typically been low to moderate (Narad et al., 2015). In this sense, it is also important to acknowledge that the current assessment method has been criticized for a lack of diagnostic precision and even differences between the most widely used manuals (ICD-10; World Health Organization [WHO], 1992; and DSM-V).

In this regard, the relevance of the present study becomes apparent. The results reveal the need for an assessment alternative that allows for greater procedural objectivity and a diagnosis based on more integrated appraisals of ADHD symptoms. The new assessment process could incorporate, for example, experimental tests that can reduce

the time of an ADHD diagnostic procedure and increase its accuracy (Hall et al., 2016). Likewise, neuropsychological tasks can contribute and complement the behavioral measures (Santana-Vidal et al., 2020).

Diagnosing ADHD is a delicate task that is the subject of much debate around its etiology and, consequently, its symptoms. Therefore, it is a multifactorial disorder and needs to be addressed as such. Regarding future research, this work suggests that more investigation is needed into the changing diagnostic criteria of the main classification manuals, the evolution of how the disorder is conceptualized, and international differences in its assessment. This is crucial because one of the possible causes of overdiagnosis and underdiagnosis is the existence of ineffective instruments (Santana-Vidal et al., 2020). Moreover, the confounding criteria for the diagnosis of ADHD represent an issue that makes detection difficult, especially if it is carried out by professionals with little experience in the area (Ferrer-Urbina et al., 2017).

Children with ADHD represent a heterogeneous population and vary greatly in the degrees and severity of symptoms (Leahy, 2017). Follow-up studies with samples of ADHD children have showed that they have a higher-grade retention rate, more participation on special educational needs programs, school suspensions, more school expulsion, and lower academic performance, compared with control groups (Pi et al., 2018). This evidence makes such investigation as proposed in this present work all the more pressing.

Considering all that has been presented herein, the implication of this study relies on both clinical and practical areas. Regarding the clinical implications, on one hand, the construction of an integrative theoretical model for ADHD, incorporating hypotheses that support the biological, genetic, environmental, cognitive, and emotional factors that compose it, will impact directly on the comprehension and the handling of the disorder, for all those health professionals who have the power of the assessment and diagnosis process. On the other hand, regarding the practical implications, a change on the conception of ADHD diagnosis and the availability of a protocol to guide the practice will have a strong impact on the daily basis of several educational professionals and, therefore, on many families who put their trust in them.

Limitations

The present study presents some limitations in its execution. One of the main limitations of the study is the sample size which makes it difficult to generalize the findings obtained. Our procedure consisted of collecting data for 1 year. Initially we proposed to incorporate more schools, but due to the social unrest in Chile, the end of 2018, followed by the COVID-19 outbreak, it became impossible for us. Although the results obtained are representative of the schools being evaluated, in the immediate future we hope to incorporate more schools and consider including other age ranges, given that attentional capacity varies according to developmental stages (Milani et al., 2022).

Another limitation presented by the study is that the anamnesis form and the ADHD diagnostic checklist have not been validated for the Spanish-speaking population. For future developments, when the sample size allows, we will conduct validity and reliability studies. Specifically: validity evidence based on test content. Expert specialists will examine whether the proposed items are relevant, useful, and feasible (Chacón Moscoso et al., 2019). Evidence based on construct validity (Holgado-Tello et al., 2018) considering all the stages of validation (Muñiz and Fonseca-Pedrero, 2019). The developments of this line of work would allow the standardization of tests in Chilean educational contexts that facilitate the application of ADHD symptom evaluation protocols.

Conclusion

Because ADHD is the most common behavioral disorder of childhood, an appropriate and sensitive evaluation of symptoms is essential (Rostain et al., 2015). Even so, guidelines used for diagnosis of ADHD are not still rigorously applied, leading to an underdiagnosis or overdiagnosis of ADHD (Manos et al., 2017). The main finding of this study indicated that the application of the clinical tests to parents of children whit ADHD diagnosis appoints to the discrimination capacity of the instruments used for the evaluation of symptoms of ADHD. This first evaluation could be very relevant as a useful guide for clinicians in the diagnosis of ADHD.

Finally, the findings of this study will allow for the implementation of essential considerations in the assessment and diagnosis of children with ADHD and contribute to advancing the discussion in the scientific community.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Comité Ético Científico de la Universidad

Autónoma de Chile. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

The initial idea was generated by DM and IF and developed by all authors. IF collected all the information. JL-L analyzed the data. All authors wrote the manuscript, made a substantial contribution to the design of the document and improving both its writing and structure, consented to this final version for publication, and agreed to be responsible for all aspects of the work, such as the accuracy of the data and the completeness of the investigation.

Funding

This work was supported by the following grants of the National Agency for Research and Development (ANID-Chile): National PhD Fellowship (2022-21220562), FONDECYT (1181472) "The contribution of arousal and vigilance systems in Attention Deficit Hyperactivity Disorders (ADHD): an experimental study" and FONDECYT (1190945) "Methodological quality and effect size in intervention programs in Chile."

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abad-Mas, L., Caloca-Català, O., Mulas, F., and Ruiz-Andrés, R. (2017). Comparación entre el diagnóstico del trastorno por déficit de atención/hiperactividad con el DSM-5 y la valoración neuropsicológica de las funciones ejecutivas. *Rev. Neurol.* 64, 95–100.

Aboitiz, F., Ossandón, T., Zamorano, F., and Billeke, P. (2012). Balance en la cuerda floja: la neurobiología del trastorno por déficit atencional e hiperactividad. *Rev. Méd. Clín. Las Condes* 23, 559–565.

Aboitiz, F., and Schröter, C. (2005). Síndrome de déficit atencional: antecedentes neurobiológicosy cognitivos para estudiar un modelo de endofenotipo. *Rev. Chil. Neuro Psiquiatr.* 43, 11–16.

Aìlvarez, M., Soutullo, C., Dìez, A., and Figueroa, A. (2013). *TDAH Y Su Comorbilidad Psiquiaìtrica*. Spain: Universidad de Navarra.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Arlington, VA: American Psychiatric Association.

Barkley, R. A. (2009). Avances en el diagnóstico y la subclasificación del trastorno por déficit de atención/hiperactividad: qué puede pasar en el futuro respecto al DSM-V. *Rev. Neurol.* 48(Suppl. 2), 101–106. doi: 10.33588/rn.48S02.2009003

Belmar, M., Holgado, T., and Navas, M. (2015). Confiabilidad y validez de la Escala de Déficit Atencional (EDAH) adaptada en estudiantes chilenos. *Salud Mental* 38, 245–252. doi: 10.17711/SM.0185-3325.2015.034

Brown, P., Capella, C., and Antivilo, A. (2014). Propiedades psicomeìtricas de la versioìn para padres del strengths and difficultiesquestionnaire. *Rev. Psicol.* 23, 28–44. doi: 10.5354/0719-0581.2014.36146

Chacón Moscoso, S., Anguera Argilaga, M. T., Sanduvete Chaves, S., Losada López, J. L., and Portell Vidal, M. (2019). Methodological quality checklist for studies based on observational methodology (MQCOM). *Psicothema* 31, 458–464. doi: 10.7334/psicothema2019.116

Chacón-Moscoso, S., Shadish, W. R., and Cook, T. D. (2008). "Diseños evaluativos de intervención media [Evaluative designs of medium intervention]," in *Evaluación de programas sociales y sanitarios. Un abordaje metodológico*, eds M. T. Anguera, S. Chacón-Moscoso, and A. Blanco (Madrid: Sìntesis), 185–218.

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

Chile Ministerio de Educación (2015). *Programa De Integración Escolar PIE: Ley De Inclusión 20.845*. Santiago. Available online at: https://bit.ly/2sjzjL9 (accessed July 24, 2021).

Chile Ministerio de Salud (2008). *Guía clínica: Atención Integral De Niñas/ Niños Y Adolescentes Con Trastorno Hipercinético/Trastorno De La Atención (THA)*. Santiago. Available online at: https://bit.ly/2FndIt2 (accessed July 24, 2021).

Conners, C. K. (1970). Symptom patterns in hyperactivity, neurotic, and normal children. *Child Dev.* 41, 667–682.

Conners, C. K. (1989). *Conners Rating Scale*. Toronto, ON: Multi-Health Systems.

Conners, C. K. (1997). *Conners' Rating Scales-Revised (CRS-R): Technical Manual*. North Tonawanda, NY: Multi-Health Systems, Inc.

De la Barra, F., Vicente, B., Saldivia, S., and Melipillaìn, R. (2013). Epidemiology of ADHD in Chilean children and adolescents. *Atten. Def. Hyp. Disord.* 5, 1–8. doi: 10.1007/s12402-012-0090-6

Drechsler, R., Brem, S., Brandeis, D., Grünblatt, E., Berger, G., and Walitza, S. (2020). ADHD: current concepts and treatments in children and adolescents. *Neuropediatrics* 51, 315–335.

Fabio, R. (2017). The study of automatic and controlled processes in ADHD: a reread and a new proposal. *Mediterr. J. Clin. Psychol.* 5, 2–34. doi: 10.6092/2282-1619/2017.5.1507

Fabio, R., Caprì, T., Mohammadhasani, N., Gangemi, A., Gagliano, A., and Martino, G. (2018). Frequency bands in seeing and remembering: comparing ADHD and typically developing children. *Neuropsychol. Trends* 2018:24. doi: 10.7358/neur-2018-024-fabi

Faraone, S., Bonvicini, C., and Scassellati, C. (2014). Biomarkers in the diagnosis of ADHD – promising directions. *Curr. Psychiatry Rep.* 16:497. doi: 10.1007/s11920-014-0497-1

Farré, A., and Narbona, J. (1989). Índice de hiperquinesia y rendimiento escolar: validación del cuestionario de Conners en nuestro medio. *Acta Pediátr. Esp.* 47, 103–109.

Farreí, A., and Narbona, J. (2000). *Escalas Para La Evaluacioìn Del Trastorno Por Deìficit De Atencioìn Con Hiperactividad*. Madrid: TEA Ediciones.

Fenollar-Cortés, J., and Fuentes, L. J. (2016). The ADHD concomitant difficulties scale (ADHD-CDS), a brief scale to measure comorbidity associated to ADHD. *Front. Psychol.* 7:871. doi: 10.3389/fpsyg.2016.00871

Ferrer-Urbina, R., Chavéz, K. A., Gallardo, C., Loredo, G. Y., and Meneses, K. (2017). Apreciación diagnóstica de profesores del Trastorno de déficit atencional con hiperactividad (TDAH) en escolares de Enseñanza Básica pertenecientes a establecimientos municipalizados de la ciudad de Arica. *Salud Sociedad* 8, 52–65. doi: 10.22199/S07187475.2017.0001.00004

Garcia-Rosales, A., Vitoratou, S., Faraone, S., Rudaizky, D., Banaschewski, T., Asherson, P., et al. (2020). Differential utility of teacher and parent-teacher combined information in the assessment of attention deficit/hyperactivity disorder symptoms. *Eur. Child Adolesc. Psychiatry* 30, 143–153. doi: 10.1007/s00787-020-01509-4

Goodman, R. (1997). SDQ-CAS. The strengths and difficulties questionnaire: a research note. *J. Child Psychol. Psychiatry* 38, 581–586. doi: 10.1111/j.1469-7610.1997.tb01545.x

Hall, C., Selby, K., Guo, B., Valentine, A., Walker, G., and Hollis, C. (2016). An objective measure of attention, impulsivity and activity reduces time to confirm ADHD diagnosis in children: a completed audit cycle. *Child Adolesc. Mental Health* 21, 175–178.

Hinshaw, S., and Scheffler, R. (2014). *The ADHD Explosion: Myths, Medication, Money, And Today's Push for Performance*. New York, NY: Oxford University Press.

Holgado-Tello, F. P., Morata-Ramírez, M. Á, and Barbero-García, M. I. (2018). Confirmatory factor analysis of ordinal variables: h simulation study comparing the main estimation methods. *Av. Psicol. Latinoam.* 36, 601–617. doi: 10.12804/revistas.urosario.edu.co/apl/a.4932

Leahy, L. (2017). Attention-deficit/hyperactivity disorder: a historical review (1775 to Present). *J. Psychosoc. Nurs.* 55, 10–16. doi: 10.3928/02793695-20170818-08

Liu, J., He, Y., Shen, Y., Zhou, Y., Meng, T., Xiao, B., et al. (2021). Association of attention deficit/hyperactivity disorder with events occurring during pregnancy and perinatal period. *Front. Psychol.* 12:707500. doi: 10.3389/fpsyg.2021.707500

Manos, M. J., Giuliano, K., and Geyer, E. (2017). ADHD: overdiagnosed and overtreated, or misdiagnosed and mistreated? *Cleve Clin. J. Med.* 84, 873–880.

Martel, M., Nikolas, M., Schimmack, U., and Nigg, J. (2015). Integration of symptom ratings from multiple informants in ADHD diagnosis: a psychometric model with clinical utility. *Am. Psychol. Assoc.* 27, 1060–1071. doi: 10.1037/pas0000088

Martella, D., Aldunate, N., Fuentes, L. J., and Sánchez-Pérez, N. (2020). Arousal and executive alterations in attention deficit hyperactivity disorder (ADHD). *Front. Psychol.* 11:1991. doi: 10.3389/fpsyg.2020.01991

Martino, G., Caprí, T., Castriciano, C., and Fabio, R. A. (2017). Automatic deficits can lead to executive deficits in ADHD. *Mediterr. J. Clin. Psychol.* 5, 1–32. doi: 10.6092/2282-1619/2017.5.1669

McKeown, R. E., Holbrook, J. R., Danielson, M. L., Cuffe, S. P., Wolraich, M. L., and Visser, S. N. (2015). The impact of case definition on attention-deficit/hyperactivity disorder prevalence estimates in community-based samples of school-aged children. *J. Am. Acad. Child Adolesc. Psychiatry* 54, 53–61. doi: 10.1016/j.jaac.2014.10.014

Milani, A., Pascual-Leone, J., and Arsalidou, M. (2022). Converging evidence for domain-general developmental trends of mental attentional capacity: validity and reliability of full and abbreviated measures. *J. Exp. Child Psychol.* 222:105462. doi: 10.1016/j.jecp.2022.105462

Muñiz, J., and Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test [Ten steps for test development]. *Psicothema* 31, 7–16. doi: 10.7334/psicothema2018.291

Narad, M., Garner, A., Peugh, J., Tamm, L., Antonini, T., Kingery, A., et al. (2015). Parent-teacher agreement on ADHD symptoms across development. *Psychol. Assess.* 27, 239–247. doi: 10.1037/a0037864

Orjales, I. (2012). "Trastorno por deìficit de atencioìn con hiperactividad (TDAH): impacto evolutivo e intervencioìn," in *Alteraciones Del Desarrollo Y Discapacidad. Trastornos del desarrollo*, ed. A. Brioso (Madrid: Sanz y Torres), 91–146.

Pelham, W., Page, T. F., Altzuler, A. R., Gnagy, E. M., Molina, B. E. M., and Pelham, W. E. (2020). The long-term financial outcome of children diagnosed with ADHD. *J. Consult. Clin. Psychol.* 88, 161–170. doi: 10.1037/ccp0000461

Pi, M., Larraguibel, M., Rojas-Andrade, R., and Aldunate, C. (2018). Comparative study of adolescents with and without ADHD. (2018). *Salud Mental* 41, 287–296. doi: 10.17711/SM.0185-3325.2018.041

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

Polanczyk, G., De Lima, M., Horta, B., Biederman, J., and Rohde, L. (2007). The worldwide prevalence of ADHD: a systematic review and metaregression analysis. *Am. J. Psychiatry* 164, 942–948. doi: 10.1176/ajp.2007.164.6.942

Rafalovich, A. (2008). *Framing ADHD Children: A Critical Examination Of The History, Discourse, And Everyday Experience Of Attention Deficit/Hyperactivity Disorder*. New York, NY: Lexington Books.

Raven, J. (1976). *Test de Matrices Progresivas*. Barcelona: Editorial Paidoìs.

Reyes, P., Cottet, P., Jimenez, A., and Jauregui, G. (2019). Rethinking medicalization: discursive positions of children and their caregivers on the diagnosis and treatment of ADHD in Chile. *Saúde Sociedade* 28, 40–54. doi: 10.1590/S0104-12902019181141

Rivera, F. (2016). La elevada prevalencia del TDAH: posibles causas y repercusiones socioeducativas. *Psicol. Educ.* 22, 81–85.

Roessner, V., Banaschewski, T., Becker, A., Buse, J., Wanderer, S., Buitelaar, J. K., et al. (2016). Familiality of co-existing ADHD and tic disorders: evidence from a large sibling study. *Front. Psychol.* 7:1060. doi: 10.3389/fpsyg.2016.01060

Rojas, S., Rojas, P., and Penþa, M. (2018). "From problematic children to problematic diagnosis: the paradoxical trajectories of child and adolescent ADHD in Chile," in *Global Perspectives on ADHD: Social Dimensions of Diagnosis and Treatment in Sixteen Countries*, eds M. R. Bergey, A. M. Filipe, P. Conrad, and I. Singh (Baltimore, MA: Johns Hopkins University Press), 310–321.

Rostain, A., Jensen, P. S., Connor, D. F., Miesle, L. M., and Faraone, S. V. (2015). Toward quality care in ADHD: defi ning the goals of treatment. *J. Atten. Disord.* 19, 99–117.

Santana-Vidal, P., Gatica-Ferrero, S., and Valdenegro-Fuentes, L. (2020). Evidencia de sobrediagnóstico en el TDAH en base a evaluación neuropsicológica: un estudio en escolares chilenos. *Psicogente* 23, 1–20. doi: 10.17081/psico.23.44.3587

Sutcubasi, B., Metin, B., Kurban, M. K., Metin, Z. E., Beser, B., and Sonuga-Barke, E. (2020). Resting-state network dysconnectivity in ADHD: a system-neuroscience-based meta-analysis. *World J. Biol. Psychiatry* 21, 662–672. doi: 10.1080/15622975.2020.1775889

Thomas, R., Sanders, S., Doust, J., Beller, E., and Glasziou, P. (2015). Prevalence of attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. *Pediatrics* 135, e994–e1001. doi: 10.1542/peds.2014-3482

Uribe, P., Abarca-Brown, G., Radiszcz, E., and López-Contreras, E. (2019). ADHD and gender: subjective experiences of children in Chile. *Saúde Sociedade* 28, 75–91. doi: 10.1590/S0104-12902019181144

Wolraich, M. (1999). Attention deficit hyperactivity disorder: the most studied and yet most controversial diagnosis. *Dev. Disabil. Res. Rev.* 5, 163–168. doi: 10.1002/(SICI)1098-277919995:3<163::AID-MRDD1<3.0.CO;2-T

World Health Organization [WHO] (1992). *The ICD-10 Classification Of Mental And Behavioral Disorders: Clinical Descriptions And Diagnostic Guidelines*. Geneva: World Health Organization.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 28 September 2022
DOI 10.3389/fpsyg.2022.948745

OPEN ACCESS

EDITED BY
Salvador Chacón-Moscoso,
Sevilla University, Spain

REVIEWED BY
Morteza Taheri,
Imam Khomeini International University, Iran
Anjani Devi Chintagunta,
Vignan's Foundation for Science, Technology and Research, India

*CORRESPONDENCE
Inmaculada Herranz Aguayo
inmaculada.Herranz@uclm.es
José Luis Gómez Ramos
joseluis.gomez@uclm.es

SPECIALTY SECTION
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

RECEIVED 20 May 2022
ACCEPTED 24 August 2022
PUBLISHED 28 September 2022

CITATION
Díaz-García O, Herranz Aguayo I, Fernández de Castro P and Ramos JLG (2022) Lifestyles of Spanish elders from supervened SARS-CoV-2 variant onwards: A correlational research on life satisfaction and social-relational praxes. *Front. Psychol.* 13:948745. doi: 10.3389/fpsyg.2022.948745

Lifestyles of Spanish elders from supervened SARS-CoV-2 variant onwards: A correlational research on life satisfaction and social-relational praxes

Orlanda Díaz-García[1], Inmaculada Herranz Aguayo[2]*, Patricia Fernández de Castro[1] and José Luis Gómez Ramos[3]*

[1]Department of Labour Law and Social Work, Faculty of Social Sciences of Talavera de la Reina, University of Castilla-La Mancha, Ciudad Real, Spain, [2]Department of Philosophy, Anthropology and Sociology, Faculty of Social Sciences of Talavera de la Reina, University of Castilla-La Mancha, Ciudad Real, Spain, [3]Pedagogy Department, Faculty of Education at Albacete, University of Castilla-La Mancha, Ciudad Real, Spain

Objective: This study aimed to analyze the influence and measurement of the relationship and interaction between the elderly lifestyles after the appearance of the SARS-CoV-2 variant and the factors analyzed comprised life satisfaction levels, social relationships, and daily-life activities.

Methods: The study population was ≥ 65 in Castile-La Mancha ($N = 390{,}221$). The research design was quantitative and arose from primary data collected via an ad hoc survey carried out through the Computer Assisted Telephone Interview system by randomly stratified sampling. The sample size was made up of 1,065 cases, and the participants were selected through a random sampling stratified by gender quotas (55.40% women; 44.60% men), age ($\bar{x} = 76.56$), province, and habitat size.

Results: The results obtained revealed two main lifestyles, from which a notable behavioral change in personal relationships led to infer toward alternative lifestyles.

Conclusion: Notwithstanding the variation in lifestyles of the elderly after the pandemic, certain relationships remained unaltered. Thus, from the analyzed variables, relatives and friends relationships were scarcely influenced by the supervened incident.

KEYWORDS

elderliness, lifestyles, satisfactoriness, leisureliness, resocialization

Introduction

The aging of the population is yet a consolidated trend worldwide, having been estimated that the number of older adults will increase considerably around the world in the nearest future, being even presumed to double by 2050 (United Nations, 2019). In this regard, the population aged 65 and over comprised 19.43% of the total inhabitants in Spain in 2020 (Instituto Nacional de Estadística, 2020), with an increasing rate of migrants (Ciobanu et al., 2017). The boundary of ≥ 65-year-olds is set in this study because it is the current official age in Spain at which

citizens' retirement starts. Thus, this transition from professional activity to leisureliness is associated with significant social, economic, and relational changes in people's lives. Apropos of such a trend, the Instituto Nacional de Estadística (2020) estimates that this age range faction will reach 31.44% by 2050. The global aging-raising trend has also favored the rapid increase of research on discovering the factors influencing optimal aging revealed through lifestyles in humankind. To Rowe and Kahn (1997), successful aging comprises the absence of disease, the preservation of functional skills, and the ongoing commitment to life, all interrelated factors toward achieving pleasant aging leading to specific activities (Rowe and Kahn, 1997; Herranz et al., 2013; Bulow and Söderqvist, 2014; Urtamo et al., 2019).

Due to migration facts and that "[t]here will be two billion people in old age category worldwide by 2050" (Pant and Subedi, 2020, p. 32), not only does pleasant aging needs to be linked to pleasing-physical wellbeing but also does the *idiosyncrasy* and the sense of living a meaningful life (VanderWeele, 2017; Steptoe and Fancourt, 2019). If we ponder the term from a heterogeneous perspective, satisfactory livingness remains the most common concept in the literature to measure the subjective wellbeing of the superannuated (Fagerström et al., 2007; Ju, 2017; Tomioka et al., 2017; Etxeberría et al., 2019; Steptoe and Fancourt, 2019). Thus, to spot how the ab ovo stay-at-home to prevent further spreading of SARS-CoV-2 influenced lifestyles, we analyzed the concept of successful aging from the type and frequency of daily activities carried out by people aged 65 and above. Congruously, this investigation also adopts a social perspective and acknowledges the social relationships established by the analyzed sample with their homonyms and heteronyms.

The myth of elderly's lifestyles

Just as the lifestyles established from interpersonal relationships are linked to the type and frequency of daily life and the activities carried out by people (Geithner and Wagner, 2022), the dimension of life satisfaction (i.e., economic situation; sexual/affective relations; health; leisure, appearance, personal abilities; family, friends, or colleagues; spare time) is mirrored in the multiple aspects of social life and the distinct areas of personal joy (Fagerström et al., 2007; Von Humboldt and Leal, 2017; Gumà and Arpino, 2021). From the empirical evidence about the connection between lifestyles and satisfaction with life, it is observed that the higher the satisfaction with life, the more the other positive life aspects increase and vice versa (Toepoel, 2013; Gutiérrez et al., 2014; Vozikaki et al., 2017). Indeed, incomes also influence lifestyles, social-relational praxes, and behaviors (Xue et al., 2021).

An example of the above can be observed in Gumà and Arpino's (2021) research, where participants' satisfaction with life is more significant when their occupational contribution (domestic or professional) increases. To these authors, such satisfaction levels become more notable when *wealthiness* and *healthiness* concur. Other factors influencing the significant relationships among life-satisfied elders are those linked to the yet mentioned health, financial situation, family, and social contacts or relationships (Fagerström et al., 2007; Schnettler et al., 2014; de Albuquerque Araújo et al., 2020; Zadworna, 2022). To Pérez-Escoda (2013), when a correlation is established between the different aspects of people's satisfaction with life (family, health, work, and personal relationships), a determined quantitative weight is established for each bounding aspect.

Yet, not only the significant correlational relationship between the different aspects of life satisfaction in elderliness have been studied but such factors are also primarily studied from those perspectives associated with lifestyles and personal relationships (Molero and Pérez-Fuentes, 2011; Sampaio and Ito, 2013; Marques et al., 2016; Von Humboldt and Leal, 2017; Han et al., 2022; Penton et al., 2022), as much as from the perspective of indoor and outdoor activities (Silverstein and Parker, 2002; Triadó et al., 2009; Lee et al., 2014; Aponte, 2015; Santaella and Bohórquez, 2017; Vozikaki et al., 2017; Li et al., 2021; Chen et al., 2022; Geithner and Wagner, 2022). Similarly, the effect of personal relationships, particularly with family and friends, on life satisfaction is widely documented (Cornwell et al., 2008; Cheng et al., 2009; Litwin and Shiovitz-Ezra, 2011; Rafnsson et al., 2015; Burholt et al., 2020; LaBorde and Williams, 2022). From the literature, another aspect worth to be mentioned is that adults get higher satisfaction from life as they have more relatives than friends (Tomini et al., 2016).

Materials and methods

The research goal consists of analyzing the influence and the relational and interactional connections between the three dimensions studied and considered fundamental in the lifestyle scrutiny of 65-year-olds onwards in the Autonomous Region of Castile-La Mancha, Spain. To say, *satisfaction levels* (D_1), *social relationships* (D_2), and *daily living activities* (D_3). From the referenced goal, a dyad of objectives emerges for the in-depth exploration of the grouped variables within each of the dimensions concerning the correlational analysis (O_1) and the enquiring on the influence and the interaction of groups between the levels of satisfaction, the frequencies in the social relations, and the frequency of activities as variables might have been originating different lifestyles (O_2).

Instruments

The instrument for data collection is an ad hoc questionnaire implemented as a survey on a Spanish sample (Table 1). The

duration of the research lasted 5 months (January to May 2021). The main categories and dimensions of the questionnaire, which are related to the objectives of this study, were the following: profile and sociodemographic variables (CD_1); living conditions (CD_2); household structure (CD_3); life habits (CD_4); socioeconomic indicators (CD_5); health and impact of the COVID-19 pandemic (CD_6); dependency and disability (CD_7); use of time (CD_8); leisure and culture (CD_9); use of information and communication technologies (CD_10); physical activity (CD_11); values and attitudes (CD_12); expectations and interests (CD_13); and perception and use of specialized services (CD_14).

Thus, because of its proximity to research goals, the results presented in this study constitute the analysis of just three of the dimensions contemplated from the full version of the questionnaire: *life satisfaction levels* (D_1); *interpersonal relationships* (D_2); and *type and frequency of day-to-day activities* (D_3). The dimensional components are detailed as follows:

TABLE 1 Data sheet.

Location	Castile-La Mancha, Spain
Population	≥65-year-olds (N = 390,221 [to date January 1, 2020])
Methodology	Computer Assisted Telephone Interviewing (CATI) survey
Sample size	n = 1,065 individuals/cases from the target population
Confidence interval	95%
Margin of error for the estimate	3%
Sampling distribution and quotas	Proportional sampling distribution by provinces. Representative survey process, stratified by sex quotas, age group, and habitat size
Fieldwork schedule	From January 5 to January 22, 2021
Questionnaire preparation time	15 min

D_1: Satisfaction with different aspects of life

This dimension is measured using gradual scales one-to-four whose values are revealed as follows: 1 = Not satisfied at all; 2 = Not very satisfied; 3 = Fairly satisfied; 4 = Very satisfied; and 98 = Not applicable. The factors measured in this item correspond to family, friends, neighbors, colleagues (residence, cultural center, social center, etc.), professional environments (gym, day center, library, and health center), sexual/affective relationships, health, physical, and cognitive abilities, appearance or personal image, financial situation, and leisure and free time. The questionnaire is widened from the scale of satisfaction with different aspects of life provided by Centro de Investigaciones Sociológicas. (2020a), as well as on the Effects and Consequences of COVID-19 scale (Centro de Investigaciones Sociológicas., 2020b,c). It also encompasses items from the Study on Public Opinion and Fiscal Policy held between 2010 and 2017 (Centro de Investigaciones Sociológicas., 2020d) and the Study 3201 parallel carried out by the Spanish General Social Survey (SGSS [in Centro de Investigaciones Sociológicas., 2018]).

D_2: Interpersonal relationships

This dimension is measured using a 1–4 personal contact frequency scale: 1 = Every day or almost every day; 2 = At least one time a week; 3 = At least one time a month; 4 = At least one time a year; and 98 = Not applicable. The personal relationships enquired about referred to a fellow club or association members; the number of offspring not living in the

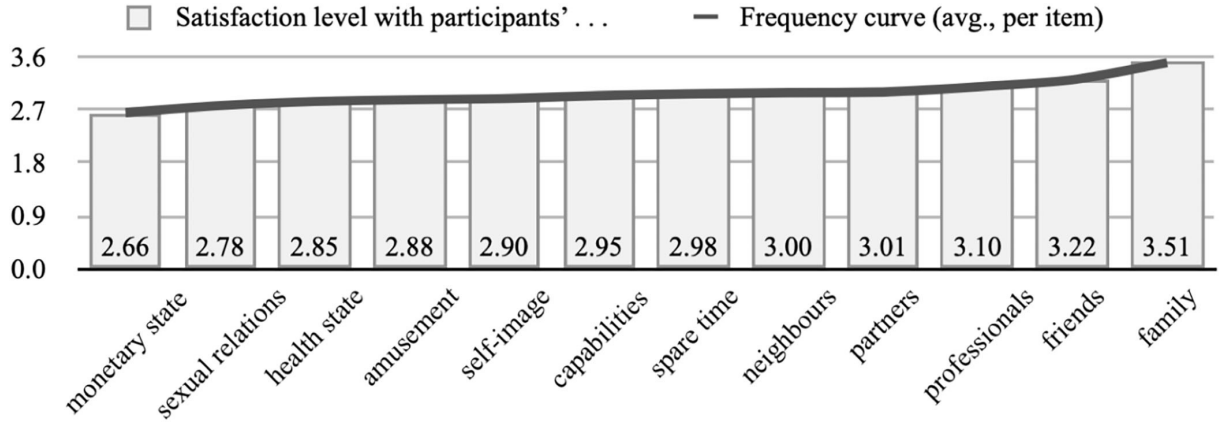


FIGURE 1
Averaged satisfaction levels with the different aspects of life.

TABLE 2 Interpersonal and intrapersonal satisfaction levels.

		Satisfaction levels with											
		I	ii	iii	iv	v	vi	vii	viii	ix	x	xi	xii
Satisfaction levels with family	(1)	1	**0.490****	0.309**	0.243**	0.282**	0.184**	0.250**	0.227**	0.111**	0.146**	0.198**	0.192**
	(2)	–	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(3)	1,062	1,055	1,024	747	862	695	1,059	1,057	1,056	1,037	1,058	1,048
Satisfaction levels with friends	(1)	**0.490****	1	**0.511****	**0.421****	0.351**	0.261**	0.277**	0.282**	0.193**	0.163**	0.236**	0.284**
	(2)	0.000	–	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(3)	1,055	1,056	1,018	745	858	691	1,053	1,051	1,050	1,031	1,052	1,043

(i) Family; (ii) Friends; (iii) Neighbors; (iv) Partners; (v) Professionals; (vi) Sexual/affective relations; (vii) Health; (viii) Skills; (ix) Self-image; (x) Financial situation; (xi) Free time; (xii) Leisure. (1) Pearson correlation, (2) Sig. (Bilateral), and (3) Bold represent the strongest relationships between variables. ** marks that the correlation is significant at the 0.01 level.

same locality; the number of offspring living in the same locality but not with the interviewee; grandchildren, siblings, cousins, and other relatives; friends who are not neighbors; neighbors; others; and caregivers (type of personal contact post-coded from the others). The scale is adapted from the frequency of contact in relationships (Centro de Investigaciones Sociológicas e Instituto de Migraciones y Servicios Sociales, 1998) and the frequency of carrying out activities with different people (Centro de Investigaciones Sociológicas e Instituto de Migraciones y Servicios Sociales, 1998).

D_3: Daily life and activities carried out and their frequency

To explore this dimension, the information is collected through two enquiries: the first on social and leisure activities (D_3.1) and the second on the frequency of activities carried out (D_3.2). The activity frequency scale keeps patterns like the previous ones (1–4 scales) whose values were as follows: 1 = Every day or almost every day; 2 = At least one time a week; 3 = At least one time a month; and 4 = At least one time a year. The scale is adapted from Attendance to Cultural and Leisure Activities and Activities of a Social Nature (Instituto Nacional de Estadística, 2011). The set of activities on which information has been extracted is 23, where some of the examples refer to visiting and receiving visitors (at home [including eating or having coffee together]); eating or dining away from home with family or friends; getting together to play cards or other games; going for tapas, drinks, having an aperitif, having a coffee, beer, or wine at a bar or a coffee shop; going dancing (orchestral dances, discos, lounge); attending a course, workshop, or seminar; attending conferences, gatherings, or discussion forums; going to the movies; going to the theater, ballet, concerts, and magic shows.

Analysis procedure

To present the main descriptive results on each of the analyzed dimensions displayed in the previous epigraph, the choice of the fitting model's central tendency (M) measure is chosen to subsequently proceed to the correlational analysis ($p < 0.01$). This correlational research is performed via Pearson's correlation coefficient (intra- and inter-dimensional). The enquiry about the consistency and the explanatory weight of the analyzed dimensions on the social relations is carried out employing the linear multiple regression analysis.

Results

Research results can be identified in two extensive blocks. The first segment (a) represents the interpretation of each dimension (satisfaction levels, social relationships, and activities of daily life), all linked to the correlations between the different aspects analyzed in each dimension to determine the main internal groupings revealed. The second segment (b), which is the aim of this study, explores the interactional relationship between the dimensions and the power or strength between the variables.

Lifestyles: Satisfaction levels, social relationships, activities, and daily life

The satisfaction levels with the different aspects of life (D_1) are formed as self-perception indicators about people's lifestyles and vital situations. Thus, when averaging data, the results revealed that (inter)personal relationships are the aspects providing the analyzed population higher satisfaction levels (family [$\bar{x} = 3.51$], friends [$\bar{x} = 3.22$], professionals [$\bar{x} = 3.10$], colleagues [$\bar{x} = 3.01$], and neighbors [$\bar{x} = 3.00$]). Contrarily, among the surveyed, the least valued aspects of life were those associated with the monetary state, sexual or affective relations, and health state (Figure 1).

Personal relationships are constituted as the aspects that generate the most remarkable satisfaction levels; satisfaction with family and friends also maintains a robust correlation with

TABLE 3 Daily activities from the age of 65-year-old and above.

Visit and receive visits	55.8%
Eating or dining out home	58.4%
Going for a walk (accompanied)	73.7%*
Talking on the phone with family and friends	74.6%*

*Every day or almost every day.

the rest of the factors analyzed ($p < 0.01$). Consequently, the high satisfaction levels with family and friends coincide with the different aspects analyzed (Table 2).

Having shown the high satisfaction levels prompted by the social interactions, the frequency of contact between the different groups is exposed (D_2). Regarding the frequency of social relationships, family relationships are the most assiduous (children in the same locality but not living together = 55.2%; grandchildren = 37.9%; children outside the locality = 28.6%). In the case of non-family social relationships, the frequency of contact is 46.7% in friends and 44.8% in neighbors. All the variables that affect personal contacts correlate significantly with each other.

Two differentiated relationships emerge from the analysis when observing the strongest correlations between variables. On the one hand, there are different family relationships with each other ("offspring living in the same locality" correlates with the "frequency of contact with offspring living outside city boundaries" [$r = 0.643$] and with "grandchildren" [$r = 0.622$]). On the other hand, there are established relationships between non-family contacts on both the "frequency of contact with friends" and "frequency of contact with neighbors (AmE)" [$r = 0.448$]).

The variables that affect the relationship with different personal contacts correlate significantly between them so that those who maintain frequent contact with one of the groups of people also maintain it with others and vice versa. However, if we stop at the most robust correlations, two distinct relationships reappear. Family relationships maintain a strong correlation among them both ("offspring living in the same locality" correlates with the frequency of contact with "offspring living outside city boundaries" [$r = 0.643$] and with "grandchildren" [$r = 0.622$]). Meanwhile, non-family relationships and friends and neighbors present the highest correlations between themselves ($r = 0.448$).

The third dimension of analysis (D_3) refers to the performance and frequency of the sets of daily activities (Table 3). If we focus on carrying out activities in the daily life of people over 65 in Castile-La Mancha, the most common activities are "talking on the phone with family and friends," "walking" (accompanied), "eating or dining out," and "visiting and entertainment." However, not all activities are carried out with the same frequency. Among them all, a higher percentage of activities carried out every, or almost every, day is "talking on the phone with family and friends" (82%) and "going for a walk" (77.7%).

Again, the correlation analysis between the performance frequencies of the different activities reveals two differentiated groups (Table 4). On the one hand, strong correlations were observed between the variables that show a high level of activity and diversity in leisure time outside the home and are the most linked to consumerism ($p < 0.01$). Contrarily, activities of a more traditional nature were linked to lower activity levels and social interactions.

Two large groups are observed from analyzing the main correlations within each dimension. In the dimension "satisfaction with different aspects of life" (D_1), we found a group of significant correlations ($p < 0.01$). Such correlations between family relationships and satisfaction levels in non-family relationships (friends and neighbors) concur. In the case of the dimension that measures the frequency in social relationships (D_2), we also found two groups of significant correlations ($p < 0.01$). On the one hand, the different family relations (children and grandchildren) are observed; on the other hand, non-family relationships (friends and neighbors) are revealed. Such correlations would indicate that, in cases in which there is a high frequency of contact with children, it also occurs in other family relationships. Similarly, when there is a high frequency of contact with friends, there is also a contact frequency in other relationships that are not family. Such results reveal a certain continuity with the results of D_1.

Finally, the correlations between the frequency of performing the group of activities of daily life (D_3.2) have also been analyzed, from which two different groups emerge between (I) the activities most linked to leisure and outside home activities and consumerism; and (II) the more traditional activities implying higher levels of inactivity or sedentariness. After the pandemic, such dimensional dualities might have justified the interactional analysis between the three proposed dimensions on the relationships between elderly groups and their lifestyles.

Interaction between dimensions

The following subsections show (i) the interaction between satisfaction levels and contact frequencies and (ii) the frequency of activities and social relationships to show how the groupings presented above provide continuity to the relationship between the dimensions.

Satisfaction levels and frequency of contacts

When connecting respondents' satisfaction levels (D_1) with the frequency of contact in their social relationships (D_2), their frequency of contact with people increases and so does

TABLE 4 Correlation on activity performance.

		Frequency levels with												
		I	ii	iii	iv	v	vi	vii	viii	ix	x	xi	xii	xiii
Frequency level on visiting and receiving visits	(1)	–	0.241**	0.381**	–	–	0.238**	–	–	**0.427****	**0.448***	–	–	–
	(2)	–	0.000	0.000	–	–	0.000	–	–	0.000	0.011	–	–	–
	(3)	–	423	286	–	–	490	–	–	189	31	–	–	–
Frequency level on eating or dining out at home	(1)	0.241**	–	**0.461****	**0.569****	0.187**	0.244**	0.308**	0.353**	–	–	0.251**	0.433**	0.268**
	(2)	0.000	–	0.000	0.000	0.000	0.000	0.002	0.001	–	–	0.000	0.004	0.000
	(3)	423	–	371	50	518	516	100	79	–	–	323	43	339

(i) Visiting and receiving visits; (ii) Have lunch or dinner off home; (iii) Go for tapas, drinks, and aperitifs; (iv) Go dancing; (v) Go for a walk (accompanied); (vi) Talk on the phone with family and friends; (vii) Go to the movies; (viii) Go to the theater, ballet, and concerts; (ix) Go to church or religious activities; (x) Go to folkloric, bullfighting, and sports shows; (xi) Go to shopping centers; (xii) Go to the library or rearing club; (xiii) Read books, newspapers, and magazines. (1) Pearson correlation, (2) Sig. (Bilateral), and (3) Bold represent the strongest relationships between variables. * marks that the correlation is significant at the 0.05 level and ** marks that the correlation is significant at the 0.01 level.

TABLE 5 Correlations between frequency of socio-family contact and satisfaction levels.

		Frequency of contact with		Satisfaction levels with					
		i	ii	iii	iv	v	vi	vii	viii
Frequency of contact with offspring living in the same city	(1)	1	**0.622****	−0.335**	−0.284**	−0.166**	−0.171**	−0.217**	0.097**
	(2)	–	0.000	0.000	0.000	0.000	0.000	0.000	0.004
	(3)	924	838	924	919	894	660	758	900
Frequency of contact with grandchildren	(1)	**0.622****	1	−0.259**	−0.229**	−0.204**	−0.202**	−0.166**	0.117**
	(2)	0.000	–	0.000	0.000	0.000	0.000	0.000	0.001
	(3)	838	881	881	875	855	619	718	857

(i) Offspring living in the same city; (ii) Grandchildren; (iii) Family; (iv) Friends; (v) Neighbors; (vi) Partners; (vii) Professionals; (viii) economic status. (1) Pearson correlation. (2) Sig. (Bilateral). (3) Bold represent the strongest relationships between variables. ** marks that the correlation is significant at the 0.01 level.

their satisfaction levels with the different aspects evaluated ($p < 0.01$). Also, it was observed that the contact frequency with relatives (offspring/grandchildren living in the same locality) establishes significant correlations with the rest of satisfaction levels: family, friends, neighbors, colleagues, professionals, etc. (Table 5). This correlation is inverse when it concerns the "level of economic satisfaction." That is, in all cases, the greater the relationship with "offspring or grandchildren living in the same locality," the lower the satisfaction level with their "financial situation."

When analyzing and comparing the set of significant correlations ($p < 0.01$) with the "frequency of contact with friends," the most robust correlations emerge from the levels of satisfaction with non-family personal relationships: friends, neighbors, colleagues, and professionals (Table 6).

Frequencies in carrying out activities and frequency of contact

Delving into the possible relationships that could be established between the frequencies of activities (D_3.2) and the frequency of contact with different groups of people (D_2), we found significant relationships between the different groups of variables (detecting two large correlation groups). On the one hand, the frequency of contact with non-family relationships (friends and neighbors) maintains a significant correlation ($p < 0.01$) with most of the leisure activities and activities leading to consumerism (Table 7).

Within the second correlational group (Table 8), it can be observed that the frequency of contact with family relationships (children and grandchildren) also correlates significantly ($p < 0.01$) with the more traditional activities carried out more individually (talking on the phone with family and

TABLE 6 Satisfaction levels from non-family interrelations.

		Satisfaction levels with					
		i	ii	iii	iv	v	vi
Satisfaction levels with friends	(1)	1	−0.065*	−0.207**	−0.145**	−0.150**	−0.179**
	(2)	–	0.037	0.000	0.000	0.000	0.000
	(3)	1,026	1,025	1,022	988	734	835

(i) Frequency of contact with friends who are not neighbors; (ii) Family; (iii) Friends; (iv) Neighbors; (v) Partners; (vi) Professionals. (1) Pearson correlation. (2) Sig. (Bilateral). (3) * marks that the correlation is significant at the 0.05 level and ** marks that the correlation is significant at the 0.01 level.

TABLE 7 Correlational significance between the groups of variables (I).

		Frequency in					
		i	ii	iii	iv	v	vi
Frequency of contact with friends who are not neighbors	(1)	0.177**	0.133**	−0.235**	−0.387**	−0.196**	−0.187**
	(2)	0.000	0.001	0.000	0.002	0.000	0.000
	(3)	582	600	460	64	765	766
Frequency of contact with neighbors	(1)	0.172**	0.128**	−0.276**	0.251	−0.119**	−0.183**
	(2)	0.000	0.002	0.000	0.058	0.001	0.000
	(3)	549	563	423	58	723	716

(i) Visiting and receiving visits; (ii) Have lunch or dinner off home; (iii) Go for tapas, drinks, or have an aperitif; (iv) Go dancing; (v) Go for a walk (accompanied); (vi) Talk on the phone with family and friends. (1) Pearson correlation. (2) Sig. (Bilateral). (3) ** marks that the correlation is significant at the 0.01 level.

friends, walking, computer communications [chat, email, video conferencing], and going to the library or reading club).

In sum, it can be said that revealed correlations point to two differential relationships and lifestyles that strengthen the frequency of different activities and social relationships in the elder. Digging deeper into such an observed lifestyle's dichotomy, variance remains explained at 23.6% in the linear regression analysis, introducing as a *dependent* variable the frequency in relationships with friends and as an *independent* variable the main variables displaying significant correlations. Consequently, the result yields an adjusted R^2 of 0.236 [23.6% of the variance explained (Table 9)].

To what happens in the Fitted model (I), if we set the frequency of contact with offspring living outside locality boundaries, the *dependent* variable, and the group of variables with the highest correlation as the *independent* variable, the linear regression analysis provides a result of the model to an adjusted R^2 of 0.451 [45.1% of the variance explained on the frequency of contact with offspring (Table 10)].

Discussion

The original research goal and the three scrutinized dimensions revealed through the analysis that personal relationships still constitute themselves as the aspects of life displaying the highest satisfaction levels to those aged 65 and above residing in Castile-La Mancha, with family interaction being the most frequent of them all (every/almost every day). Such an aspect is initially worth to be mentioned because, despite the respiratory issues provoked by the SARS-CoV-2 retrovirus, the 40,000,000 diagnosed cases and the exceeding 1,100,000 deaths by 2020, and the social distancing and confinement or shelter-in-place measures (Palialol et al., 2020), the social relations prevail over other supposed essential aspects like health, self-skills, or wealth to the elders. So much so is this phenomenon that the high level of relational satisfaction with family and friends shows, at all times, a high correlation with satisfaction levels with the rest of the daily aspects of life. In this regard, there is coincident research pointing out the importance of social relationships in satisfaction with life and wellbeing (Cornwell et al., 2008; Litwin and Shiovitz-Ezra, 2011; Molero and Pérez-Fuentes, 2011; Pérez-Escoda, 2013; Rafnsson et al., 2015; Tomini et al., 2016; Burholt et al., 2020; LaBorde and Williams, 2022). However, in other studies, the time spent in social relationships reveals no correlation with life satisfaction (Triadó et al., 2009).

Despite the frequency of contact, the relationships with family scores as the most frequent, and to a lesser extent the socialization with friends or neighbors, those polled showing a high frequency of contact with any of these groups (family, friends, and neighbors) also revealed a high frequency with the other variables from the set. In this sense, although the adopted social distancing policies and measures were (and indeed still are) beneficial to stopping the spread out of the SARS-CoV-2 infection and saving lives (Thunström et al., 2020), we observed the use of alternative ways of interaction, i.e., in the study, a

TABLE 8 Correlational significance between the groups of variables (II).

		Frequency to			
		i	ii	iii	iv
Frequency of contact with friends who are not neighbors	(1)	0.136**	0.197**	−0.210**	−0.521**
	(2)	0.002	0.000	0.001	0.000
	(3)	535	530	242	50
Frequency of contact with neighbors	(1)	–	0.090*	−0.204**	−0.340**
	(2)	–	0.020	0.001	0.007
	(3)	–	673	271	61

(i) Go for a walk; (ii) Talk on the phone with family and friends; (iii) Manage computer communications; (iv) Go to the library or reading club. (1) Pearson correlation. (2) Sig. (Bilateral). (3) * marks that the correlation is significant at the 0.05 level and ** marks that the correlation is significant at the 0.01 level.

strong relationship between the variables "talking on the phone with relatives and friends" and "computer communications" (chat, email, video conferencing) with the variable "frequency of family relationships" was observed. From this particular piece of information, we can infer that, regardless of the social situation of each moment, there are interactional or socializing factors remaining constant or unaltered (even reinforced).

The above-referenced observation leads to deepening into the interactional differences revealed by participants. That is to say, those polled displaying higher punctuations in the frequency of contact with the "non-relatives" item conveyed an activity more related to *consumerism* and *leisureliness*. This result of leisure activities is coincident with Toepoel's (2013) research, which revealed greater social integration by the people connected to this variable. The post-pandemic scenario would represent a problem for this population because these activities tend to occur outside the home and in indoor locations (shopping centers, bars, restaurants, etc). Concerning this tendency, various studies proved the influence of both social relationships and activity and leisure on satisfaction with life in general (Cheng et al., 2009; Molero and Pérez-Fuentes, 2011; Li et al., 2021; Chen et al., 2022; Geithner and Wagner, 2022), as well as the connection between the performance of leisure activities and an increase in social connectivity (Toepoel, 2013), but the present research has just focused the attention on the interrelation between participant's relationships and the activities carried out. Outdoor activities do also imply significant cultural differences (Silverstein and Parker, 2002; Lee et al., 2014). Diametrically opposed, in our research, it was found that those polled manifesting higher contact frequency with "relatives" or "family" members correlated with the more *traditional* activities.

The differentiation of activities shown in the previous paragraph leads to the different lifestyles observed in participants. Previously, it was mentioned that, regarding the main activities of daily life, there was a group with a high correlation between the more traditional activities and a lower level of activity and another group with a correlation between activities more linked to consumption outside the home and more significant social activity. These dualities were kept when the three dimensions were analyzed two-by-two. For example, regarding the relationship between satisfaction levels and the frequency of contact with relationships, it was observed that personal contact increased satisfaction with life. Similarly, it was detected that the satisfaction level with the family was closely connected to the frequency of contact with family relationships (mainly children and grandchildren), while the satisfaction level with friends arises, to a greater extent, from the frequency of contact with friends and neighbors.

Regarding the connection between social relationships and satisfaction levels, other studies coincided with the present research (Fagerström et al., 2007; Triadó et al., 2009; Toepoel, 2013; Rafnsson et al., 2015; Steptoe and Fancourt, 2019). Two correlation groups showed two differentiated lifestyles concerning the relationship between frequency of contact and the activities carried out. That is, the frequency of contact with non-family relationships (friends and neighbors) maintained a strong influence on a large part of the leisure and consumption

TABLE 9 Fitted model (I).

Model	R	R^2	R^2_{adj}	Std. error of the estimate
1	0.630[a]	0.397	0.236	0.71046

[a]Explanation of the regression model; Frequency of having lunch or dinner away from home; Frequency of visiting and receiving visitors; Frequency of talking on the phone with family and friends; Frequency of going dancing.

TABLE 10 Fitted model (II).

Model	R	R^2	R^2_{adj}	Std. error of the estimate
1	0.710[a]	0.504	0.451	0.70446

[a]Explanation of the regression model; Frequency of computer communications (chat, email, videoconferences); Frequency of going to the library, reading club; Frequency of talking on the phone with family and friends.

activities, while, in the second correlational group, it was observed that the frequency of contact with family relationships manifested a more substantial influence on the more traditional and solo activities.

In addition to the relationships found, other factors regarding this line of research could be investigated, which could be considered limitations in this study field, such as nutrition and physical activity. The relationship between physical activity and active and healthy aging is yet agreed upon. However, scientific evidence warns of the risks of physical activity that require significant effort. What has been pointed out is essential because 7.5% of the elderly in Castile-La Mancha who would carry out tasks that require great physical effort would be affected. Also, Vancini et al. (2021) showed that high-volume physical exercise with moments of maximum intensity, in the long term, is associated with an increased risk of myocardial infarction and other serious health problems. Regarding nutrition, although it is not part of this research, it could be interesting to incorporate the relationship between ingested food and fine motor and cognitive function. Concerning future and necessary research, Akbari et al. (2021) dug into an exciting topic in the field as the specific intake of certain micronutrients in a healthy diet and proved its correlation with *cognitive* performance and the level of physical activity in older adults.

Conclusion

In sum, this study's contribution to science confirms the family's value for those aged 65 years and above and how they, far from substantially altering their life habits in the face of adversity, tend toward new forms of socialization and maintenance of leisure activities. In other words, social relations in the post-pandemic scenario remain as before the appearance of the virus and reappear with more force. Such relationships are constituted as the motor of both the satisfaction level with life and the diversity and frequency of daily life activities.

Although the previous aspect is general and applies to the entire population analyzed, there are two main lifestyles linked to the frequency and type of social relationships that occurred: one of a more traditional nature and linked to a high frequency in family relationships causing activities more traditional and with less social interaction and another with greater relevance and frequency in the relationships with non-family causing leisure activities, consumerism, and more significant social and relational activity.

In our view, such lifestyles should not be interpreted as exclusive but as intrinsic and underlying aspects of the elders' diversity, beliefs, and values. Regarding future research and shortcomings detected in the study, a larger-scale investigation incorporating participants from other cultures would have added additional value. In this sense, more studies would be necessary to verify the results in other countries and communities or cultures.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors. Requests to access the datasets have to be directed to the corresponding author.

Author contributions

OD and PF: research design and data collection. IH: statistical analysis. JG: writing. All authors contributed to the article and approved the submitted version.

Funding

This article is a by-product of a broader study entitled "Nuevos perfiles de la población mayor en Castilla-La Mancha", a result of the collaboration between the European Network for the Fight against Poverty in Castilla-La Mancha (EAPN) and the University of Castilla-La Mancha (UCLM) from the core project SBPLY/19/270802/000377 named "Impulso y desarrollo de acciones de voluntariado en el marco de Envejecimiento Activo de los Centros de Mayores" and financed by the Regional Ministry of Social Welfare in Castilla-La Mancha, Spain.

Conflict of interest

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.

Publisher's note

References

Akbari, A., Mirakhori, F., Ashouri, M., and Tehrani, S. N. (2021). The effect of micronutrient intake on cognitive function and physical activity of the elderly. *Int. J. Sport Stud. Health*. 4, 1. doi: 10.5812/intjssh.121360

Aponte, V. C. (2015). Calidad de vida en la tercera edad. *Ajayu* 13, 152–182. https://ajayu.ucb.edu.bo/a/article/view/78

Bulow, M. H., and Söderqvist, T. (2014). Successful ageing: a historical overview and critical analysis of a successful concept. *J. Aging Stud.* 31, 139–149. doi: 10.1016/j.jaging.2014.08.009

Burholt, V., Winter, B., Aartsen, M., Constantinou, C., Dahlberg, L., Feliciano, V., et al. (2020). A critical review and development of a conceptual model of exclusion from social relations for older people. *Eur. J. Ageing* 17, 3–19. doi: 10.1007/s10433-019-00506-0

Centro de Investigaciones Sociológicas e Instituto de Migraciones y Servicios Sociales (1998). *Estudio La soledad en las personas mayores, 2279*. Available online at: https://cutt.ly/fUp9kD8 (accessed February 22, 2022*)*.

Centro de Investigaciones Sociológicas. (2015). *Barómetro de septiembre de 2015, 3109*. Available online at: https://cutt.ly/2Up9YpC (accessed December 9, 2021).

Centro de Investigaciones Sociológicas. (2018). *Encuesta Social General Española (ESGE) 2017, 3201*. Available online at: https://cutt.ly/GUp9FrO (accessed December 9 2021).

Centro de Investigaciones Sociológicas. (2020a). *Estudio Bienestar Emocional (Piloto Cati), 3285*. Available online at: https://cutt.ly/9Up94Ix (accessed December 9, 2021).

Centro de Investigaciones Sociológicas. (2020b). *Estudio Efectos y Consecuencias del Coronavirus (I), 3298*. Available online at: https://cutt.ly/RUp92QH (accessed December 9, 2021).

Centro de Investigaciones Sociológicas. (2020c). *Estudio Efectos y Consecuencias del Coronavirus (II), 3302*. Available online at: https://cutt.ly/FUp9BQC (accessed December 9, 2021).

Centro de Investigaciones Sociológicas. (2020d). *Estudio Opinión pública y política fiscal (XXXVII), 3290*. Available online at: https://cutt.ly/zUp9LA0 (accessed December 9, 2021).

Chen, S., Malete, L., and Ling, J. (2022). An examination of physical activity guidelines and health-related quality of life among U.S. older adults. *Prev. Med.* 156, 106986. doi: 10.1016/j.ypmed.2022.106986

Cheng, S. T., Lee, C. K. L., Chan, A. C. M., E. M. F., and Leung, E. M. F. and Lee, J. J. (2009). Social Network Types and Subjective Well-being in Chinese Older Adults. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 64B, 713–722. doi: 10.1093/geronb/gbp075

Ciobanu, R. O., Fokkema, T., and Nedelcu, M. (2017). Ageing as a migrant: vulnerabilities, agency and policy implications. *J. Ethn. Migr. Stud.* 43, 164–181. doi: 10.1080/1369183X.2016.1238903

Cornwell, B., Laumann, E. O., and Schumm, L. P. (2008). The social connectedness of older adults: a national profile. *Am. Sociol. Rev.* 73, 185–203. doi: 10.1177/000312240807300201

de Albuquerque Araújo, L., Álvarez, A. J., and Palomo, I., Bustamante, M. A. (2020). Determinantes de la satisfacción con la alimentación en adultos mayores chilenos. *Nutr. Hosp.* 36, 805–812. doi: 10.20960/nh.02481

Etxeberría, I., Etxebarría, I., and Urdaneta, E. (2019). Subjective well-being among the oldest old: The role of personality traits. *Pers. Individ. Dif.* 146, 209–216. doi: 10.1016/j.paid.2018.04.042

Fagerström, C., Borg, C., Balducci, C., Burholt, V., Wenger, C. G., Ferring, D., et al. (2007). Life Satisfaction and Associated Factors Among People Aged 60 Years and Above in Six European Countries. *Appl. Res. Qual. Life* 2, 33–50. doi: 10.1007/s11482-007-9028-0

Geithner, L., and Wagner, M. (2022). Old-age lifestyles: Patterns of participation in leisure activities and their associations with different forms of capital. *J. Aging Stud.* 61, 101022. doi: 10.1016/j.jaging.2022.101022

Gumà, J., and Arpino, B. (2021). Satisfacción con la vida según la contribución a la esfera pública y privada en las parejas españolas adultas. *Rev. Int. Sociol.* 79, e177. doi: 10.3989/ris.2021.79.1.19.045

Gutiérrez, M., Galiana, L., Tomás, J. M., Sancho, P., and Sanchís, E. (2014). La predicción de la satisfacción con la vida en personas mayores de Angola: El efecto moderador del género. *Psychosoc. Interv.* 23, 17–23. doi: 10.5093/in2014a2

Han, J., Lee, S., and Kwon, Y. (2022). Can social capital improve the quality of life satisfaction for older adults? Focusing on the 2016 Quality of Life Survey in Gyeonggi Province, Korea. *Cities* 130, 103853. doi: 10.1016/j.cities.2022.103853

Herranz, I., Lirio, J., Portal, E., and Arias, E. (2013). La actividad física como elemento de participación y calidad de vida en las personas mayores. *Escr. Psicol.* 6, 13–19. doi: 10.5231/psy.writ.2013.1906

Instituto Nacional de Estadística (2011). *Encuesta de uso del tiempo 2009-2010*. Available online at: https://cutt.ly/0Up2ACI (accessed January 15, 2022).

Instituto Nacional de Estadística (2020). *Proyecciones de población 2020-2070*. Available online at: https://cutt.ly/qREcMQE (accessed December 9, 2021).

Ju, H. (2017). The relationship between physical activity, meaning in life, and subjective vitality in community-dwelling older adults. *Arch. Gerontol. Geriatr.* 73, 120–124. doi: 10.1016/j.archger.2017.08.001

LaBorde, P. J., and Williams, V. (2022). The surprising effects of social isolation and loneliness on physical health in older adults. *Adv. Fam. Pract. Nurs.* 4, 13–25. doi: 10.1016/j.yfpn.2021.12.001

Lee, J. H., Lee, J. H., and Park, S. H. (2014). Leisure activity participation as predictor of quality of life in Korean urban-dwelling elderly. *Occup. Ther. Int.* 21, 124–132. doi: 10.1002/oti.1371

Li, M., Sun, H., Xu, W., Ma, W., Yuan, X., Wu, H., et al. (2021). Leisure activity and cognitive function among Chinese old adults: The multiple mediation effect of anxiety and loneliness. *J. Affect. Disord.* 294, 137–142. doi: 10.1016/j.jad.2021.07.051

Litwin, H., and Shiovitz-Ezra, S. (2011). Social network type and subjective well-being in a national sample of older Americans. *Gerontologist* 51, 379–388. doi: 10.1093/geront/gnq094

Marques, L. P., Schneider, I. J. C., and d'Orsi, E. (2016). Quality of life and its association with work, the Internet, participation in groups and physical activity among the elderly from the EpiFloripa survey, Florianópolis, Santa Catarina State, Brazil. *Cad. Saúde Pública* 32, e00143615. doi: 10.1590/0102-311x00143615

Molero, M. M., and Pérez-Fuentes, M. C. (2011). Salud y calidad de vida en adultos mayores institucionalizados. *Int. J. Dev. Educ. Psychol.* 1, 249–258. https://dialnet.unirioja.es/ejemplar/433268

Palialol, B., Pereda, P., Azzoni, C. (2020). Does weather influence COVID-19 transmission? *Regional Sci. Policy Prac.* 12:981–1004. doi: 10.1111/rsp3.12367

Pant, S., and Subedi, M. (2020). Impact of COVID-19 on the Elderly. *J. Patan Acad. Health Sci.* 7, 32–38. doi: 10.3126/jpahs.v7i2.31104

Penton, H., Dayson, C., Hulme, C., and Young, T. (2022). A qualitative investigation of older adults' conceptualization of quality of life and a think-aloud content validation of the EQ-5D-5L, SF-12v2, Warwick Edinburgh Mental Wellbeing Scale, and Office of National Statistics-4. *Value Health*. doi: 10.1016/j.jval.2022.04.1735

Pérez-Escoda, N. (2013). *Variables predictivas de la satisfacción con la vida en estudiantes universitarios. XVI Congreso Nacional/II Internacional Modelos de Investigación Educativa de AIDIPE. Alicante, Spain*. Available online at: https://cutt.ly/XREvcwp (accessed December 9, 2021).

Rafnsson, S. B., Shankar, A., and Steptoe, A. (2015). Longitudinal influences of social network characteristics on subjective well-being of older adults: findings from the ELSA study. *J. Aging Health* 27, 919–934. doi: 10.1177/0898264315572111

Rowe, J. W., and Kahn, R. L. (1997). Successful Aging. *Gerontologist* 37, 433–440. doi: 10.1093/geront/37.4.433

Sampaio, P. Y., and Ito, E. (2013). Activities with higher influence on quality of life in older adults in Japan. *Occup. Ther. Int.* 20, 1–10. doi: 10.1002/oti.1333

Santaella, A., and Bohórquez, M. R. (2017). *¿Afecta la realización de actividad física en la percepción de la satisfacción con la vida en personas mayores de 65 años? XV Congreso Andaluz y II Luso-Andaluz de Psicología de la Actividad Física y el Deporte. Granada, Spain*. Available online at: http://dx.doi.org/10.13140/RG.2.2.31186.35522 (accessed December 9, 2021).

Schnettler, B., Miranda, H., Sepúlveda, J., Orellana, L., Denegri, M., Mora, M., et al. (2014). Variables que influyen en la satisfacción con la vida de personas de distinto nivel socioeconómico en el sur de Chile. *Suma Psicol.* 21, 54–62. doi: 10.1016/S0121-4381(14)70007-4

Silverstein, M., and Parker, M. G. (2002). Leisure activities and quality of life among the oldest old in Sweden. *Res. Aging* 24, 528–547. doi: 10.1177/0164027502245003

Steptoe, A., and Fancourt, D. (2019). Leading a meaningful life at older ages and its relationship with social engagement, prosperity, health, biology, and time use. *Proc. Nat. Acad. Sci.* 116, 1207–1212. doi: 10.1073/pnas.1814723116

Thunström, L., Stephen, C., Newbold, D. F., Madison, A., and Shogren, J. F. (2020). The benefits and costs of using social distancing to flatten the curve for COVID-19. *J. Benefit Cost Anal.* 11:179–195. doi: 10.1017/bca.2020.12

Toepoel, V. (2013). Ageing, leisure, and social connectedness: how could leisure help reduce social isolation of older people? *Soc. Indic. Res.* 113, 355–372. doi: 10.1007/s11205-012-0097-6

Tomini, F., Tomini, S. M., and Groot, W. (2016). Understanding the value of social networks in life satisfaction of elderly people: a comparative study of 16 European countries using SHARE data. *BMC Geriatr.* 16, 203. doi: 10.1186/s12877-016-0362-7

Tomioka, K., Kurumatani, N., and Hosoi, H. (2017). Association between social participation and 3-year change in instrumental activities of daily living in community-dwelling elderly adults. *J. Am. Geriatr. Soc.* 65, 107–113. doi: 10.1111/jgs.14447

Triadó, C., Villar, F., Solé, C., Celdrán, M., and Osuna, M. J. (2009). Daily activity and life satisfaction in older people living in rural contexts. *Span. J. Psychol.* 12, 236–245. doi: 10.1017/S1138741600001645

United Nations. (2019). *World Population Prospects 2019*. Available online at: https://cutt.ly/WUpMhbV (accessed January 10, 2022*)*.

Urtamo, A., Jyväkorpi, S. K., and Strandberg, T. E. (2019). Definitions of successful ageing: a brief review of a multidimensional concept. *Acta Biomed.* 90, 359–363. doi: 10.23750/abm.v90i2.8376

Vancini, R. L., dos Santos Andrade, M., de Lira, C. A., Nikolaidis, P. T., and Knechtle, B. (2021). Is It Possible to Age Healthy Performing Ultra-endurance Exercises? *Int. J. Sport Stud. Health.* 4, 1. doi: 10.5812/intjssh.122900

VanderWeele, T. J. (2017). On the promotion of human flourishing. *Proc. Nat. Acad. Sci.* 114, 8148–8156. doi: 10.1073/pnas.1702996114

Von Humboldt, S., and Leal, I. (2017). Predictors of satisfaction with life among older adults. *Eur. Psychiatry* 41, S177–S178. doi: 10.1016/j.eurpsy.2017.01.2080

Vozikaki, M., Linardakis, M., Micheli, K., and Philalithis, A. (2017). Activity participation and well-being among european adults aged 65 years and older. *Soc. Indic. Res.* 131, 769–795. doi: 10.1007/s11205-016-1256-y

Xue, Y., Lu, J., Zheng, X., Zhang, J., Lin, H., Qin, Z., et al. (2021). The relationship between socioeconomic status and depression among the older adults: the mediating role of health promoting lifestyle. *J. Affect. Disord.* 285, 22–28. doi: 10.1016/j.jad.2021.01.085

Zadworna, M. (2022). Pathways to healthy aging – exploring the determinants of self-rated health in older adults. *Acta Psychol.* 228, 103651. doi: 10.1016/j.actpsy.2022.103651

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 13 October 2022
DOI 10.3389/fpsyg.2022.998658

OPEN ACCESS

EDITED BY
Fco. Pablo Holgado-Tello,
National University of Distance Education (UNED), Spain

REVIEWED BY
Miguel Pic,
South Ural State University, Russia
José Ángel Martínez-Huertas,
Universidad Nacional de Educación a Distancia, Spain

*CORRESPONDENCE
Eulàlia Arias-Pujol
eulaliaap@blanquerna.url.edu

SPECIALTY SECTION
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

RECEIVED 20 July 2022
ACCEPTED 30 August 2022
PUBLISHED 13 October 2022

CITATION
Arias-Pujol E, Mestres M, Miralbell J, Bachs N and Anguera MT (2022) Implementation and evaluation in low intensity intervention programs from the CONNECT perspective of mixed methods: Application in a case of an autistic child.
Front. Psychol. 13:998658.
doi: 10.3389/fpsyg.2022.998658

Implementation and evaluation in low intensity intervention programs from the CONNECT perspective of mixed methods: Application in a case of an autistic child

Eulàlia Arias-Pujol[1]*, Marina Mestres[2], Júlia Miralbell[2,3], Natalia Bachs[1] and M. Teresa Anguera[3]

[1]Faculty of Psychology, Education and Sports Sciences Blanquerna, Ramon Llull University, Barcelona, Spain, [2]Carrilet Research and Education Center, Barcelona, Spain, [3]Faculty of Psychology, Institute of Neurosciences, University of Barcelona, Barcelona, Spain

There has been a comprehensive development over the last few years of low intensity intervention programs that are implemented within a user context and that are made up of everyday life activities, and it has been necessary to adapt the necessary methodological channels in order to guarantee an adequate resolution pathway. The mixed method perspective offers a suitable framework, and observational methodology – in itself considered mixed method – is appropriate for studying the implementation and evaluation of low intensity intervention programs, allowing the development of the QUAL-QUAN-QUAL stages that correspond to the connect integration pathway of mixed methods. In this work it was applied to a single case, in a low intensity intervention, retrieving valuable information obtained, but systematizing it and applying quantitizing to the qualitative data that was treated quantitatively in a rigorous manner. The aim was to analyze the psychotherapist-patient interaction in psychoanalytic psychotherapy, in which we sought to identify which of the therapist's techniques stimulated actions of reciprocal social interaction in the child, and which techniques inhibited non reciprocal social interactions. The observational design was nomothetic, follow-up, and multidimensional. The patient was a 4-year-old boy with a diagnosis of severe autism spectrum disorder. 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. Polar coordinate analysis was carried out using the free program HOISAN 2.0. Polar coordinate analysis allows us to obtain an inter-relational map of the connections detected between focal behavior established in each case and the different categories. The results provide objective evidence – backed up by the application of polar-coordinate-based data analysis – that within a framework of psychoanalytic psychotherapy, the techniques

of "verbalization" and "vocalization" significantly activate reciprocal social interaction behaviors and inhibit non-social reciprocal behaviors in a child with severe autism spectrum disorder with no language. On the other hand, direct gaze promotes the child's withdrawal. The results are of key importance as they show the therapist behaviors most useful for promoting social interaction in a child with severe autism.

KEYWORDS

mixed methods, case study, polar coordinate analysis, CONNECT, autism, psychoanalytic psychotherapy

Introduction

It is an undoubtedly complex task to make decisions about the implementation and evaluation of intervention programs which, in any case, should be conditional on the applied methodology. The structural dimensions of the evaluation of a program are established in scientific literature (Chacón et al., 2000; Chacón-Moscoso et al., 2002, 2013, 2014, 2021), and the existing correspondence between these dimensions and methodological quality in the chosen procedural option is extensively relevant.

The broad definition of the aim of intervention program evaluation leads to judgments concerning the value of such programs or of some of their elements (Anguera and Chacón-Moscoso, 2008; Chacón-Moscoso et al., 2013). It has been observed that the intensity of such intervention can be extremely varied depending on the situation, taking into account that it will have considerable bearing on the procedure to be followed. It is precisely for this reason that minimum basic and common aspects have been specified that must be stated in the evaluation report of any program, regardless of its approach (Chacón-Moscoso et al., 2013), and shape a formative and summative evaluation in continual interaction, throughout its phases.

In the last few years, it has been established that the greatest shortcomings in the implementation of intervention programs are procedural, and that some of these have barely been addressed. This is mainly due to the complex nature of the reality, with many superimposed levels, and the fact that individuals or collectives who experience the actions of an intervention program may be heterogeneous. Furthermore, the dynamic of the processes is not uniform, making it difficult to collate data in a way that fulfills the requirements demanded by rigor.

In this sense, we are aware of the need to pay special attention to the implementation and evaluation of bespoke intervention programs for each specific case, aiming for an increased effectiveness in the actions carried out, without disturbing the daily life of the user, as an essential leitmotif. This special attention implies that the intervention itself blends into daily life, searching for an alternative to traditional conventionality. In this study, we focus on a low *level of intervention intensity*, an expression coined by Anguera (2008, p. 154), in the face of an "incessant increment in cases in which intervention programs are implemented without the imposition of instructions, and in contexts that are natural and/or habitual for the users of the program, taking advantage of activities that are spontaneous and/or everyday for them." There have been numerous studies over the last few decades that have dealt with the handling of an intervention in the habitual context of a user, with the different programs having very diverse aims (Weil, 1985; Bryant and Bickman, 1996; Lapresa et al., 2020), and with beneficial effects in terms of the 'normalization' of the user's life (Dvoskin and Steadman, 1994; Roustan et al., 2013; Alcover et al., 2019), studying the interaction in the therapeutic conversation (Arias-Pujol and Anguera, 2017, 2020a,b; Del Giacco et al., 2020).

In this study we present a low intensity level program – a psychoanalytic psychotherapy intervention for a child with severe autism. It is carried out in a natural context and stems from the child's predominantly sensory spontaneous behavior, disconnected from social interaction. The therapeutic technique seeks to help children establish connections between their sensory experiences, their emotions, language and thought, following the approach developed by the prestigious Coromines (Coromines, 1991; Viloca, 2011; Farrés et al., 2020). It is a dyadic program based on the verbal and non-verbal communication between the child and the therapist. An adequate evaluation requires the flexibility and scientific rigor of systematic observation from the mixed method perspective.

In Chacón-Moscoso et al. (2021) an adaptation was carried out of the structural dimensions of low intensity level designs when observational methodology is applied (Anguera, 2003). This is characterized as a scientific method that allows for the study of spontaneous behavior in habitual contexts. Its defining elements are the following: (1) delimiting the aim from a prior needs analysis; (2) building the program depending on the theoretical framework and the intervention context; (3) guaranteeing its usability and sufficiency; (4) proposing a

suitable design; (5) implementing the actions of the program in such a way as to obtain a diachronic record throughout the recommended monitoring; (6) building a non-standard instrument adapted to the object of study; (7) systemizing the information; (8) controlling data quality; and (9) assessing the program with adequate analysis techniques. It should be highlighted that in the last two decades the mixed method perspective has been developed exponentially, generating an important cross-fertilization process in terms of procedure.

We would like to point out one final aspect to be developed, relating to the fact that we are dealing with a case study. Here we attempt to transform something that has acquired negative connotations in the literature of the last few years into a worthwhile opportunity that makes a rigorous, intensive study of human behavior possible.

Contributions of the mixed method perspective in the evaluation of low intensity programs

The mixed method perspective has had a significant impact all over the world throughout the last few years, awakening an exponential interest and generating an extremely high volume of scientific production. As is widely known, a harsh dialectic was produced over decades between those in favor of qualitative and quantitative approaches. Almost a quarter of a century ago, Newman and Benz (1998), set out to explore the interactive qualitative-quantitative *continuum* in research. They considered false and without foundation the dichotomy that many other authors have tried to show, by presenting both options as opposing paradigms and refusing to accept that both qualitative and quantitative strategies are always found in any study. Their approach, focusing on the feedback between qualitative and quantitative analysis, can be considered as an accurate forerunner to the current, complex state of the issue (Anguera, 2022), that forces us to be vigilant whilst moving forward.

Observational studies were initially qualitative, proof of which lies in the main works published in the 1970s and 1980s (Weick, 1968; Hutt and Hutt, 1974; Mucchielli, 1974; Anguera, 1979; Martin and Bateson, 1991), whose aim was to capture the reality in a descriptive way just as it was produced – but which our approach challenges – which led to the belief that qualitative methodology fit with the first half of the process, whilst quantitative methodology should be used in the second half of the process (Sánchez-Algarra and Anguera, 2013). Furthermore, in the last few years things have gone a step further, with the consideration that observational methodology was in itself mixed method (Anguera and Hernández-Mendo, 2016; Anguera et al., 2017) – also in indirect observation (Anguera et al., 2018b) –, and proposing a form of quantitizing as an integration path (Anguera et al., 2020).

The specification of *quantitizing* in observation methodology is founded on Creswell and Plano Clark's, 2011 contribution (3rd ed., 2011), that we especially value:

> 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 dataset within the other so that one type of data provides a supportive role for the other dataset. (p. 7) (the underlining is ours).

This *mixing*, in the CONNECT option, taken both literally and from a wider perspective (Anguera, 2022), is a strong basis for carrying out a reconsideration of *quantitizing* that fits very well within observational methodology.

Indeed, in literal terms, "*connecting* the two datasets by having one build on the other," would imply that one dataset can give rise to another via its transformation. Such a transformation must guarantee the maintenance of its informative quality, whilst modifying the appearance. From a wider perspective, *connecting* allows the alternation of the QUAL-QUAN-QUAL stages; this legitimizes the generic mixed method approach, such that a total integration of qualitative and quantitative elements is achieved (Anguera et al., 2020).

The materialization of *quantitizing* takes place between the first QUAL stage and the QUAN stage (Anguera et al., 2020; Izquierdo and Anguera, 2021), and is accomplished from two fundamental decisions (Anguera, 2017, 2022): (1) establishment of the design dimensions (or response levels, or criteria) (Weick, 1968), that can be deployed in sub-dimensions on different levels; and (2) proposing segmentation criteria of the observed behavior episodes (or textual material, in indirect observation) in observation units (Anguera, 2020, 2021). The two decisions (dimensions and observation units) were developed later, both in direct observation (Anguera, 2017) and indirect observation (Anguera, 2021).

Once these decisions have been taken and the *ad hoc* observation instrument built, the record is created, made up of qualitative data, and will preferably be structured in the form of a code matrix, with columns containing the dimensions (or, where appropriate, the sub-dimensions of the most molecularized level); with each row of the matrix containing the codes corresponding to the co-occurrence of the different dimensions in each unit of behavior. This matrix is essential for the process of *quantitizing* the qualitative data.

Once the data matrix has been obtained, its quality must be controlled via one of the concordance/agreement indexes (Blanco-Villaseñor and Anguera, 2000); and once this is done then the quantitative analysis of the systematized qualitative data is possible (and, among others, the analysis of polar coordinates), thus allowing a complete integration between qualitative and quantitative elements.

Taking this reconsideration into account, the innovative form of *quantitizing* in the implementation and evaluation of low intensity programs implies important methodological benefits (Anguera, 2022). The mixed method greatly vitalizes the collation, management and analysis of information obtained via observation, which previously played a trivial, superficial and incoherent role. An important strengthening of range is achieved from observational methodology (both direct, as in this study, and also indirect), due to it being considered in itself mixed method; hence, in the Introduction we refer to the process of cross-fertilization.

We strongly defend this approach (Anguera, 2022), that has been progressively structured over the last two decades, without evading the attraction of the mixed methods being developed (Sánchez-Algarra and Anguera, 2013; Portell et al., 2015; Chacón-Moscoso et al., 2019; Anguera et al., 2020; Izquierdo and Anguera, 2021); these differing from multi-method studies (Anguera et al., 2018a), constituting an already consolidated culture of research (Anguera et al., in press).

The case study in the evaluation of low intensity programs

The case study has traditionally been considered marginal and with little convening power; in addition to renowned authors such as Gerring (2004, p. 341) stating that "the case study survives in a curious methodological limbo."

It is generically accepted as a study that is intensive, detailed, in-depth, centered on one "case," and focusing on "the particular"; and undoubtedly influenced by the context in which it is located and the theoretical framework that covers it (Anguera, 2018). The case study implies an "intensive study of a single unit for the purpose of understanding a larger class of (similar) units . . . observed at a single point in time or over some delimited period of time" (Gerring, 2004, p. 342), provided that it is carried out with guarantees of suitability (Kuyken et al., 2005; Flinn et al., 2015).

For Stake (1994), the case study does not represent a methodological option, but an object of study option. However, Campbell, in his *Foreword* to Yin (2014), and the thematic development itself of this author, do openly consider it a research method. There have been many dissonant voices in different publications (Anguera, 2018), and there is no consensus concerning its methodological range, although there is a majority who do not consider it a methodology. Tight (2010, p. 329) asks the following question: "Case study is widely referred to and applied within social research, but its status remains unclear. Is it a method, a methodology, a strategy, a design, an approach, or what?".

In our view (Anguera, 2018), it is not a methodology, but it is possible to apply diverse methodologies to a single case. Aside from existing typologies (Stake, 1994; Thomas, 2011; Yin, 2014), the logic of the single case is intra-case by nature (Hilliard, 1993) and permits the consideration of a diachronic perspective, whilst at the same time emphasizing the richness of the context in the real world in which the phenomenon is produced.

There is obviously an inherent weakness in the case study, relating to the non-replication of results, which is totally logical according to its own aim. Its focus of attention is found precisely in the opposing situation, focused on the results of one single case which is studied in depth, and is shown in this study. In our view, the strength of a methodology that is appropriate and adapted to the characteristics of the case study and the profile of the case itself compensates for this weakness.

The methodological criticism that the case study has received focuses essentially on the problem of a lack of representativeness; although the case is an individuality with an initial presumption of singularity, which is studied intensively. As Sandelowski states (Sandelowski, 1996, p. 527), "the analyst works to discern what elements comprise the case and, more importantly, the way they come together uniquely to characterize the case," which suggests that researchers should establish the profile of the case (Anguera, 2018).

The profile of the case we present below corresponds to that of a child with severe autism. The therapist provokes interaction by initiating communicative turns of protoconversation, verbalizing sensations and emotions that the child feels or could feel, and also dramatizing them (Trevarthen and Aitken, 2001; Viloca, 2003; Alcàcer and Viloca, 2014). Unlike other studies in which we narrate the intervention and clinical improvement of the patient (Arias-Pujol et al., 2015a,b), on this occasion we are interested in focusing on the technique used by the therapist. The suitable methodology is systematic observation, this being equipped with design that supports an intensive approach in the study of perceptible behaviors; carrying out continuous recordings throughout the observation sessions, and being able to manage behaviors arising from different dimensions/sub-dimensions, some specific to the therapist's role and others to the child's actions. In other words, we transform into a methodological opportunity that which has traditionally been the biggest weakness in case studies.

For their part, Edwards et al. (2004) state that they do not wish to differentiate between qualitative and quantitative approaches in the case study. As Stake (1994, 1995) indicates, the case study can be qualitative (Hilliard, 1993), or quantitative (Kent, 2009), or a combination of both (mixed methods are currently conceived within a continuum between the qualitative and quantitative poles); although he leans more toward a clear qualitative predominance (Stake, 2005). Our proposal, in line with the above, involves taking advantage of the appearance of a third way, that of the aforementioned mixed methods, which places us in a privileged position of integrating qualitative and quantitative elements.

Aim

To analyze, from the mixed methods perspective, a case study consisting of low intensity psychotherapeutic intervention focused on a child with a diagnosis of severe autism who interacts with the psychotherapist; and in which we aim to identify which actions on the part of the therapist stimulate social interaction from the child.

Design

Observational methodology was applied. The observational design is Nomothetic/Follow-up/Multidimensional (N/F/M) (Anguera et al., 2001; Sánchez-Algarra and Anguera, 2013): nomothetic because we studied the interaction between therapist and autistic child, with inter-session follow-up (three sessions) and intra-session follow-up (because each session was recorded continuously from start to finish); and multidimensional since the complexity of the aim required the application of various dimensions that were included in the observation instrument.

Participants

There were two participants:

The patient was a 4-year-old child with a diagnosis of autism spectrum disorder (ASD), according to the clinical criteria of DSM-5 (American Psychiatric Association, 2015), of a severe type according to the results obtained from the ADOS (Autism Diagnostic Observation Schedule; Lord et al., 2000). The child had no language, although did emit sounds and some syllables forming echolalia.

The therapist was a clinical psychologist with training and experience in psychoanalytic psychotherapy with children.

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 child and the child's family were informed that they were being filmed. They were shown the location of the video cameras, which were positioned discretely to minimize reactivity bias. Written informed consent was also obtained from the parents of the minor.

Program intervention plan

The psychotherapeutic intervention was designed for a child with severe ASD. It consisted of 20 weekly sessions of 45 min in length, and was focused on stimulating reciprocal social interaction in the child, arising from the relationship. We adapted the psychoanalytic technique described by Coromines that promotes a process of differentiation and of interest in the "Other" via a shared emotional experience between therapist and patient (Viloca, 2003; Farrés et al., 2020). The sessions took place in Carrilet Treatment Center (Barcelona, Spain), attached to the educational and therapeutic center.

Instruments

Recording instrument

All the sessions were filmed using a video camera installed in the therapy office of the Carrilet Treatment Center attached to the Educational and Therapeutic Center where the child was schooled.

Observation instrument

We used an *ad hoc* observation instrument, as a field format modality combined with category systems, adapted by Bachs and Arias-Pujol (Bachs, 2019) from a previous study (Arias-Pujol et al., 2015b) and recoded for this new study. The instrument has two dimensions for the child: reciprocal social interaction (RSI) and non-reciprocal social interaction (N_RSI) and 12 for the therapist (see **Table 1**).

With the aim of studying the therapist-patient relationship, the sessions were broken down into units, with the adoption of a primary dialogic criterion, and a secondary criterion that differentiated verbal, vocalized and non-verbal behavior in the transcript (Anguera, 2020). **Table 2** shows some fragments of the coded clinical material from sessions 9, 16, and 20.

Procedure

Inter observer agreement

The inter observer agreement was calculated via Cohen's kappa coefficient (κ) (Cohen, 1960). Data quality control was performed using the free program GSEQ. An agreement of 85.6% was obtained in the codification of the child's behavior and of 90.4% in that of the therapist, values considered "almost perfect" according to the criteria of Landis and Koch (1977) and Bachs (2019).

Data analysis

Analysis of polar coordinate

The aim is to apply the analytical technique of polar coordinate analysis, seeking a possible relationship of activation/inhibition between the behaviors of the therapist and the child that will be quantitatively calculated from the qualitative recordings carried out initially at three different points of the intervention, plotting them via vectors.

Polar coordinate analysis allows us to obtain an inter-relational map of the connections detected between the different

TABLE 1 Observation instrument of ASD child and psychotherapist in psychoanalytic psychotherapy.

Name and definition of categories	Code
Observation instrument of ASD child	
Non-reciprocal social interaction (N_RSI): Actions carried out by the child in relation to an object or toward himself/herself.	
(1) Stereotypes: repetitive behaviors that always follow the same pattern. They can be of different types: (1.1) **Motor:** highly repetitive movements with consistent action patterns (by emotion or sensation) *Example*: jumps [on themselves], arm flapping and gesticulation, rocking. (1.2) **Vocal:** highly repetitive vocalizations with consistent intonation patterns such as humming, various vocal repetitions (By emotion or sensation) *Example*: tititi, hmmm, ah ahah, d-gæ-d-gæ-d-gæ, iiiiii, tacataca. (1.3) **Visual:** strange eye movements, such as looking sidelong at the camera, suddenly diverting the gaze after maintaining eye contact with the therapist, objects, etc. *Example*: looking sidelong from one side to the other.	MS VS VIS
(2) Erratic Behavior: constant wandering with no set purpose. It occurs when the child moves, walks around the office without a clear intent. When the child grasps an object when there is no subsequent functional intent. He/she takes it and then leaves it. *Example*: wandering without exploring, jumping around the room, grasping and abruptly throwing an object.	EB
(3) Auto Sensory: an action carried out by the child that provokes a sensation, with no exploratory functional purpose. *Example*: hitting with an object, abruptly pulling an object, continuously and persistently touching different objects, touching the face, taking a handkerchief and wiping the face with it, tracing their sneakers with the hands or the edge of the chair with the fingers, keeping one hand on the wall when walking or leaning on it with the whole body. Despite appearing as repetitive behavior, it is not considered stereotypical because: (a) the behavioral pattern is not always identical. (b) it is exploratory via sensation, and c. it is of low frequency *Covering the ear would be included due to being exploratory via sensation.*	AS
(4) Functional Intent: normal actions with objects, the objects are used with a coherent purpose for which they were created. *Example*: order, collect, take out, put on, take off, shake a box of blocks, take crayons from a box, open a box, arrange a chair. . . *In the case of continuous repetition of behaviors FI, consider the category "Solitary play (SP)".*	INT
(5) Gaze: (5.1) Attentive object: when the child gazes attentively at an object. The object has entered the field of vision and is noticed. It is an action that has exploratory aspects. *Example*: "He/she takes the box with both hands + **stares at it** + turns it upside down to shake out the blocks + puts one hand inside to remove them (he/she achieves this) + continues shaking with force". *Example*: "He/she sits down + touches the dice with his/her fingers and makes them move + "oh" + continues touching the dice with the fingers + **looks at them very attentively.**" (5.2) Blank stare: when the child remains staring at the floor, the ceiling or the wall when there is no specific object at which to direct the gaze.	AGO BS
(6) Solitary Play: there is an appropriate use of the toy. The relationship with the object (such as blocks for building a tower) is more structured and sequenced than in the functional intent category. This category includes those sequences that are repeated in various turns, becoming a ritualized game despite the therapist's intervention. It also includes exploratory play, i.e., when the child manipulates and relates to the object, also permitting the predominance of sensory self-stimulation throughout the action. The child shows an interest, his/her behavior has a purpose in itself and at the same time is very sensory (wants to build the tower, wants to make it taller, wants to hit the balloon, scribble with the crayons. . .). *Example*: the child begins a tower + puts down two blocks + takes out a block + changes one block for another + mmmmm (SP + VS). *Example*: the child moves the trucks from one side to the other + "brumbrum" + lifts them up and leaves them on the floor, organizing them.	SP
(7) Normal Actions: actions that the child carries out, that are related to being in one place (with no auto sensory purpose). *Example*: gets up, sits down, turns, crouches, lifts the head, and repositions the body.	NA
Reciprocal social interaction (RSI): They are actions that the child carries out, bearing in mind the therapist or in response to an action by the therapist.	
(8) Demand: when the child addresses the therapist for the purpose of asking for something. *We differentiate between a continuum of more fusion to more differentiation ME – NOT ME* It can be: (8.1) **Instrumentalization/Instrumentalized Demand**: undifferentiated demand (no me–you difference). It occurs when the child uses the other as an object in order to achieve a purpose, in something which he/she knows how to do motively. Also when the child uses the other as an object to achieve something he/she can't do for himself/herself. *Example*: the child wants to open the bubble container. He/she takes the therapist's hand with both hands and moves them toward the container for the therapist to open it and then steps back waiting for T to do it (the child knows how to open it, but takes the therapist's hand as his/her own limb). *If the instrumentalization is followed by a VS or MS, consider categories VOCD or NVD.* (8.2) **Non-verbal Demand**: when the child communicates an intention or desire and uses his/her body to achieve it; for example touching the object, showing it, pointing to it, etc. *Example*: the child looks at T and takes the bubble container from the table. He/she moves it (making a noise) and gives it to the therapist. *Example*: the child stands behind the therapist's chair placing both hands on her/him. He/she uses force to move the therapist because he/she wants T to move. Then he/she steps back, waiting for a response. (8.3) **Vocal Demand:** a demand from the child accompanied by a vocalization, albeit stereotyped. *Example*: "He/she moves away to go and look for the balloon + takes it + gives it to T (who keeps hold of it) + guttural sound" (VS) (8.4) **Verbal Demand:** a demand from the child accompanied by verbalization or word approximation.	I NVD VOCD VD
(9) Response to a demand (9.1) **Active response/following proposal**: when the child responds actively to a demand or to the therapist's stimulation proposal. *Example*: the therapist inflates a balloon, moves it (making a noise) and the child approaches and takes it very carefully. *Example*: the therapist says: "oh, have we changed?" (referring to how P has changed the blocks around) and the child looks at T and changes them back again. *Example*: the therapist: partly opens the box + "Now you can open it" + looks at P, and the child: looks at the box + bites lip + smiles + opens the box with both hands. (9.2) **No active response/no following proposal**: when the child does not respond to the demand or the therapist's stimulation proposal. *Example*: T tries to give the child the bubbles, but P rejects them abruptly. *Example*: T: "Tc-tc-tc" (paused) while T tickles the child (from lower back up to the neck); P: Turns body forwards to avoid it + "iiii" (with a high-pitched intonation).	AF NAF

(Continued)

TABLE 1 (Continued)

Name and definition of categories	Code
(10) Proxemic Behavior: the child's trajectory with the therapist, in response or not to an action or verbalization from the therapist. (10.1) Moves closer. (10.2) Moves away.	APRO DPB
(11) Physical Contact. (11.1) Brief *Example*: child touches the therapist. (11.2) Maintained *Example*: child hugs the therapist, sits on T's lap.	BPC
(12) Joint Attention: when the child and the therapist share the same focus of attention, whether it be an object or an activity sequence. There is a shared pleasure and the action takes turns: (12.1) **Child shows** or points with the purpose of sharing the attention to an object. Thus, a shared experience with the therapist is created. Example: placing an object where the therapist can see it, holding the object in front of the therapist, pointing to an object with the aim of the therapist seeing it, or giving an object. 12.2 **Child draws.** (12.3) **Protoconversation:** a type of dialogue via vocalizations and/or verbalizations that follow turns of intervention and that make sense within the context in which they are expressed. They are different from the verbal stereotype in that there is a communicative intention. Types: (12.3.1) Vocalization/word approximation: The production has at least a vocal similarity with a word which makes sense in the context. It includes onomatopoeia. Example: Ahh! Example: "aul" (azul - *blue*). (12.3.2) Word: The production is clear. (12.3.3) Sentence approximation: 2 or more consecutive, related words. "Esto e aul" ("*This is blue*").	MPC SH DRA WA WO SAP
(13) Gaze: when the child's gaze is in relation to the therapist. Eye contact. The child looks at the therapist to establish eye contact. It is categorized as EC although brief – a "fleeting glance" (Viloca, 2003).	EC
(14) Imitation: This category should be produced in response to a stimulation, imitation or proposal by the therapist. (14.1) Vocal – Verbal (14.2) Non-verbal	VI NVI
(15) Facial Expression (FEx). (15.1) Rejection: when the child shows a facial expression of displeasure, anger, upset or sighs. (15.2) Joy: when the child shows a facial expression of joy, laughs or smiles.	FEXR FEXJ
Observation instrument of the psychoanalytical psychotherapist	
(1) **Verbalize:** Put into words. Therapist uses language to bring the child closer to the symbolization process (following the psycho-pedagogical scheme of Dra. Coromines, 1991; Viloca, 1998, 2003). (1.1) **Describes** a behavior or an object. Suggests to the child a feeling or a desire. It puts words to the child's behaviors, feelings, emotions, desires, and thoughts. Given the hypersensitivity of these children, it includes naming interferences and giving information from the context such as ambient noises, if the material makes noise or if an object fails. Examples: (a) Name object and sensation: "Oh! It's a sponge, it's hard and scratchy," (b) Behavior: "Ah, you want it all for yourself!", "You have seen the bubbles and left the dice," "You have thrown them all," (c) Desire: "Shall we blow?", "More?", (d) Context: "What a loud noise, huh? Tocotoco. They made noise," "What noise do you make, huh? You hear noise outside and you make a noise, tacataca!". (1.2) **Offers help:** the therapist verbalizes an offer of help to the child when T sees that the child needs it. Example: "Let's see, can I help you . . . open the box?" + opens the box, "Do you want to open **the box?"** + pauses + "the crayons?" (1.3) **Anticipates:** the therapist anticipates actions that will be carried out soon, in the immediate or distant future *Example*: "Do you want to open the box?" + pauses + "**The crayons?**"; The therapist says: "One, two and. . ." singing + inhaling air + blowing the bubbles; The child takes the cup and puts it in his/her mouth. Meanwhile, the therapist says: "Let's put some water in it" + takes another cup from the box. (1.4) **Reminds:** the therapist verbalizes actions that have been carried out previously. *Example*: T looks at the child + **"The other day Biel was singing."** + starts to sing a melody. (1.5) **Gives support:** the therapist supports, encourages, congratulates, grants. T expresses approval of the child's behavior with words or gestures. Example: "Well done!", "Thank you!", "Very good!" (1.6) **Repeats to show understanding:** the therapist repeats what the child has just said in a similar manner in order to start a dialogue, although the child doesn't follow T's lead, T tries to make sense of the child's verbal communications, even though they might not be very clear. There is recognition of the child's verbalizations. Similar to the protoconversation that takes place between a parent and baby. *Example*: the child looks at the therapist + holds up the object + "the blocks"; the therapist looks at the child + nods head + "The blocks". (1.7) **Interpretation** (INT): the therapist verbalizes the emotion and goes further, giving meaning, a possible explanation of what the child does, says or feels. T connects what the child does with what she/he imagines the child could be feeling. T goes a step further in understanding the facts and emotional experience the child is living. Examples: The child takes a cup from the box and starts to drink. When he/she stops he/she says "hm" with a disgusted face. The therapist then laughs and says: "Ecs, is it not good?"; The child is out of camera shot, but the therapist says: "Ui, ui what a face!"; The child turns his/her head and laughs; the therapist: "Ah, you're smiling" and also laughs.	VDE VOH VAN VREM VSUP RMU INT
(2) Vocalization: (2.1) Exclamatory elocution: Sounds or onomatopoeia that the therapist makes, expressing or highlighting an emotion. Examples: "Oh!", "¡Ala!", "Oi!", "Ehi!!". (2.2) Sings	EE, SI
(3) Imitation: the therapist copies the child's behavior. This category includes the three levels of imitation described in (annex 1): sensorial resonance, significant imitation and double material. It also includes actions in which the therapist follows the child's activity or adds to what the child is doing (looking for a turn-taking game). Example: The child is piling up blocks and the therapist starts to pile up blocks next to him, as in taking turns (the child does not modify his/her behavior, it is the therapist who adapts). (3.1) Verbal imitation. (3.2) Non-verbal imitation.	VIT NVIT
(4) Stimulation: it can be verbal or non-verbal, here we do not differentiate. At times the therapist makes an active change in the setting. (4.1) **Gives:** the therapist gives the child and object to facilitate play, attract the child's attention or initiate a game. *Examples*: The therapist says: "Shall I open the crayons?" + opens them + "There you go" + **stretches out his/her arm to give them to child;** "Here you are" + **T gives the balloons to child.** (4.2) **Shows:** the therapist shows the child an object or an action, with the aim of directing his/her attention. *Example:* "Sheets of paper" + **points to them** + sits in the chair. (4.3) **Directs attention**: the therapist directs the child's attention toward an object, a noise or an action in order to initiate a new activity, a game, or to interrupt something repetitive. *Example*: when the therapist shakes an object (balloons, blocks, etc.) to get the child's attention. (4.4) **Directs behavior:** when the therapist gives an order, asks for, guides, shows or encourages the child to do something in a particular way. (4.5) **Proposes:** the therapist starts an activity, expecting the child to follow or, faced with the child's indecision, proposes a joint game with material that the child has not yet used. Or also when the child is doing nothing or performing a repetitive action. *Example*: The child moves the blocks + places a block (yellow) on the floor + takes another block (red) with one hand and with the other takes the therapist's hand and moves it closer + "eh." Thus, T responds: "Aaah" + takes the red block and puts it on top of the red block + "Here?"	SG SSH SDA SDB SPRO

(Continued)

TABLE 1 (Continued)

Name and definition of categories	Code
(5) Non-verbal Behavior (NVB). (5.1) Facial expression: the therapist shows with a facial expression (surprise, disgust, smiles.) or the movement of hands and torso (gesture). (5.2) Gaze: We consider that the therapist tends to look at the child, but if we see in the transcript that it is specified, then we will code it.	FEXT GA
(6) Auxiliary Functions: facilitating actions carried out by the therapist, such as moving an object closer that the child needs, opening a bag, uncovering the box. When there is no clear demand from the child, since in that case it would be "responds to demand." When the therapist senses the need, functioning as an auxiliary Me.	AUXF
(7) Responds to the instrumentalized demand/proposal: takes notice by responding to an action of demand more or less instrumentalized or to a proposal from the child.	RID
(8) Does not respond to the instrumentalized demand: the therapist does not agree to do what the child is expressing with a clear petition or proposal (albeit non-verbal).	NRID
(9) Proxemic Behavior: moving closer to or away from the child. (9.1) Closeness *Example*: T lies on the floor next to the child. (9.2) Distance *Example*: T moves away from the child.	APROT DT
(10) Physical Contact: the therapist touches, holds, tickles, caresses, hugs. It can be a brief moment or a maintained contact that can form part of a game or cuddles. (10.1) Brief *Example*: the therapist tickles the child, touching him/her briefly. (10.2) Maintained *Example*: T strokes hand over the child's back, head. T holds the child in her/his lap.	BPCT MPCT
(11) Normal Actions: actions carried out by the therapist that are related with preserving the "*setting,*" such as adjusting her/his posture to that of the child, curling up, getting comfortable, etc.	NAT
(12) Phatic Function: the therapist encourages the child to continue speaking or waits for a reply from him/her, giving over her/his turn. *Example*: nods head waiting for a response + "hmhm"; Child takes the bubbles from the box; T smiles widely and assents.	PF

Adapted from Bachs (2019).

TABLE 2 Fragments of the coded clinical material.

Session 9
They are playing with a balloon which has deflated and fallen onto the floor. The child wants the T to inflate it again. **C:** He/she bends down (NA) + picks up the balloon (INT) + turns to the T (NA) + and goes towards her/him (APRO) + makes the gesture of giving the balloon (NVD) **T:** He/she looks at the child (GA) + makes a surprised face (FEXT) + "the pink balloon?" (VDE) + picks up the balloon (RID) **C:** "Tii" (WA) + brushes the wall with his/her back (SA) + "Buh-ah!" (WA) + looks at the T (EC) + takes T's hand that is holding the balloon and moves it towards the therapist's mouth (I) **T:** "Oh" (EE) + "I'm going to blow up the balloon" (VAN) + inflates it (RID)
Session 16
The child places blocks on top of each other, building a tower. The T attempts to join in the game. **T:** "Oh" (EE) + "a very high tower" (VDE) + takes one of the blocks and moves it towards the child to give it to him/her (SG) **C:** "Oh" (WA) + ignores the block (NAF) **T:** "Here you are" (SG) **C:** He/she moves his/her arm and takes the T's block with one hand (AF) and picks up another with the other hand (SP) **T:** "Very good" (VSUP) + Smiles (FEXT) **C:** "TcTcTc" (VS) + Keeps both blocks, one in each hand, and adds one of the blocks to the tower (SP) **T:** Look (GA) **C:** He/she adds the other (SP) + "TcTcTc" (VS)
Session 20
Through imitation, a sort of dance is created between the child and the T of moving closer together and further apart. They are holding hands. The T imitates the child's vocal stereotypies. **T:** "Taaa taaaan." (VIT) **C:** Looks at the T (EC) + Drops his/her hands and moves away toward the desk (DPB) + "Taaa taaa" (VS) **T:** He/she walks toward the other side, moving away (DT) + looks at the child (GA) + "Taaa taaa." (VIT) + smiles (FEXT) **C:** Walks toward the T (APRO) + takes her/his hand (BPC) + twists around (SA) + takes T's hand with his/her hand (BPC)

T, therapist; C, child.

categories. It is a robust analytical technique developed by Sackett (1980) that is based on the sequentiality of the qualitative records obtained. The focal behavior – located in the center of each of the 'maps' that will be created – must be identified in each analysis, and is proposed depending on the desired aim and the conditioned behaviors; these being all those about which we want to know whether there is an associative relationship with the focal behavior. The associative relationships between the focal behavior and each one of the conditioned behaviors incorporate two perspectives: prospective

(from the focal behavior forwards) and retrospective (from the focal behavior backwards). We should clarify that we are not referring to a classical retrospective perspective, but to the genuine retrospectivity proposed by Anguera (1997), which has since been consolidated. The calculation required to apply the prospective and retrospective perspectives generates a huge volume of partial results, and Sackett (1980) knew how to exploit the possibilities of the Z_{sum} parameter proposed by Cochran (1954) as an important data reducer. The calculation of the prospective and retrospective Z_{sum} values, in accordance with Sackett's approach (Sackett, 1980), allows us to obtain the values that correspond to the length of the vector and its angle, which can be graphically represented. The vector angle (that will correspond to one of the quadrants I, II, III, and IV), allows us to interpret the nature of the relationship that exists between the focal behavior and the respective conditioned behavior, while from the length of the vector we can interpret the intensity of said relationship depending on statistic significance.

We have presented all these calculations systematized in our case study (see **Table 3**), in such a way that the conditioned behaviors appear for each focal behavior; and for each conditioned behavior information is presented, correspondent to the quadrant in which the vector is found: prospective and retrospective values of the Z_{sum}, ratio, length of vector (that is crucial for knowing its significance, since if it is > 1.96 it is significant, and if it is > 2.58 it is very significant), significance, and vector angle.

Polar coordinate analysis was carried out using the free program HOISAN 2.0 (Hernández-Mendo et al., 2012), and additionally R (Rodríguez-Medina et al., 2022) in order to obtain a graphic optimization in the vector representation.

In this case study, the focal behaviors were selected from the therapist's highest frequency categories or dimensions (see **Table 1**). The analysis was carried out with seven focal behaviors: "Verbalization," "Vocalization," "Stimulation" (combining all the categories of each dimension); "Verbal imitation," "Non-verbal imitation," "Facial expression" and "Gaze." In terms of conditioned behaviors, the child's 28 categories were included from the observation instrument, excluding the codes SAP (sentence approximation) and VD (verbal demand) due to low frequency.

For the analysis, the behaviors of the therapist and the child were recorded in the first 20 min of sessions 9, 16, and 20.

Descriptive study of clinically favorable behaviors

From the results obtained in the polar coordinate analysis, the conditioned behaviors (the child's actions) were selected, that were significantly activated or inhibited prospectively. They were grouped by dimensions of reciprocal social interaction (RSI) or non-reciprocal social interaction (N_RSI).

The dimension "Reciprocal social interaction" consists of the categories: Instrumentalized Demand (I), Non-verbal demand (NVD), Vocal demand (VOCD), Active response (AF), No active response (NAF), Moves closer (APRO), Moves away (DPB), Brief Physical Contact (BPC), Maintained Physical Contact (MPC), Child Shows (SH), Child Draws (DRA), Word Approximation (WA), Word (WO), Gaze (EC), Vocal-verbal Imitation (VI), Non-verbal imitation (NVI), Facial Expression Rejection (FEXR), Facial Expression Joy (FEXJ). The dimension "Non-reciprocal social interaction" consists of the categories: Motor Stereotypes (MS), Vocal Stereotypes (VS), Visual Stereotypes (VIS), Erratic Behavior (EB), Auto Sensory (SA), Functional Intent (INT), Gaze (AGO), Blank Stare (BS), Solitary Play (SP), Normal Actions (NA). The following categories were excluded: NA, referring to normal actions in relation with the therapeutic framework without a sensory purpose, and INT, referring to actions with objects appropriate to the purpose for which they were created.

For each of the therapist's focal behaviors the percentage of RSI and N_RSI behaviors that were activated and inhibited was calculated. The fact that one of the therapist's focal behaviors activated RSI behaviors and inhibited N_RSI behaviors was considered clinically favorable. The chi-squared test was used to determine whether the clinically favorable behaviors activated by the therapist were statistically significant, ($p < 0.05$).

Results and discussion

In the sections below, we describe the relationships detected between interventions by the therapist and the child's behaviors using polar coordinate analysis and a descriptive study of clinically favorable behaviors.

Analysis of polar coordinate

Significant results were obtained in activation/inhibition relationships between all the therapist's and the child's behaviors.

Table 3 show the level of significance of the focal behavior "Verbalizes," "Vocalizes," "Verbal imitation," "Non-verbal imitation," "Stimulates," "FEXT," "GA" as the main analysis of sessions 9, 16, and 20.

Figures 1–4 shows the significant vectors for all the focal behaviors in each of the therapist's seven actions.

Descriptive study of clinically favorable behaviors

Table 4 shows the type and percentage of behaviors that prospectively activated or inhibited each of the therapist's dimensions.

10.3389/fpsyg.2022.998658

TABLE 3 Table of parameters corresponding to the analysis of polar coordinates, with the focal behavior "Verbalizes," "Vocalizes," "Verbal imitation," "Nonverbal imitation," "Stimulates," "FEXT," and "GA" as the center of the analysis and all the others as conditional of sessions 9, 16, and 20.

Focal behavior (therapist)	Conditioned behavior (child)	Quadrant	Prospective zsum	Retrospective zsum	Ratio	Length	Significance	Angle
Session 9								
Verbalizes	Erratic behavior_EB	n	−0,4	2,15	0,98	2,19	*	100,59
	Functional intent FI	i	0,22	3,07	1	3,08	**	85,9
	Attentive object_AGO	I	4,09	0,54	0,13	4,13	**	7,52
	Normal actions_NA	I	3,4	0,59	0,17	3,45	**	9,78
	Demanda_DNV	II	−1,08	4,57	0,97	4,69	**	103,27
	No active response/no following proposal_NAF	I	0,01	2,29	1	2,29	*	89,68
	Joint attention_WO	IV	1,77	−0,88	−0,44	1,98	*	333,6
Vocalizes	Stereotypes_VS	HI	−0,97	−2,46	−0,93	2,65	**	248,37
	Erratic behavior_EB	IV	3,49	−0,25	−0,07	3,5	**	355,87
	Attentive object_AGO	N	−0,81	2,32	0,94	2,46	*	109,33
	Solitary play_SP	HI	−1,83	−1,21	−0,55	2,19	*	213,42
	Instrumentalization/instrumentalized Demand_I	I	1,55	2,46	0,85	2,91	**	57,68
	Demanda_DNV	I	2,6	1,8	0,57	3,16	**	34,73
	Active response/following proposal_AF	I	1,01	3,38	0,96	3,53	**	73,37
	Proxemic behavior DPB	IV	2,94	−0,43	−0,15	2,97	**	351,61
Verbal imitation	Stereotypes_VS	I	1,27	2,41	0,88	2,73	**	62,15
	Autosensorialidad AS	N	−0,79	1,89	0,92	2,05	*	112,7
	Solitary play_SP	HI	−1,74	−1,65	−0,69	2,4	*	223,53
	Demanda_DNV	i	2,01	0,5	0,24	2,07	*	13,86
	Joint attention_WA	IV	2.48	−0,85	−0,32	2,62	**	341,06
	Facial expression_FEXR	n	−0,45	2	0,98	2,05	*	102,78
Nonverbal imitation	Stereotypes VS	i	3,44	0,38	0,11	3.46	**	6,29
	Erratic behavior_EB	HI	−1,72	−1	−0,5	1,99	*	210,21
	Autosensorialidad_AS	i	1,16	2.94	0,93	3,16	**	68,39
	Attentive object_AGO	II	−0,13	1,97	1	1,97	*	93,71
	Solitary play_SP	HI	−1,55	−1,54	−0,7	2,18	*	224,78
	No active response/no following proposal_NAF	i	0,78	2,27	0,95	2,4	*	70,95
	Joint attention_DRA	i	5,66	6,96	0,78	8,96	**	50,89
	Joint attention_WO	II	−0,4	2,16	0,98	2,19	*	100,59
Stimulates	Stereotypes_MS	i	1	2,18	0,91	2,4	*	65,2
	Stereotypes VS	HI	−2,57	−1,44	−0,49	2,95	**	209,32
	Attentive object AGO	i	4,15	0,72	0,17	4,21	**	9,79
	Solitary play_SP	HI	−1,09	−2,32	−0,91	2,56	*	244,87
	Instrumentalization/instrumentalized	n	−0,52	2,01	0,97	2,07	*	104,56
Demand_I								
	Vocal demand VOCD	I	2,34	0,81	0,33	2,48	*	19,01
	Active response/following proposal_AF	I	5,09	1,78	0,33	5,39	**	19,28
	Proxemic behavior_DPB	I	1	2,18	0,91	2,4	*	65,22
	Joint attention_WA	n	−0,74	3,33	0,98	3,41	**	102,61
	Imitation_NVI	i	0,81	2,35	0,95	2,48	*	70,96
FEXT	Stereotypes_MS	i	2,02	1,87	0,68	2,76	**	42,81
	Instrumentalization/instrumentalized Demand_I	i	2,28	0,06	0,03	2,28	*	1.47

(Continued)

10.3389/fpsyg.2022.998658

TABLE 3 (Continued)

Focal behavior (therapist)	Conditioned behavior (child)	Quadrant	Prospective zsum	Retrospective zsum	Ratio	Length	Significance	Angle
	Vocal demand_VOCD	IV	4,76	−0,4	−0,08	4,77	**	355,15
	Active response/following proposal_AF	i	1,64	2,53	0,84	3,02	**	57,03
	Proxemic behavior_APRO	i	2,15	1,26	0,51	2,5	*	30,41
	Proxemic behavior_DPB	i	1,96	2,85	0,82	3,46	**	55,44
	Joint attention_WO	IV	2,18	−0,4	−0.18	2,22	*	349,53
	Imitation_NVI	IV	2,19	−0,4	−0,18	2,22	*	349,55
GA	Stereotypes_VS	I	4,49	5,02	0,75	6,73	**	48,19
	Autosensori alidad_AS	I	1,72	3,14	0,88	3,58	**	61,25
	Functional intent_FI	I	2,13	1,23	0,5	2,46	*	30,06
	Attentive object AGO	III	−1,97	−0,8	−0,38	2,12	*	202,05
	Blank stare_BS	I	1,68	1,74	0,72	2,42	*	45,91
	Solitary play_SP	I	5,69	4,81	0,65	7,45	**	40,21
	No active response/no following proposal_NAF	III	−1,09	−1,65	−0,84	1,98	*	236,72
	Joint attention DRA	III	−2,12	−1,04	−0,44	2,36	*	206,2
	Joint attention_WA	I	2,64	0,6	0,22	2,71	**	12,79
	Facial expression_FEXR	I	1,77	1,81	0,71	2,53	*	45,53
Session 16								
Verbalizes	Stereotypes_MS	n	−0,62	2,21	0,96	2,3	*	105,57
	Stereotypes_VIS	II	−1,06	2,29	0,91	2,52	*	114,88
	Functional intent_FI	HI	−2,73	−0,93	−0,32	2,89	**	198,78
	Blank stare_BS	i	4,52	4,01	0,66	6,04	**	41,55
	Solitary play_SP	HI	−0,72	−1,91	−0,94	2,04	*	249,34
	Normal actions_NA	i	2,82	1,9	0,56	3,4	**	33,99
	Proxemic behavior_APRO	i	1,27	2,58	0,9	2,88	**	63,77
	Joint attention_WA	IV	2.47	−0,5	−0,2	2,52	*	348,56
	Gaze_EC	1	2,09	0,83	0,37	2,25	*	21,6
Vocalizes	Functional intent FI	HI	−2,06	−1,84	−0,67	2,76	**	221,76
	Attentive object_AGO	II	−0,25	2,12	0,99	2,13	*	96,73
	Blank stare_BS	i	2,41	0,65	0,26	2,5	*	15,13
	Normal actions_NA	IV	2,77	−0,09	−0,03	2,77	**	358,09
	Active response/following proposal_AF	I	2,71	0,28	0,1	2,72	**	5,94
	No active response/no following proposal_N7	\l I	1,71	1,71	0,71	2,42	*	45,02
	Gaze_EC	I	0,12	2,28	1	2,28	*	87,04
Verbal imitation	Functional intent_FI	I	2,66	1,7	0,54	3,15	**	32,54
	Solitary play SP	I	2,31	0,37	0,16	2,34	*	9,18
	Instrumentalization/instrumentalized demand_I	I	2,36	2,36	0,71	3,34	**	45,09
	Joint attention_WA	n	−0,94	2,45	0,93	2,62	**	111,06
	Joint attention_WO	II	−0,38	2,36	0,99	2,39	*	99,11
Nonverbal imitation	Solitary play_SP	i	0.14	2,76	1	2,77	**	87,17
	No active response/no following proposal_NAF	IV	2,7	−0,34	−0,12	2,72	**	352,87
Stimulates	Stereotypes_MS	HI	−1,35	−1,46	−0,74	1.99	*	227,31
	Stereotypes_VIS	i	0.62	2	0,95	2,09	*	72,66
	Autosensori alidad_AS	1	0,2	2,28	1	2,29	*	84,9
	Solitary Play_SP	i	1,82	1,3	0,58	2,23	*	35,53
	Active response/following proposal_AF	IV	3,21	−1,23	−0,36	3,44	**	339,05
	Proxemic behavior_APRO	III	−1,67	−1,67	−0,71	2,36	*	225

(Continued)

TABLE 3 (Continued)

Focal behavior (therapist)	Conditioned behavior (child)	Quadrant	Prospective zsum	Retrospective zsum	Ratio	Length	Significance	Angle
FEXT	Active response/following proposaLAF	I	0,6	1,99	0,96	2,08	*	73,12
	Joint attention_WO	n	−0,34	2,72	0,99	2,74	**	97,08
GA	Stereotypes_MS	i	2,08	0,85	0,38	2,25	*	22,13
	Stereotypes_VS	i	2,54	2,59	0,71	3,63	**	45,54
	Autosensorialidad_AS	i	2,56	3,07	0,77	3,99	**	50,17
	Functional intent FI	i	2,55	2,54	0,71	3,6	**	44,94
	Attentive object_AGO	i	2.16	2,18	0,71	3,07	**	45,33
	Blank stare_BS	HI	−1,93	−2.46	−0,79	3,13	**	231,82
	Solitary play_SP	i	2,42	4	0,86	4,68	**	58,79
	Normal actions_NA	i	1,31	2,22	0,86	2,58	*	59,52
	Proxemic behavior APRO	1	1,29	1,51	0,76	1,99	*	49,5
Session 20								
Verbalizes	Stereotypes_VS	in	−1,73	−1,28	−0,59	2,15	*	216,45
	Erratic behavior_EB	HI	−1,4	−1,38	−0,7	1,97	*	224,51
	Solitary play_SP	i	1,06	1,68	0,85	1,98	*	57,69
	Normal actions NA	HI	−1,82	−1,61	−0,66	2,43	*	221,48
	Vocal demand_VOCD	i	0,31	2.1	0,99	2,12	*	81,52
	Active response/following proposal_AF	i	2,08	1,65	0,62	2,65	**	38,43
	Physical contact_BPC	i	0,46	3,61	0,99	3,64	**	82,76
	Joint attention_DRA	HI	−1,4	−1,38	−0,7	1,97	*	224,51
Vocalizes	Stereotypes_MS	i	3,74	1,52	0,38	4,04	**	22,16
	Solitary play SP	i	2,04	1,16	0,49	2,34	*	29,63
	Joint attention_WA	IV	2,92	−1,41	−0,43	3,24	**	334,26
	Gaze_EC	i	2,26	1,14	0,45	2,53	*	26,79
Verbal imitation	Autosensori alidad_AS	IV	2,69	−0,53	−0,19	2,74	**	348,92
	Blank stare_BS	I	1,65	3,75	0,92	4,1	**	66,32
	No active response/no following proposal_NAF	I	4,08	1,62	0,37	4,39	**	21.64
	Proxemic behavior_APRO	I	1,42	1.41	0,71	2	*	44.89
	Physical contact_BPC	I	6,52	1,15	0,17	6,62	**	10
	Joint attention_DRA	IV	2,31	−0,7	−0,29	2,41	*	343,1
Joint attention_WO		I	1,77	2,84	0,85	3,34	**	58,11
	Gaze_EC	I	0,04	2,53	1	2,53	*	89,15
	Imitation_NVI	I	1,62	7,79	0,98	7,95	**	78,27
Nonverbal imitation	Stereotypes MS	1	3,02	3,39	0,75	4,54	**	48,27
	Stereotypes_VIS	1	1,35	4.14	0,95	4,35	**	71,89
	Normal actions_NA	1	0,61	2.16	0,96	2,25	*	74.12
	Proxemic behavior_APRO	n	−0,6	1,97	0,96	2,06	*	107,07
	Proxemic behavior_DPB	i	2,74	1,12	0,38	2,96	**	22,21
Stimulates	Stereotypes VIS	HI	−1.91	−1,94	−0,71	2,73	**	225,46
	Solitary play_SP	i	1,17	3,53	0,95	3,72	**	71,73
	Instrumentalization/instrumentalized demand_I	i	4,34	0,7	0,16	4,39	**	9,22
	Active response/following proposal_AF	IV	2,4	−0,46	−0,19	2,45	*	349,08
	No active response/no following proposal_NAF	IV	1,69	−1,5	−0,66	2,26	*	318,54
	Proxemic behavior DPB	n	−1,05	2,33	0,91	2,56	*	114,22
	Joint attention_WA	i	3	0,7	0,23	3,08	**	13,2
	Gaze_EC	IV	2,05	−0,89	−0,4	2,23	*	336,37
	Facial expression_FEXR	n	−1,2	2,6	0,91	2,86	**	114,78

(Continued)

TABLE 3 (Continued)

Focal behavior (therapist)	Conditioned behavior (child)	Quadrant	Prospective zsum	Retrospective zsum	Ratio	Length	Significance	Angle
FEXT	Stereotypes_VS	IV	2,31	−0,35	−0,15	2,34	*	351,49
	Autosensori alidad_AS	n	−0,45	2,37	0,98	2,41	*	100,72
	Functional intent_FI	i	3,15	3,24	0,72	4,52	**	45,83
	Attentive object AGO	n	−1,12	2,36	0,9	2,61	**	115,36
	Solitary play_SP	i	3,27	0,56	0,17	3,32	**	9,79
	Normal actions_NA	i	1,55	1,44	0,68	2,12	*	42,9
	Joint attention_WO	i	1,91	1,15	0,52	2,23	*	31,08
	Gaze_EC	ni	−1.47	−1,5	−0,71	2,1	*	225,48
GA	Autosensorialidad AS	i	2,18	1,81	0,64	2,83	**	39,65
	Solitary play_SP	in	−1,57	−1,51	−0,69	2,18	*	223,9
	Instrumentalization/instrumentalized Demand_I	HI	−2,59	−0,49	−0,18	2,63	**	190,65
	Proxemic behavior APRO	i	2,72	0,23	0,08	2,72	**	4,77
	Proxemic behavior_DPB	i	2,25	1,7	0,6	2,82	**	37,18
	Joint attention_DRA	i	2,53	1.11	0,4	2,76	**	23,65
	Gaze_EC	i	1,68	2,12	0,78	2,71	**	51,68

Only the conditioned behaviors that generate significant (*) and very significant (**) vectors have been selected.

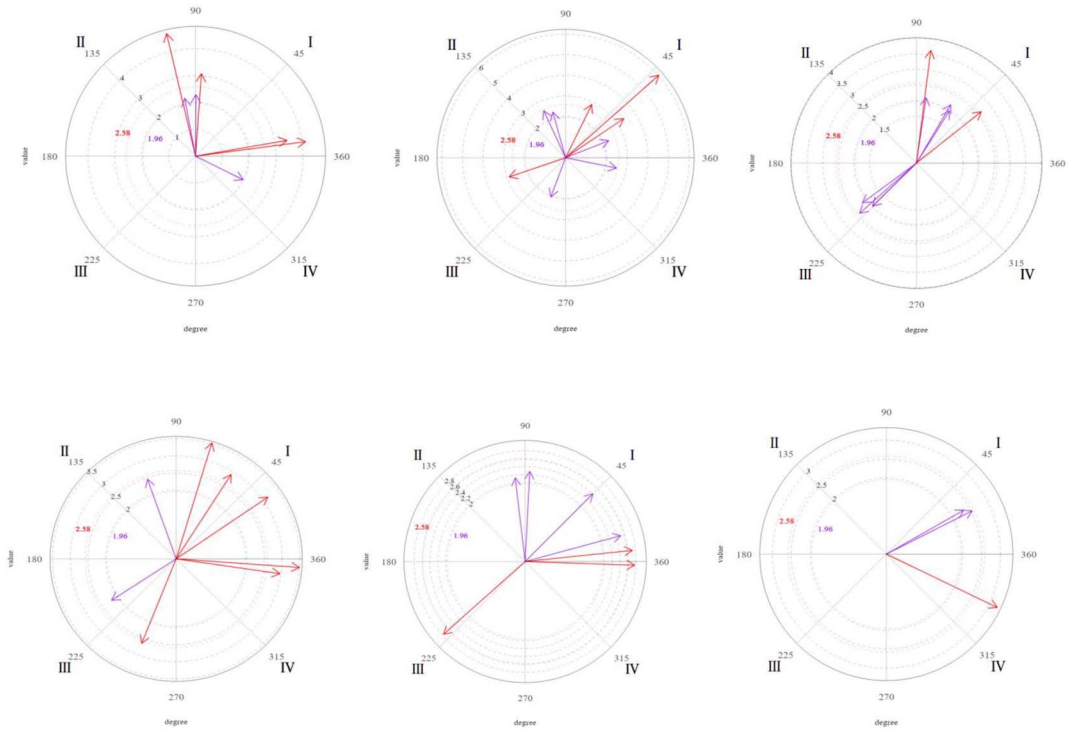


FIGURE 1
Graphic representation of the significant (purple) and very significant (red) vectors obtained in the polar coordinate analysis. Focal behavior "Verbalizes" as the center of the analysis and all the others as conditional of sessions 9, 16, and 20 (from left to right). Focal behavior "Vocalizes" as the center of the analysis and all the others as conditional of sessions 9, 16, and 20 (from left to right).

In the analysis of the child's clinically favorable behaviors, the therapist's verbalization was related with 78.9% of clinically favorable behaviors ($\chi^2 = 6.53$; df = 1; $p = 0.01$) and vocalization with 76.5% ($\chi^2 = 5.87$; df = 1; $p = 0.02$). Of the therapist's imitation behaviors, verbal imitation was related with 53.3% ($\chi^2 = 0$; df = 1; $p = 1$) and non-verbal also with 53.3% ($\chi^2 = 0.13$;

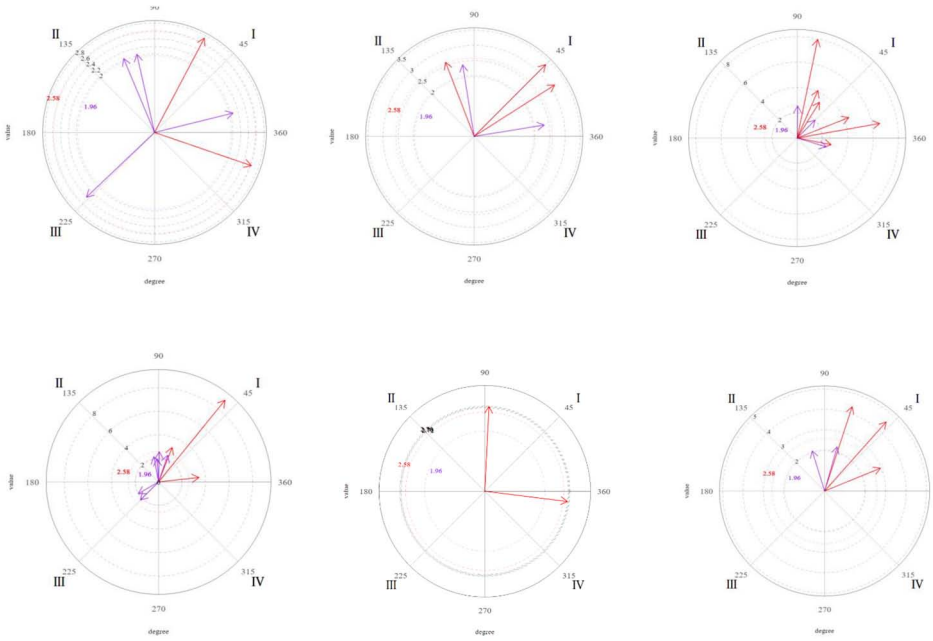


FIGURE 2
Graphic representation of the significant (purple) and very significant (red) vectors obtained in the polar coordinate analysis. Focal behavior "Verbal imitation" as the center of the analysis and all the others as conditional of sessions 9, 16, and 20 (from left to right). Focal behavior "Non-verbal imitation" as the center of the analysis and all the others as conditional of sessions 9, 16, and 20 (from left to right).

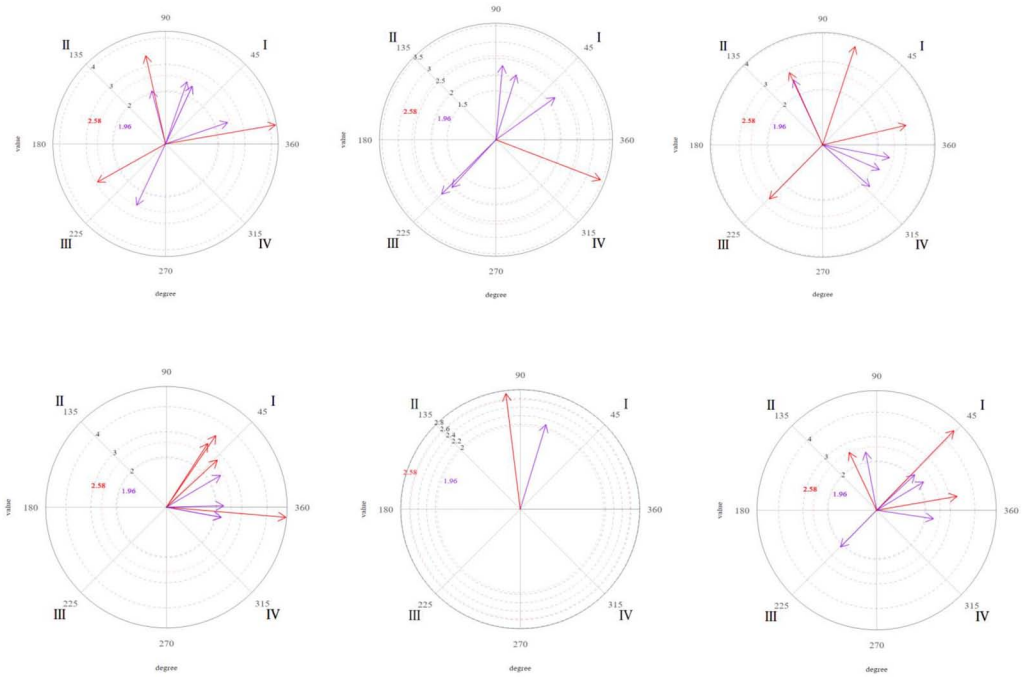


FIGURE 3
Graphic representation of the significant (purple) and very significant (red) vectors obtained in the polar coordinate analysis. Focal behavior "Stimulates" as the center of the analysis and all the others as conditional of sessions 9, 16, and 20 (from left to right). Focal behavior "Facial expression" as the center of the analysis and all the others as conditional of sessions 9, 16, and 20 (from left to right).

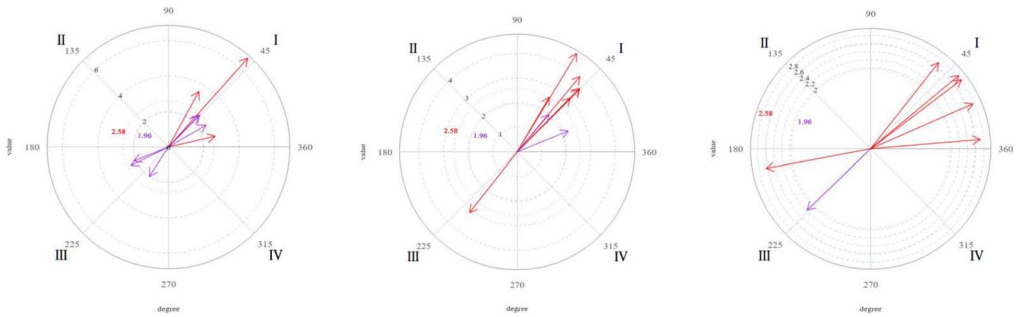


FIGURE 4
Graphic representation of the significant (purple) and very significant (red) vectors obtained in the polar coordinate analysis. Focal behavior "Gaze" as the center of the analysis and all the others as conditional of sessions 9, 16, and 20 (from left to right).

df = 1; $p = 0.71$). Similarly, stimulation with 54.5% ($\chi^2 = 0.10$; df = 1; $p = 0.74$) and facial expression with 71.4% ($\chi^2 = 1.59$; df = 1; $p = 0.20$), while gaze produced 38.1% of clinically unfavorable behaviors ($\chi^2 = 0.5$; df = 1; $p = 0.47$).

In this single case study, we have applied an approach that allows observational methodology and mixed methods as the main analysis (Anguera et al., 2017).

Systematic and meta-analysis reviews show that although child psychoanalytic psychotherapy evidence is increasing, high quality research is needed in order to better understand the effectiveness of such interventions (Midgley and Kennedy, 2011; Abbass et al., 2013; Midgley et al., 2017; Target, 2018). However, there are currently very few centers that systematically evaluate the results of psychoanalytic interventions (Arias-Pujol, 2020). Researchers and therapists, aware of the complexity of each person and the singularity of each psychopathological expression, continue to search for what works and for whom (Fonagy et al., 2016). Recent studies suggest a need to carry out process research in order to identify predictive interaction structures in child psychotherapy (Halfon et al., 2020). At present, different methods and methodologies can be applied, with mixed method methodology representing a new and potentially useful approach for evaluating low level psychoanalytic interventions (Arias-Pujol and Anguera, 2020b).

In the present study, by using a mixed method, we aimed to identify which of the therapist's techniques elucidated clinically favorable behaviors in a child with severe autism, specifically those that promoted RSI actions and inhibited non-RSI actions. From all of them, verbalization and vocalization by the therapist produced significant clinically favorable behaviors, whereas direct gaze promoted the child's withdrawal.

Specifically, "Verbalization" (formed by VDE, VOH, VAN, VREM, VSUP, INT) consists of describing a behavior or an object, offering help, anticipating actions happening in the near future, remembering aloud, encouraging or repeating what the child says in a similar way to promote certain dialogue. It is used to bring the child closer to the process of symbolization, following Coromines' aforementioned psycho-pedagogic technique (Coromines, 1991; Viloca, 1998, 2003; Farrés et al., 2020).

Results show that verbalization activates different behaviors of the RSI dimension related to producing a vocal demand (VOCD), the response to a demand (AF, NAF), proxemic behavior (APRO), eye contact (EC), physical contact (BPC); and in joint attention and also protoconversation behaviors (WA, WO). On the other hand, it inhibits stereotypes (MS, VS, EM, EV, MSI), erratic behavior (EB) and solitary play.

On balance, verbalization is shown to produce significant clinically favorable behaviors in the child, suggesting that this technique is appropriate for activating resources of the child's RSI dimension. It is of key relevance that verbalization prospectively activates the categories "word" (WO) and "word approximation" (WA), as the child in the study is non-verbal. Results suggest that verbalization promotes language and communication, which are also important in developing the symbolization process.

"Vocalization" is the technique used when the therapist seeks interaction by using exclamations, singing, laughing or encouraging the child to express himself/herself vocally. It has been suggested as especially useful for children with verbal communication difficulties. Results show that use of the "Vocalization" (consisting of EE, SI, L, PF) activates behaviors of vocal and non-vocal demand (I, VOCD); and response to the demand (AF), proxemic behaviors (APRO, DPB), eye contact (EC) and word approximation (WA). It also inhibits the attentive gaze at the object (AGO), the verbal stereotype (VS) and solitary play (SP). Therefore, according to our results, "Vocalization" is an appropriate technique for activating interaction.

Imitation involves mirroring the child's actions. It has been described as a technique that allows the child to see "outside" in a specific way what should be mentally represented "inside" (Farrés et al., 2020). In this type of intervention, verbal and non-verbal imitation is not used as a way of modeling the child's behavior but as a way of making contact with the child. Within the framework of mirror neurons theory, the results of

10.3389/fpsyg.2022.998658

TABLE 4 Type and percentage of behaviors that prospectively activated or inhibited each of the therapist's dimensions.

Behavior criterion (therapist)	Conditioned behavior (child)			
	Activates (I, IV)		Inhibits (II, III)	
	RSI	N_RSI	RSI	N_RSI
Verbalizes				
s.9	NFI, WO	AGO	NVD	EB
s.16	APRO, EC, WA	BS		MS, VIS, SP
s.20	VOCD, FI, NFI, BPC	SP		VS, EB
n (%)	9 (47.3%)	3 (15.8%)	1 (5.3%)	6 (31.6%)
Vocalizes				
s.9	I, NVD, FI, DPB	EB		AGO, VS, SP
s.16	FI, NFI, EC	BS		AGO
s.20	EC, WA	MS, SP		
n (%)	9 (52.9%)	4 (23.5%)	0 (0%)	4 (23.5%)
Verbal imitation	RSI	N_RSI	RSI	N_RSI
s.9	NVD, WA	VS	FEXR	AS, SP
s.16	I	SP	WA, WO	
s.20	NFI, APRO, BPC	BS, AS		
n (%)	6 (40%)	4 (26.7%)	3 (20.0%)	2 (13.3%)
Non-verbal imitation	RSI	N_RSI	RSI	N_RSI
s.9	NFI, DRA, EC	VS, AS	WO	EB, AGO, SP
s.16	NFI	SP		
s.20	DPB	MS, VIS	APRO	
n (%)	5 (33.3%)	5 (33.3%)	2 (13.3%)	3 (20.0%)
Stimulates	RSI	N_RSI	RSI	N_RSI
s.9	VOCD, FI, DPB, INV	MS, AGO	WA, I	VS, SP
s.16	FI	VIS, AS, SP	APRO	MS
s.20	I, FI, NFI	SP	DPB	VIS
n (%)	8 (36.6%)	6 (27.3%)	4 (18.2%)	4 (18.2%)
Facial Exp T	RSI	N_RSI	RSI	N_RSI
s.9	I, FI, APRO, DPB, VOCD, WO, INV	MS		
s.16	FI		WO	
s.20		SP, VS		AS, AGO
	8 (57.1%)	3 (21.4%)	1 (7.1%)	2 (14.3%)
Gaze T	RSI	N_RSI	RSI	N_RSI
s.9	WA, EXFI	VS, AS, BS,SP	NFI, DRA	AGO
s.16	APRO	MS, VS, AS,AGO, SP		BS
s.20	APRO, DPB	AS	I	SP
n (%)	5 (23.8%)	10 (47.6%)	3 (14.3%)	3 (14.3%)

RSI, reciprocal social interaction; N_RSI, non-reciprocal social interaction; n, number of behaviors.

a previous study by our group suggest that the systematic use of significant imitation in psychoanalytic psychotherapy with autistic children improves their RSI capacity (Arias-Pujol et al., 2015b).

The present results show that verbal imitation, consisting of VIT, activates the actions of instrumentalization (I), non-verbal demand (NVD), word approximation (WA) and lack of response to the demand (NAF), approximation (APRO) and brief physical contact (BPC). It also inhibits sensory action (SA) and solitary play (SP).

Non-verbal imitation, consisting of NVIT, activates actions of no response to the demand (NAF), drawing (DRA), eye contact (EC) and distancing behavior (DPB). There were some moments when the child appeared to enjoy a kind of dance with music with the therapist. For example, while the child was moving drawing or repeating sounds (ti-ti-ti-ti) he/she gazed with curiosity at how the therapist repeated it. Moreover, non-verbal imitation prospectively inhibited erratic behavior (EB), an attentive gaze at the object (AGO) and the action of solitary play (SP).

However, the behaviors were not statistically significant in the analysis of clinically favorable responses in the child. We believe that these results differ from those of the previous study because on that occasion the use of verbal and non-verbal imitation by the therapist was applied systematically and not depending on the child's spontaneous behavior.

Stimulation (consisting of SDA, SSH, SDB, SPRO) is used when the child is disconnected from the relationship and the therapist tries to seek his/her attention by giving or showing an object, directing the child's attention toward something or toward him/herself, or proposing an activity. Results show that "stimulation" activates instrumentalized and vocal demand (I, VOCD), response to the demand (AF, NAF), distancing proxemic behavior (DPB) and the child's non-verbal imitation (NVI). It also inhibits the attentive gaze at the object (AGO), the blank stare (BS) and solitary play (SP). However, stimulation did not produce statistically significant clinically favorable behaviors. These results suggest that stimulating the child when he/she is very disconnected from the relationship can be more intrusive than verbalizations or vocalizations, and does not always promote reciprocal social interaction.

Dramatization of an emotion consisting of FEXT means making faces or moving the arms and torso, expressing oneself through the body (Viloca, 2003; Farrés, 2014). Results show that it activates behaviors of the RSI dimension. Specifically, in sessions 9 and 16 it activates the instrumentalized and vocal demand (I, VOCD), response to the demand (AF), approximation and distance proxemic behavior (APRO, DPB), the use of words (WO) and the child's non-verbal imitation (NVI); and prospectively inhibits, in session 20, sensory action (SA) and the attentive gaze at the object (AGO). However, the results did not achieve significance as clinically favorable behaviors.

Finally, direct gaze (consisting of GA) refers to when the therapist is observing the child, allowing him/her to express himself/herself freely. Despite the child's tendency to disconnect and avoid closeness and physical contact with the therapist, the GA activates the child's behavior of word approximation (WA), facial expression (FAEX) and of moving close to or away from the therapist (APRO, DPB). Likewise, it inhibits the attentive gaze at the object (AGO), the blank stare (BS) and solitary play (SP). Overall, direct gaze promotes significant withdrawal behaviors, which are not clinically favorable. This result is in line with previous studies suggesting the extreme sensitivity shown by children with autism to small breaks in the therapeutic partnership (Goodman et al., 2017).

Conclusion

This study shows how observational methodology, and specifically the use of mixed methods, can be useful for the evaluation of a low intensity intervention program. The greatest advantage is that the mixed method perspective allows us to capture the reality just as it happens, systemize it, guarantee its quality and treat it quantitatively in a rigorous way.

From a clinical perspective, our results provide objective evidence – backed up by the application of polar-coordinate-based data analysis – that within a framework of psychoanalytic psychotherapy of a child with severe ASD and no language, the therapeutic techniques of "verbalization" and "vocalization" significantly activate reciprocal social interaction behaviors and inhibit non-social reciprocal behaviors. On the other hand, direct gaze promotes child withdrawal. The results are of key importance as they show the therapist behaviors most useful for promoting social interaction in a child with severe autism.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

EA-P and MTA contributed to conception and design of the study. MTA conducted the method section and polar coordinate analysis. JM performed the descriptive analysis. EA-P, MM, and JM performed the interpretation of the data. NB adapted and validated the observational instrument. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

MTA gratefully acknowledges the support of a Spanish government subproject *Integration ways between qualitative and quantitative data, multiple case development, and synthesis review as main axis for an innovative future in physical activity and sports research* (PGC2018-098742-B-C31) (2019-2021) (Ministerio de Ciencia, Innovación y Universidades/AgenciaEstatal de Investigación/Fondo Europeo de Desarrollo Regional), that is part of the coordinated project *New approach of research in physical activity and sport from mixed methods perspective* (NARPAS_MM) (SPGC201800 × 098742CV0). In addition, MTA thanks

for the support of the Generalitat de Catalunya Research Group, GRUP DE RECERCA I INNOVACIÓ EN DISSENYS (GRID). *Tecnología i aplicació multimedia i digital als dissenys observacionals* (Grant number: 2017 SGR 1405). Finally, we also acknowledge support from Ramon Llull University (PGRiD of FPCEE Blanquerna).

Acknowledgments

We thank our colleagues in Carrilet for their help and suggestions, especially Elena Fieschi, Cristina Castelló, Anna Soldevila, and Elisabet Sánchez-Caroz, for their commitment and generosity. We also thank the child and his/her family, without whom this study would not have been possible.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abbass, A., Rabung, S., Leichsenring, F., Refseth, J., and Midgley, N. (2013). Psychodinamic Psychotherapy for Children and Adolescents. A Meta-Analysis of Short-Term Psychodynamic Models. *J. Am. Acad. Child Adolesc. Psychiatry* 52, 863–875. doi: 10.1016/j.jaac.2013.05.014

Alcàcer, B., and Viloca, Ll. (2014). La psicoterapia psicoanalítica en personas con Trastorno Autista. Revisión histórica [Psychoanalytic psychotherapy in people with Autistic Disorder]. *Temas de psicoanál.* 7, 1–29.

Alcover, C., Mairena, M. A., Mezzatesta, M., Elías, N., Díez, M., Balañá, G., et al. (2019). Mixed methods approach to describe social interaction during a group intervention for adolescents with Autism Spectrum Disorders. *Front. Psychol.* 10:1158. doi: 10.3389/fpsyg.2019.01158

American Psychiatric Association (2015). *Manual diagnóstico y estadístico de los trastornos mentales [Diagnostic and statistical manual of mental disorders dsm-5]*, 5th edn. Washington DC: Editorial Médica Panamericana.

Anguera, M. T. (1979). Observational Typology. *Qual. Quant.* 13, 449–484. doi: 10.1007/BF00222999

Anguera, M. T. (1997). "From prospective patterns in behavior to joint analysis with a retrospective perspective," in *Colloque sur Invitation Méthodologied'Analyse des Interactions Sociales*. (Paris: Université de la Sorbonne).

Anguera, M. T. (2003). "Observational Methods (General)," in *Encyclopedia of Psychological Assessment*, ed. R. Fernández-Ballesteros (Thousand Oaks: Sage), 632–637.

Anguera, M. T. (2008). "Diseños evaluativos de baja intervención [Low intervention evaluation designs]," in *Evaluación de Programas Sanitarios y Sociales. Abordaje Metodológico*, eds M. T. Anguera, S. Chacón, and A. Blanco-Villaseñor (Madrid: Síntesis), 153–184.

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

Anguera, M. T. (2018). "Del caso único al caso múltiple en el estudio del comportamiento humano [From the single case to the multiple case in the study of human behavior]," in *Academia de Psicología de España, Psicología para un Mundo Sostenible*. (Zenda Rd: Pirámide), 31–50.

Anguera, M. T. (2020). "Is It Possible to Perform "Liquefying" Actions in Conversational Analysis? The Detection of Structures in Indirect Observation," in *The Temporal Structure of Multimodal Communication. Intelligent Systems Reference Library*, eds L. Hunyadi and I. Szekrényes (Berlin: Springer), 45–67. doi: 10.1007/978-3-030-22895-8_3

Anguera, M. T. (2021). "Desarrollando la observación indirecta: Alcance, proceso, y habilidades metodológicas en el análisis de textos [Developing indirect observation: Reach, process and methodological abilities in the analysis of texts]," in *Patrones de Habilidades Metodológicas y Conceptuales de Análisis, Planeación, Evaluación e Intervención en Ciencias de la Conducta*, eds C. Santoyo and Y. L. Colmenares (Ciudad de México: UNAM), 189–217.

Anguera, M. T. (2022). *Profundizando en el análisis en mixed methods: Integración de elementos cualitativos y cuantitativos en elmarco de la observación sistemática del comportamiento [Going deeper into analysis in mixed methods: Integration of qualitative and quantitative elements within the framework of the systematic observation of behavior]*. Doctoral thesis. Tenerife: University of La Laguna.

Anguera, M. T., and Chacón-Moscoso, S. (2008). "Aproximación conceptual en evaluación de programas [A conceptual approach to program evaluation]," in *Evaluación de Programas Sanitarios y Sociales. Abordaje Metodológico[An Evaluation of Health and Social Programs]*, eds M. T. Anguera, S. Chacón, and A. Blanco-Villaseñor (Madrid: Síntesis), 17–36.

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 sports sciences from a mixed method approach]. *Cuad. Psicol. Deporte* 16, 17–30. doi: 10.4321/S1578-84232015000100002

Anguera, M. T., Blanco-Villaseñor, A., Losada, J. L., Sánchez-Algarra, P., and Onwuegbuzie, A. J. (2018a). Revisiting the Difference Between Mixed Methods and Multimethods: Is It All in the Name? *Qual. Quan.* 52, 2757–2770. doi: 10.1007/s11135-018-0700-2

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

Anguera, M. T., Arias-Pujol, E., Molinero, F., and Del Giacco, L. (in press). "Examining the influences of Spanish research culture in systematic observation with mixed methods," in *The SAGE Handbook of Mixed Methods Research Designs*, ed. C. Poth (Thousand Oaks: Sage).

Anguera, M. T., Blanco-Villaseñor, A., and Losada, J. L. (2001). Diseños Observacionales, cuestión clave en el proceso de la metodología observacional [Observational designs, a key question in the observational methodology process]. *Metodol. Cienc. Comport.* 3, 135–161.

Anguera, M. T., Blanco-Villaseñor, A., Losada, J. L., and Sánchez-Algarra, P. (2020). Integración de elementos cualitativos y cuantitativos en metodología observacional [Integration of qualitative and quantitative elements in observational methodology]. Ámbitos. *Rev. Int. Comun.* 49, 49–70. doi: 10.12795/Ambitos.2020.i49.04

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

Arias-Pujol, E. (2020). "Psicoterapia psicoanalítica de niños, adolescentes y familias: la investigación que une teoría y práctica [Psychoanalytic psychotherapy in children, adolescents and families: research that combines theory and practice]," in *Psicoterapia Psicoanalítica*, eds J. A. Castillo Garayoa and J. Mercadal Rotger (Barcelona: Herder), 131–152. doi: 10.2307/j.ctv16zjhhm.12

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

Arias-Pujol, E., and Anguera, M. T. (2020a). A Mixed Methods Framework for Psychoanalytic Group Therapy: From Qualitative Records to a Quantitative Approach Using T-Pattern, Lag Sequential, and Polar Coordinate Analysis. *Front. Psychol.* 11:1922. doi: 10.3389/fpsyg.2020.01922

Arias-Pujol, E., and Anguera, M. T. (2020b). "Investigación en psicoterapia psicoanalítica: metodología observacional desde la perspectiva mixedmethods [Research into psychoanalytic psychotherapy: observation methodology from a mixed method perspective]," in *Psicoterapia Psicoanalítica [Psychoanalytic Psychotherapy]*, eds J. A. Castillo Garayoa and J. Mercadal Rotger (Barcelona: Herder), 225–249. doi: 10.2307/j.ctv16zjhhm.17

Arias-Pujol, E., Fieschi, E., and Mestres, M. (2015a). La imitació del nen autista en la psicoteràpia psicoanalítica: Fonaments, disseny i aplicació [The imitation of the autistic child in psychoanalytic psychotherapy: Fonaments, design and application]. *Rev. Catal. Psicoanàl.* 32, 123–137.

Arias-Pujol, E., Fieschi, E., Miralbell, J., Castellò, C., Soldevila, A., Anguera, M. T., et al. (2015b). Efectos de la imitación en la interacción social recíproca en un niño con Trastorno del Espectro Autista Grave [Effects of imitation in reciprocal social interaction in a child with Severe Autism Spectrum Disorder]. *Rev. Psicopatol. Salud Mental Niño Adolesc.* 25, 9–20.

Bachs, N. (2019). *Autisme infantil: Adaptació i validació d'un instrument observacional [Child autism: Adaptation and validation of an observational instrument]*. Unpublished Ph.D. thesis. Barcelona: Universitat Ramon Llull.

Blanco-Villaseñor, A., and Anguera, M. T. (2000). "Evaluación de la calidad en el registro del comportamiento: Aplicación a deportes de equipo [Evaluation of quality in the recording of behavior: Application to team sports]," in *Métodos Numéricos en Ciencias Sociales*, eds E. Oñate, F. García-Sicilia, and L. Ramallo (Catalonia: CIMNE), 30–48.

Bryant, D. M., and Bickman, L. (1996). Methodology for evaluating mental health case management. *Eval. Program Plan.* 19, 121–129. doi: 10.1016/0149-7189(96)00003-1

Chacón, S., Anguera, M. T., and López Ruiz, J. (2000). Diseños de evaluación de programas: Bases metodológicas [Program evaluation designs: Methodological bases]. *Psicothema* 12, 122–126.

Chacón-Moscoso, A., Anguera, M. T., Sanduvete-Chaves, S., Losada, J. L., Lozano-Lozano, J. A., and Portell, M. (2019). Methodological quality checklist for studies based on observational methodology (MQCOM). *Psicothema* 31, 458–464.

Chacón-Moscoso, S., Anguera, M. T., Pérez-Gil, J. A., and Holgado-Tello, F. P. (2002). A Mutual Catalytic Model of Formative Evaluation: The Interdependent Roles of Evaluators and Local Programme Practicioners. *Eval. Int. J. Theory Res. Pract.* 8, 413–432. doi: 10.1177/135638902606 20612

Chacón-Moscoso, S., Anguera, M. T., Sanduvete-Chaves, S., and Sánchez-Martín, M. (2014). Methodological convergence of program evaluation designs. *Psicothema* 26, 91–96.

Chacón-Moscoso, S., Sanduvete-Chaves, S., Lozano-Lozano, J. A., Portell, M., and Anguera, M. T. (2021). From randomized control trial to mixed methods: A practical framework for program evaluation based on methodological quality / Del ensayo controlado aleatorizado a los métodos mixtos: Un marco práctico para la evaluación de programas basado en la calidad metodológica. *An. Psicol. Ann. Psychol.* 37, 599–608. doi: 10.6018/analesps. 470021

Chacón-Moscoso, S., Sanduvete-Chaves, S., Portell, M., and Anguera, M. T. (2013). Reporting a program evaluation: Needs, program plan, intervention, and decisions. *Int. J. Clin. Health Psychol.* 13, 58–60. doi: 10.1016/S1697-2600(13) 70008-5

Cochran, W. G. (1954). Some methods for strengthening the common $\chi 2$ tests. *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

Coromines, J. (1991). Psicopatologia i Desenvolupament Arcaics. Assaig Psicoanalític. Espaxs. Reissued in Spanish (1998), Psicopatología Arcaica y Desarrollo: Ensayo Psicoanalítico [Archaic Psychopathology and Development: a Psychoanalytical Essay]. Paidós Ibérica. Barcelona: Espaxs.

Creswell, J. W., and Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research*, 3rd Edn. Thousand Oaks: Sage.

Del Giacco, L., Anguera, M. T., and Salcuni, S. (2020). The Action of Verbal and Non-verbal Communication in the Therapeutic Alliance Construction: A Mixed Methods Approach to Assess the Initial Interactions With Depressed Patients. *Front. Psychol.* 11:234. doi: 10.3389/fpsyg.2020.00234

Dvoskin, J. A., and Steadman, H. J. (1994). Using case management to reduce violence by mentally ill persons in the community. *Hosp. Community Psychiaty* 45, 679–684. doi: 10.1176/ps.45.7.679

Edwards, D. J. A., Dattilio, F. M., and Bromley, D. B. (2004). Developing evidence-based practice: The role of case-based research. *Prof. Psychol.* 35, 589–597. doi: 10.1037/0735-7028.35.6.589

Farrés, N. (2014). Confluència de les vessants educative I terapéutica per al treball d'infants amb TEA: De l'emoció a la cognició [Confluence of educational and therapeutic dimensions in working with children with ASD: from emotion to cognition]. *Rev. Desenvol.* 38, 1–15.

Farrés, N., Cornadó, M., Arias-Pujol, E., Alejos, C., Viloca, L., Blanch, J. M., et al. (2020). Júlia Coromines Award 2019-2020 Psicoteràpia de grup amb infants amb TEA: Proposta d'ampliació de l'esquema psicopedagògic de Júlia Coromines [Group psychotheray with children with ASD: A proposal for the widening of Júlia Coromines' psychopedagogic framework]. *Rev. Catal. Psicoanàl.* 37, 143–161.

Flinn, L., Braham, L., and Das Nair, R. (2015). How reliable are case formulations? A systematic literature review. *Br. J. Clin. Psychol.* 54, 266–290. doi: 10.1111/bjc.12073

Fonagy, P., Cottrell, D., Phillips, J., Bevington, D., Glaser, D., and Allison, E. (2016). *What Works for Whom? A Critical Review of Treatments for Children and Adolescents. (2a ed)*. New York, NY: Guilford Press.

Gerring, J. (2004). What is a case study and what is it good for? *Am. Polit. Sci. Rev.* 98, 341–354. doi: 10.1017/S0003055404001182

Goodman, G., Chung, H., Fischel, L., and Athey-Lloyd, L. (2017). Simulation modeling analysis of sequential relations among therapeutic alliance, symptoms, and adherence to child-centered play therapy between a child with autism spectrum disorder and two therapists. *Clin. Child Psychol. Psychiatry* 22, 455–466. doi: 10.1177/1359104517691082

Halfon, S., Goodman, G., and Bulut, P. (2020). Interaction structures as predictors of outcome in a naturalistic study of psychodynamic child psychotherapy. *Psychother. Res.* 30, 251–266. doi: 10.1080/10503307.2018.1519267

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

Hilliard, R. B. (1993). Single-case methodology in psychotherapy process and outcome research. *J. Consult. Clin. Psychol.* 61, 373–380. doi: 10.1037/0022-006X. 61.3.373

Hutt, S. J., and Hutt, C. (1974). *Direct Observation and Measurement of Behavior*. Illinois: Charles C. Thomas.

Izquierdo, C., and Anguera, M. T. (2021). The analysis of interpersonal communication in sport from mixed methods strategy: The integration of qualitative-quantitative elements using systematic observation. *Front. Psychol.* 12:637304. doi: 10.3389/fpsyg.2021.637304

Kent, R. (2009). "Case-centered methods and quantitative analysis," in *The Sage Handbook of Case-Based Methods*, ed. D. Byrne (Thousand Oaks: Sage), 184–206. doi: 10.4135/9781446249413.n11

Kuyken, W., Fothergill, C. D., Musa, M., and Chadwick, P. (2005). The reliability and quality of cognitive case formulation. *Behav. Res. Therapy* 43, 1187–1201. doi: 10.1016/j.brat.2004.08.007

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

Lapresa, D., Gutiérrez, I., Pérez-de-Albéniz, A., Merino, P., and Anguera, M. T. (2020). Teacher-student-task-interactions in a motor skills programme for an adolescent boy with autism spectrum disorder: A systematic observation study / Interacción profesor-alumno-tarea en un programa de desarrollo de capacidades motrices en un adolescente con TEA: Un estudio de observación sistemática). *J. Study Educ. Dev. Infancia Aprendiz.* 44, 553–585. doi: 10.1080/02103702.2020. 1802148

Lord, C., Rutter, M., Dilavore, P. C., and Risi, S. (2000). *Autism Diagnostic Observation Schedule (ADOS)*. Bilbao: TEA Editions. doi: 10.1037/t17256-000

Martin, P., and Bateson, P. (1991). *Measuring Behavior: An Introductory Guide*. Cambridge: Cambridge University Press.

Midgley, N., and Kennedy, E. (2011). Psychodynamic psychotherapy for children and adolescents: A critical review of the evidence base. *J. Child Psychother.* 37, 232–260. doi: 10.1080/0075417X.2011.614738

Midgley, N., O'Keeffe, S., French, L., and Kennedy, E. (2017). Psychodynamic psychotherapy for children and adolescents: An updated narrative review of the evidence base. *J. Child Psychother.* 43, 307–329. doi: 10.1080/0075417X.2017.1323945

Mucchielli, R. (1974). *L'observation Psychologique et Psychosociologique*. Strasbourg: ESF.

Newman, I., and Benz, C. R. (1998). *Qualitative-Quantitative Research Methodology: Exploring the Interactive Continuum*. Carbondale: Southern Illinois University Press.

Portell, M., Anguera, M. T., Chacón-Moscoso, S., and Sanduvete-Chaves, S. (2015). Guidelines for Reporting Evaluations based on Observational Methodology (GREOM). *Psicothema* 27, 283–289.

Rodríguez-Medina, J., Arias, V., Arias, B., Hernández-Mendo, A., and Anguera, M. T. (2022). *Polar Coordinate Analysis, from HOISAN to R: A Tutorial Paper*. Available Online at: https://jairodmed.shinyapps.io/HOISAN_to_R/. (accessed August 2, 2022).

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

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.Quan. Int. J. Methodol.* 47, 1237–1257. doi: 10.1007/s11135-012-9764-6

Sandelowski, M. (1996). One is the liviest number: The case orientation of qualitative research. *Res. Nurs. Health* 19, 525–529. doi: 10.1002/(SICI)1098-240X(199612)19:6<525::AID-NUR8>3.0.CO;2-Q

Stake, R. E. (1994). "Case studies," in *Handbook of Qualitative Research*, eds N. K. Denzin and Y. S. Lincoln (Thousand Oaks: Sage), 236–247.

Stake, R. E. (1995). *The Art of Case Study Research*. Thousand Oaks: Sage.

Stake, R. E. (2005). *Multiple Case Study Analysis*. New York, NY: Guilford Press.

Target, M. (2018). 20/20 Hindsight: A 25-year programme at the Anna Freud Centre of efficacy and effectiveness research on child psychoanalytic psychotherapy. *Psychother. Res.* 28, 30–46. doi: 10.1080/10503307.2017.1349351

Thomas, G. (2011). A typology for the case study in social science following a review of definition, discourse, and structure. *Qual. Inq.* 17, 511–521. doi: 10.1177/1077800411409884

Tight, M. (2010). The curious case of case study: A viewpoint. *Int. J. Soc. Res. Methodol.* 13, 329–339. doi: 10.1080/13645570903187181

Trevarthen, C., and Aitken, K. J. (2001). Infant intersubjectivity: Research, theory, and clinical applications. *J. Child Psychol. Psychiatry* 42, 3–48. doi: 10.1111/1469-7610.00701

Viloca, Ll. (1998). Ansietat catastròfica: De la sensorialitat a la comunicació. [Catastrophic anxiety: From sensoriality to communication]. *Rev. Catal. Psicoanàl.* 15, 35–60.

Viloca, L. (2003). *El niño autista: Detección, evolución y tratamiento. Reissued in 2012 by Col·leccions Carrilet [The autistic child: Detection, evolution and treatment]*. Barcelona: CEAC.

Viloca, Ll. (2011). Aportacions punteres de la Dra. Júlia Coromines a la comprensió de l'autisme. [Dr. Júlia Coromines' cutting edge contributions to understanding autism]. *Rev. Catal. Psicoanàl.* 28, 55–69.

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

Weil, M. (1985). "Adapting case management to specific programs," in *Case Management in Human Service Practice*, eds M. Weil and J. Karls (Hoboken: Jossey-Bass).

Yin, R. K. (2014). *Case Study Research. Design and Methods*, 5th Edn. Thousand Oaks: Sage.

frontiers | Frontiers in Psychology

TYPE Systematic Review
PUBLISHED 03 November 2022
DOI 10.3389/fpsyg.2022.971617

OPEN ACCESS

EDITED BY
Salvador Chacón-Moscoso,
Seville University, Spain

REVIEWED BY
Iwan Wopereis,
Open University of the Netherlands, Netherlands
Roman Freunberger,
Institut des Bundes für Qualitätssicherung im österreichischen Schulwesen, Austria

*CORRESPONDENCE
Héctor Tronchoni
hector.tronchoni@gmail.com

SPECIALTY SECTION
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

RECEIVED 17 June 2022
ACCEPTED 21 September 2022
PUBLISHED 03 November 2022

CITATION
Tronchoni H, Izquierdo C and Anguera MT (2022) A systematic review on lecturing in contemporary university teaching.
Front. Psychol. 13:971617.
doi: 10.3389/fpsyg.2022.971617

A systematic review on lecturing in contemporary university teaching

Héctor Tronchoni[1,2]*, Conrad Izquierdo[3] and M. Teresa Anguera[4]

[1]Faculty of Teacher Training, University of Valencia, Valencia, Spain, [2]Faculty of Education, Florida Universitària, Valencia, Spain, [3]Faculty of Psychology, Autonomous University of Barcelona, Barcelona, Spain, [4]Faculty of Psychology, Institute of Neurosciences, University of Barcelona, Barcelona, Spain

Introduction: Articles published in scientific journals, concerning the present and future of the lecture format in university education in the twenty-first century are framed within organizational settings that drive teaching methodologies in line with educational policies. The following two research questions have arisen from articles in which debate the continuity of this teaching modality and propose improvements of a different nature: (1) Is there an interest in renovating the lecture format among the international research community whose remit is university teaching methods? and (2) What improvements to the lecture format do the reviewed articles suggest, within the framework of the communicative matrix of interactive learning?

Method: We have carried out a systematic review guided by the PRISMA approach, emphasizing the interest in methodological conceptual commitment, paying attention to documents published in journals with an impact factor. The search strategy was applied homogeneously in three databases: ERIC, PsycInfo, and Web of Science, following the systematic process of inclusion/exclusion.

Results: Forty-five articles were selected with a range of 0–78 quotations, from different fields of knowledge and five continents; 12 articles are from journals with a JCR impact factor. The journal articles cover communicative (21), cognitive (13) and active-practical perspectives (11); the predominant governing aim of the analyzed improvements is connected with the attendees' academic performance results (24); the reviewed studies belong mainly to the quantitative paradigm (42). The considerations derived from the results (45) cover formative, technical and/or critical aspects.

Discussion and conclusions: Whilst positively valuing all these efforts promoted by the European Higher Education Area, we have also verified the lack of contributions in line with our concerns that embrace the need to

develop an in-depth conceptualization, supported by a methodology that is sensitive to the complexity of the oral communication format between an expert actor and non-specialized actors who wish to connect and collaborate with the expert in the production of knowledge.

KEYWORDS

systematic review, lecture, PRISMA, higher education, university teaching

Introduction

Within the context of the new vision of higher education (UNESCO Declaration, 1998; The Bologna Declaration, 1999) we propose to contribute to the renovation of teaching methodology by systematically reviewing the case of the university lecture format (Tronchoni et al., 2018, 2021; Tronchoni, 2019). We agree with the view that the expository-lecture format based on the programming of subject lessons should be reassessed both from a communicative standpoint and from the angle of the shared production of academic knowledge during university lessons.

The lecture is effectively one of the most used teaching methods in universities (Fortanet-Gómez and Ruíz-Madrid, 2014), and at first glance there does not appear to be an issue between the use of this teaching format and the institutional commitment to the development of democratic values and the promotion of social welfare. In fact, the study of the lecture as an improved expository format in higher education has its own place within the area of Instructional Communication within the field of interpersonal communication (Mazer and Hess, 2017).

In the last two decades there has been a proliferation of publications that deal with the lecture-type expository format (Pérez-Llantada and Ferguson, 2006; Deroey and Taverniers, 2011; O'Callaghan et al., 2017), noting the multiple functions and wide diversity of knowledge areas to which it is applied (Steinert and Snell, 1999; Dolnicar et al., 2009; Stacy, 2009; Tanahoung et al., 2009; Özcan, 2013), whilst highlighting the positive opinion that students have of this teaching format (Bates et al., 2017; Buchanan and Palmer, 2017).

Approval (or disapproval) of this teaching method ranges from emphasizing or questioning its effectiveness in small and large groups (Steinert and Snell, 1999; Kramer, 2017), to appraising the development of students' listening and note-taking skills (Meyer and Hunt, 2017).

With the incorporation of active pedagogies in university teaching, different studies have shown a concern for the role played by the lecture in the students' learning process (Barr and Tagg, 1995; Dannels, 2016; Darling, 2017; Tronchoni et al., 2021). This is giving rise to a change of direction in terms of understanding how active listening can benefit from other self-directed cognitive and emotional processes, whilst not forgetting the interpersonal communication skills that may mobilize the participants (Darling-Hammond et al., 2017; Mallin, 2017; Hayden and Chory, 2018; Stockard et al., 2018; Thwin and Lwin, 2018). Whilst it is assumed that the lecture is a face-to-face format, the incorporation of the Internet into formal teaching has led to the lecture format being increasingly present in different online educational modalities or synchronous hybrid contexts (Raes et al., 2020), with studies appearing concerning the use of interactive webinars (Gegenfurtner and Ebner, 2019) and pre-recorded lecture classes (O'Callaghan et al., 2017). Furthermore, over the last 2 years its synchronous virtual and online use has been propelled by the COVID pandemic (Younis and Elbanna, 2022).

This tendency has led to the publication of studies committed to the transformation of the lecture into what could be called the *new expert lecture*, a subject currently under debate within international higher education forums (French and Kennedy, 2017; Buzzanell, 2017; Darling, 2017; Meyer and Hunt, 2017; Sciullo, 2017; Stearns, 2017; Waldeck and Weimer, 2017; Samarasekera et al., 2018).

The synthetic review narrative that precedes the current situation of the lecture in higher education has led us to pose the central exploratory questions of this synthetic systematic review, reducing the PICO strategy to three elements: population (P), intervention (I), and result (O). Firstly, we seek answers to the following questions: (Research Question 1) Is there an interest in renovating the lecture format among the international research community whose remit is university teaching methods? And secondly (Research Question 2), what improvements to the lecture format do the reviewed articles suggest, within the framework of the communicative matrix of interactive learning (Ruesch and Bateson, 1951)?

Method

This work follows the updated protocol of *The Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA) for the transparent, complete and precise presentation of systematic review reports (Page et al., 2021).

Search strategy

The search was carried out across three databases: PsycInfo, Web of Science, and ERIC, and the search strategy for identifying material was homogeneously applied: title containing the term *lecture*, AND *teaching methods* OR *lecture method* among the key words. This search tool was completed with the filters: articles in journals, university level education, peer review, complete text available, English language and time range from 2012 to 2021.

Eligibility criteria

The eligibility criteria applied were: articles of an empirical nature applied to university level education, with a conventional summary and methodological structure (introduction, method, results, and discussion). The aim was to give priority to studies that deal with specific experiences of changes in the lecture, with a defined organization. Reviews of any kind (narrative, bibliographical or meta-analysis) were discarded on the assumption that the established period of analysis was insufficient to allow for the production of additional elaborative material.

Selection process

The selection process involved firstly a review of the titles and articles, and secondly a detailed review of the complete texts of the remaining articles taking into consideration the eligibility criteria (see Figure 1).

Sample obtained

Table 1 includes all the selected articles (n = 45) in chronological order, with the year of publication, the name of the journal and the title of each article appearing from left to right.

Analytical framework

The constructed analytical framework consists of two dataframes:

(I) Firstly, the scientific visibility and institutional backing of the selected research was coded (see Figure 2). The geographical origin indicator was taken into consideration since it provides information about the existing educational policies

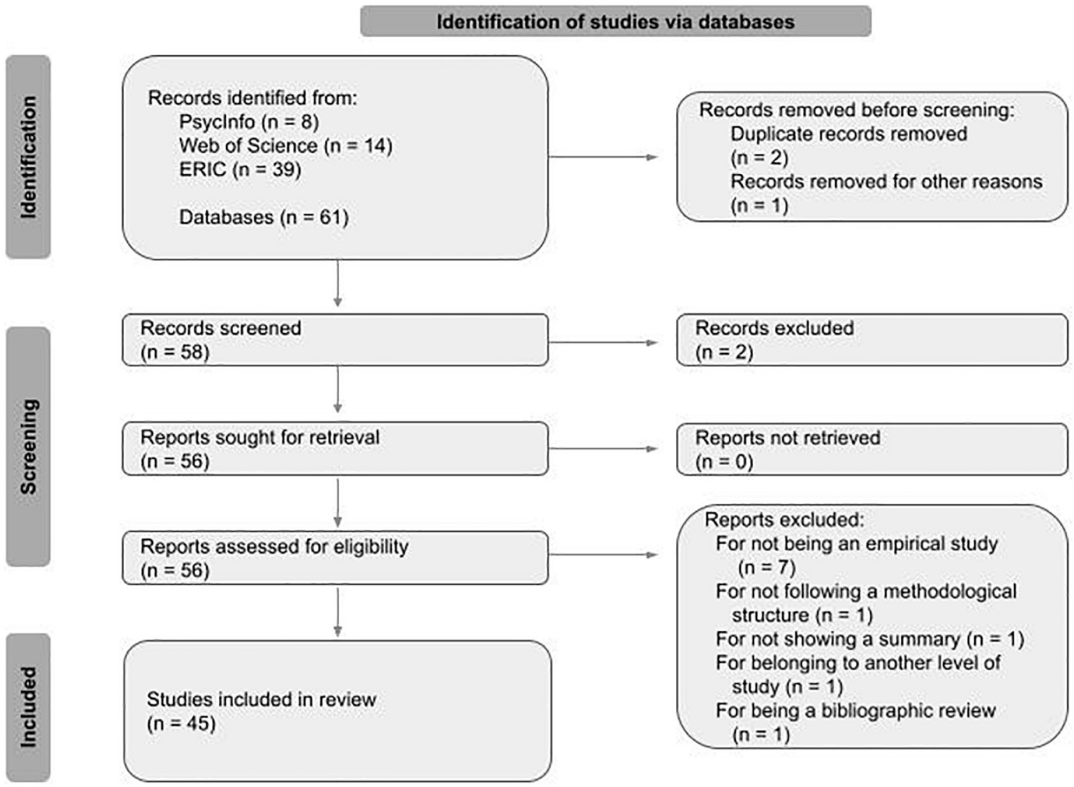


FIGURE 1
Flow diagram (PRISMA 2020) with selection process.

TABLE 1 Selected articles in chronological order.

Year	Journal	Title
2012	Journal of Pragmatics	The interdependence of repetition and relevance in university lectures
2012	The Behavior Analyst Today	The relative effects of traditional lectures and guided notes lectures on university student test scores
2012	The Behavior Analyst Today	The Relation between GPA and Exam Performance during Interteaching and Lecture
2012	Interactive Learning Environments	Explaining students' appraisal of lectures and student-activating teaching: Perceived context and student characteristics
2012	Biochemistry and Molecular Biology Education	Learning transferable skills in large lecture halls: Implementing a POGIL approach in biochemistry
2013	International Journal for the Scholarship of Teaching and Learning	Finding the Right Fit: Assessing the Impact of Traditional v. Large Lecture/Small Lab Course Formats on a General Education Course
2013	Advances in Language and Literary Studies	Developing a Specialized Vocabulary Word List in a Composition Culinary Course through Lecture Notes
2013	International Journal of Higher Education	Features of Application of Classroom Response System at the Lectures in Russia and Israel
2013	Novitas-ROYAL (Research on Youth and Language)	A Study on Perception of Lecturer-Student Interaction in English Medium Science Lectures
2013	Turkish Online Journal of Distance Education	A Comparison of Internet-Based Learning and Traditional Classroom Lecture to Learn CPR for Continuing Medical Education
2013	British Journal of Educational Technology	An augmented lecture feedback system to support learner and teacher communication
2014	Journal of the Scholarship of Teaching and Learning	Teacher Immediacy and Student Learning: An Examination of Lecture/Laboratory and Self-Contained Course Sections
2014	Teaching of Psychology	If you record it, some won't come: Using lecture capture in introductory psychology
2015	Online Learning	Using Instructor-Generated Video Lectures in Online Mathematics Courses Improves Student Learning
2015	Informatics in Education	Using Short Video Lectures to Enhance Mathematics Learning–Experiences on Differential and Integral Calculus Course for Engineering Students
2015	Journal of Evolution of Medical and Dental Sciences-JEMDS	Comparison of the traditional chalk and board lecture system versus power point presentation as a teaching technique for teaching gross anatomy to the first professional medical students
2015	Journal of Interactive Media in Education	Digital Voting Systems and Communication in Classroom Lectures–An Empirical Study Based around Physics Teaching at Bachelor Level at Two Danish Universities
2015	Canadian Journal for the Scholarship of Teaching and Learning	Sustainability: Teaching an Interdisciplinary Threshold Concept through Traditional Lecture and Active Learning
2015	Journal of Evolution of Medical and Dental Sciences-JEMDS	Comparison of Problem Based Learning with Traditional Lectures among First Year Medical Students in Philosophy
2015	The Mathematics Educator	Research on Group Learning and Cognitive Science: A Study of Motivation, Knowledge, and Self-Regulation in a Large Lecture College Algebra Class
2016	International Journal of Higher Education	Integration of Histology Lectures and Practical Teaching in China
2016	Turkish Journal of Emergency Medicine	The comparison of the efficiency of traditional lectures to video-supported lectures within the training of the Emergency Medicine residents
2016	Journal of College Teaching & Learning	Preparing Students for Class: A Clinical Trial Testing the Efficacy between Multimedia Pre-Lectures and Textbooks in an Economics Course
2016	Journal of Curriculum and Teaching	The Use of Pre-Recorded Lectures on Student Performance in Physiology
2017	International Journal of Evaluation and Research in Education	Students' Critical Thinking Improvement through "PDEODE" and "STAD" Combination in the Nutrition and Health Lecture
2017	Bali Medical Journal	Effectiveness of teaching: Jigsaw technique vs. lecture for medical students' Physics course
2017	Advances in Engineering Education	Large Lecture Transformation: Improving Student Engagement and Performance through In-Class Practice in an Electrical Circuits Course
2017	GIST-Education and Learning Research Journal	Questions in English as a Medium of Instruction versus Non-English as a Medium of Instruction Lectures

(Continued)

TABLE 1 (Continued)

Year	Journal	Title
2017	Journal of Education and Practice	The Effect of Instructional Methods (Lecture-Discussion versus Group Discussion) and Teaching Talent on Teacher Trainees Student Learning Outcomes
2018	Journal of Learning in Higher Education	Building a Case for Active Learning: The Use of Lecture vs. Other Classroom Activities at LMBC
2018	Computers & Education	Impact of slide-based lectures on undergraduate students' learning: Mixed effects of accessibility to slides, differences in note-taking, and memory term
2018	Journal of Chiropractic Education	Comparison of student performance and perceptions of a traditional lecture course versus an inverted classroom format for clinical microbiology
2018	International Journal for the Scholarship of Teaching and Learning	Role-Play in Literature Lectures: The Students' Assessment of Their Learning
2018	International Journal of Higher Education	Use of a Scaffolded Case Study Assignment to Enhance Students' Scientific Literacy Skills in Undergraduate Nutritional Science Education: Comparison between Traditional Lecture and Distance Education Course Formats
2019	Journal of Learning Analytics	Diversity of Online Behaviours Associated with Physical Attendance in Lectures
2019	Anatomical Sciences Education	Interactive Lecture in the Dissection Hall: Transforming passive lecture into a dynamic learning experience
2019	English Language Teaching	Micro-Lecture Teaching for Improving the Learning Effect of Non-English Majors at North China Electric Power University
2019	European Journal of Contemporary Education	Three Scientific Facts about Ukrainian and Polish Law-Students: Verification of Statistical Hypotheses about their Preferences of Learning at Lectures
2019	International Review of Research in Open and Distributed Learning	Diversity in Video Lectures: Aid or Hindrance?
2020	Research in Learning Technology	The Effect of Adding Same-Language Subtitles to Recorded Lectures for Non-Native, English Speakers in E-Learning Environments
2020	International Journal of Higher Education	A UTAUT Evaluation of WhatsApp as a Tool for Lecture Delivery during the COVID-19 Lockdown at a Zimbabwean University
2020	Sage Open Nursing	Comparison of the Conceptual Map and Traditional Lecture Methods on Students' Learning Based on the VARK Learning Style Model: A Randomized Controlled Trial
2020	Advances in Medical Education and Practice	Comparison Between Problem-Based Learning and Lecture-Based Learning: Effect on Nursing Students' Immediate Knowledge Retention
2020	Bulletin of the University of Karaganda-Chemistry	Presenting lecture materials in English using CLIL technologies
2020	Journal of E-Learning and Knowledge Society	Does the sequence of flipped and lecture-based classes affect the academic achievement and satisfaction of medical students?

and quality demands in higher education in the universities of the countries of reference. Along with the country of reference, the knowledge area or discipline of the academic subject matter is indicated in those which generated some type of renewal proposal of the lecture format based on empirical evidence. The codes of the knowledge areas/disciplines are: sciences (CEX); biological sciences (BIO), medicine and health sciences (MED), social sciences (SOC), economic and business administration sciences (EAD), humanities (HUM) and diverse or indeterminate (DIV).

A distinction can be made between countries and continents, –and, ultimately, between universities and institutions that recognize the relevance of, or have financed, research in this field. This generates an ordered record of institutional recognition of the origin of said research. This data is completed by two relevance indicators of the knowledge produced and disseminated: we are referring to the number of quotations taken from the article, and the scientific evaluation received by the journal responsible for the publication from some of the most respected platforms concerning the assessment and analysis of performance and scientific research quality. The number of quotations is an indicator of the professional repercussion that the article has had, both in the area of university education and in the pedagogic and didactic research of teaching methods—this was obtained via Google Scholar and identified until the end of 2021; the positioning of the journal includes the impact factor (JCR-WoS) and the quartile (Q) to which it belongs according to the year the selected article was published.

(II) The second dataframe refers to the *multidimensional classification of the structural components* that produce significant differences in the way of conceiving and structuring

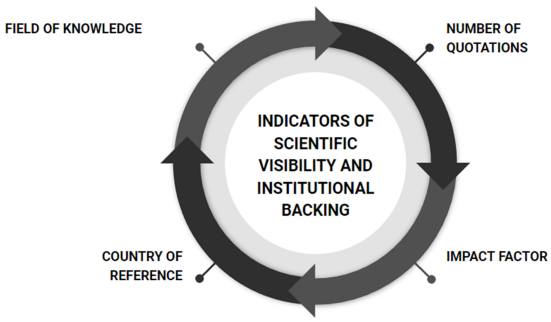


FIGURE 2
Criteria for the analysis of institutional backing and the visibility achieved with the publication of an empirical study aimed at improving the functioning of the university lecture.

the research object. As all the studies that make up the dataframe refer to how to relate and drive teaching and learning in the renovated use of the university lecture, we would like to point out that the practical and technical proposals do not always entail a theoretical justification identifiable as belonging to a recognized and named learning paradigm. Rather what is produced is a free use of concepts and techniques that can respond to different theoretical focuses (Entwistle, 2018). Taking this into account, the distinctive features considered were: conceptual perspective, guiding aim, type of study and applied result.

a) We identified three perspectives of a technical nature applied to the improvement of the lecture format: the communicative perspective (COM), the cognitive perspective (COG), and the participative-practical perspective (ACT).
b) The consolidated guiding aim that routed the selected research was conceived in terms of aptitude-treatment interaction (Cronbach and Snow, 1977) and the criteria derived to characterize the dominant concern were: student potential or aptitude (DIS), the strategies and conditions of the teaching to be carried out (INS), performance (REN) and the combination (COB) of criteria (INS-REN, INS-DIS, DIS-REN).
c) The methodological option that structures the research object on epistemological, ontological, and procedural levels can be specified with the widely argued and accepted proposal of quantitative (QUAN) and qualitative (QUAL) paradigms, and mixed-method (MM).
d) Finally, the applied results or conclusive recommendations can be understood as being aimed at assessing the education fostered by the lecture (FOR), technology for learning and knowledge (TEC), and the need to compare the use of the lecture with other teaching methods (CRI). These criteria can be presented combined in the same article (MIX).

Table 2 contains the symbols assigned to the categories used in the content analysis of the sample obtained via the PRISMA procedure.

Results

Table 3 shows the analysis of the scientific production relevance indicators (Quotations and IF JCR) and the supported relevance in origin (country, university, disciplinary knowledge) of the reviewed empirical articles:

Scientific visibility and institutional backing

Number of quotations: presence/absence of links criterion

The number of quotations ranges from 0, a study by Shabani et al. (2020), to 78 (see **Figure 3**), an article by Bailey et al. (2012) that deals with the transformation of the expository lecture in a large group, within a teaching format focused on the student. As the number of quotations rises, the number of articles diminishes, and therefore the most recently published articles show a lower number of quotations than the articles published in the first years of the time span used in this study (2012–2021). The number of accumulated quotations from all the articles as a whole was 759. It is worth assessing the presence of links among researchers concerned about the same issues rather than the quantity received, given that the materialization of synergies is more sensitive to the conditions in which the quotations are produced rather than the quantity of quotations received.

Impact factor: scientific reliability criterion

Of the 45 articles selected, 12 (27%) belong to journals with an impact factor: 5 of Q1, 3 of Q2, 3 of Q3, and 1 of Q4 (see **Figure 4**). The greatest impact factor is 5.627 and corresponds to an article published in the Q1 journal, *Computers & Education*, about the effects of access to projected slides during lectures using *Powerpoint* (Kim, 2018). The recognition that research groups and communities give to the need for an external assessment of their material before publishing, and the aspiration of being assessed by highly qualified journals, are two points that reflect the concern for obtaining applied results and a good path to achieving tangible applied results based on rigorous studies.

Geographical-academic distribution: institutional backing criterion

The geographical distribution (see **Figure 5**) of the selected articles is presented, from highest to lowest incidence, as follows: USA (14), Iran (4), India (3), Spain (3), South Korea (2), China (2), Indonesia (2), and the rest of the 15 countries

TABLE 2 Dimensions of the analytical framework with components, symbols, categories, and examples for the analysis of the obtained data.

Dimensions of the analytical framework							
I. Scientific visibility and institutional backing			**II. Multidimensional classification of the structural components of the analyzed research**				
Component	**Symbol**	**Category**	**Component**	**Symbol**	**Category**	**Example: Author, year/Fragment of article**	
No of quotations (Google Academic)			Perspective	COM	Communication	Giménez-Moreno, 2012	Examine (.) the interdependence between relevance and repetition in current lecturing by firstly reviewing the main communicative strategies.
Impact factor JCR (sanctioning index of the relative relevance of the scientific journal depending on the quotations received)				COG	Cognition	Monk and Newton, 2018	Is the change in students SL skills related to their learning approach (i.e., deep versus surface learning approaches)?
Country (university) of reference relevance sanction				ACT	Active	Bailey et al., 2012	Keywords: active learning, cooperative/collaborative education
Field of knowledge/Studies	MED	Medicine and Health Sciences	Aim	REN	Performance	Kinnari-Korpela, 2014	Students' learning is assessed
	SOC	Social Sciences: Education, Psychology		DIS	Dispositional	Starichenko et al., 2013	Reveal students' attitude to CRS abilities
	CEX	Exact Sciences		INS	Instructional	Werner et al., 2018	Study about teachers
	BIO	Biological Sciences		COB	Combination	Chimmalgi, 2019	Students' satisfaction and performance are assessed
	HUM	Humanities	Methodology	QUAN	Quantitative	Vázquez and Chiang, 2016	Controlled clinical experiment
	EAD	Economic and Business Administration Sciences		QUAL	Qualitative	Ní Riain et al., 2018	Qualitative research on the use of role-play
	DIV	Diverse or unidentified		MM	Mixed-Method	Navaz, 2013	Mixed-method perspective
Continent	AME	America	Result	FOR	Formative	Nordin et al., 2013	Title: Developing a Specialized Vocabulary Word List in a Composition Culinary Course through Lecture Notes
	EUR	Europe		TEC	Technological	Mathiasen, 2015	Title: Digital Voting Systems and Communication in Classroom Lectures

(Continued)

10.3389/fpsyg.2022.971617

TABLE 2 (Continued)

Dimensions of the analytical framework						
I. Scientific visibility and institutional backing			II. Multidimensional classification of the structural components of the analyzed research			
Component	**Symbol**	**Category**	**Component**	**Symbol**	**Category**	**Example: Author, year/Fragment of article**
	ASI	Asia		CRI	Critical	Burnham and Mascenik, 2018 Title: Comparison of student performance and perceptions of a traditional lecture course versus an inverted classroom format for clinical microbiology
	AFR	Africa		MIX	Mixed	Drouin, 2014 Title: If You Record It, Some Won't Come: Using Lecture Capture in Introductory Psychology
	ACE	Oceania				

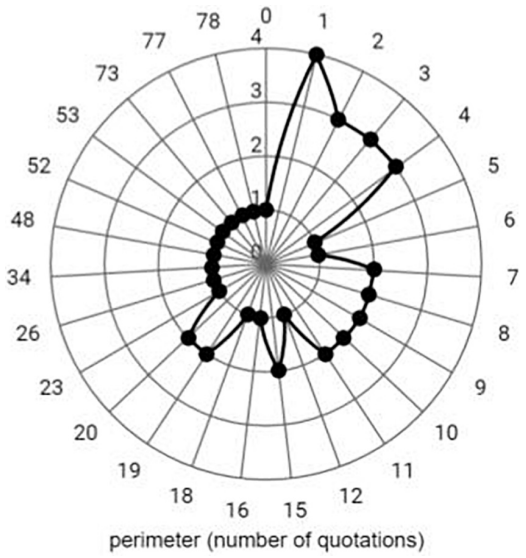


FIGURE 3
Relationship between number of quotations and number of articles.

(1). The geographical distribution covers the continents: Asia (18), North America (15), Europe (8), Africa (3), and Oceania (1), and highlights the absence of articles from Central and South America. The diversity of countries and continents reflects the diversity of educational policies and proposals for the improvement of university education in the twenty-first century. However, the seminal ideas of a teaching founded on the attention to differences and on the proposal of active teaching methods with a vision tinged by constructivism and support for a spirit of collaboration, appear to emerge in the discourse of the international communities and groups dedicated to educational research within higher education. Since the UNESCO Declaration (1998) and the The Bologna Declaration (1999) on innovative educational methods, the lines of improvement converge on an intercontinental level in pedagogic and didactic terms. Another matter entirely is the availability of means (economic, equipment, teacher training, etc.) and the cultural codes involved in the regulation of the complex factors present in situations of interactive learning induced by expository and highly specialized formats of teaching, as is the case of postgraduate and doctoral lectures.

Field of knowledge: plasticity of innovative teaching methods criterion

In the group of selected articles, the studies deal with different areas of knowledge (see **Table 4**): MED (11), SOC (11), CEX (9), HUM (5), EAD (4), DIV (4), and BIO (1). Whilst all the curricular disciplines are involved in creating a renovation in the field of cognitive strategy communication, teamwork, or

10.3389/fpsyg.2022.971617

TABLE 3 Analytical framework of the sample of reviewed empirical articles.

No. of article	Frame of the scientific visibility and institutional backing				Multidimensional classification of structural components			
	No. of quotations	Impact factor	Country	Field of knowledge	Perspective	Guiding aim	Type of study	Results
1	78	0.70/Q4	USA	CEX	ACT	DIS	QUAN	FOR
2	77	–	USA	CEX	COG	REN	QUAN	TEC
3	73	1.394/Q1	Spain	SOC	COM	DIS	QUAN	TEC
4	53	–	USA	SOC	COG	INS REN	QUAN	FOR
5	52	–	Finland	CEX	COM	REN	MM	TEC
6	48	0.667/Q3	USA	SOC	COM	DIS REN	QUAN	TEC CRI
7	34	–	USA	SOC	COM	REN	QUAN	FOR
8	26	–	USA	SOC	COM	REN	QUAN	FOR
9	23	1.302/Q1	Belgium	SOC	ACT	DIS	QUAN	FOR
10	20	–	Sri Lanka	DIV	COM	INS DIS	MM	FOR
11	20	–	Iran	CEX	COM	REN	QUAN	FOR TEC
12	19	–	Turkey	MED	COG	REN	QUAN	FOR TEC
13	19	–	India	MED	ACT	REN	QUAN	FOR TEC
14	18	5.627/Q1	South Korea	SOC	COG	REN	QUAN	FOR TEC
15	16	–	Indonesia	BIO	ACT	REN	QUAN	FOR
16	15	–	Denmark	CEX	COM	REN	QUAN	TEC
17	15	–	USA	SOC	ACT	DIS	QUAN	FOR
18	12	–	USA	EAD	COG	REN	QUAN	FOR TEC
19	11	–	USA	MED	COM	DIS REN	QUAN	CRI
20	11	–	Russia/Israel	DIV	COM	DIS	QUAN	TEC
21	10	–	China	MED	COM	REN	QUAN	FOR
22	10	–	Australia	DIV	COM	DIS	QUAN	FOR TEC
23	9	–	Iran	CEX	COM	REN	QUAN	CRI
24	9	Q2	South Africa	HUM	COG	REN	QUAN	FOR TEC
25	9	–	Zimbabwe	DIV	COM	DIS	QUAN	TEC
26	8	3.759/Q1	India	MED	COG	DIS REN	QUAN	FOR
27	8	–	Malaysia	HUM	COM	INS	QUAN	FOR
28	7	2.297/Q1	South Korea	HUM	COG	REN	QUAN	FOR TEC
29	7	–	USA	CEX	ACT	DIS REN	QUAN	CRI
30	6	–	USA	SOC	COG	REN	QUAN	FOR
31	5	–	Poland/ Ukraine	SOC	COM	DIS REN	QUAN	FOR TEC
32	4	Q3	Iran	MED	COG	REN	QUAN	CRI
33	4	–	USA	CEX	COG	REN	QUAN	FOR
34	4	–	USA	EAD	ACT	INS	QUAN	FOR
35	3	0.7/Q2	Spain	EAD	COM	INS	QUAN	FOR
36	3	–	China	HUM	COM	REN	QUAN	FOR TEC
37	3	–	Spain	EAD	COM	INS	QUAN	FOR
38	2	–	USA	MED	ACT	REN	QUAN	FOR TEC
39	2	–	Indonesia	SOC	COM	REN	QUAN	CRI
40	2	–	Canada	MED	COG	DIS REN	QUAN	FOR
41	1	Q2	Ethiopia	MED	COG	DIS REN	QUAN	FOR
42	1	–	Kazakhstan	CEX	COM	REN	QUAN	FOR
43	1	–	India	MED	ACT	REN	QUAN	CRI
44	1	–	Ireland	HUM	ACT	REN	QUAL	FOR
45	0	Q3	Iran	MED	ACT	DIS REN	QUAN	CRI

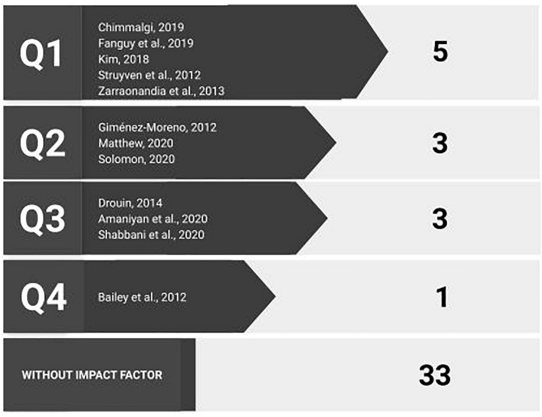


FIGURE 4
Placement of articles in journals with impact factor (JCR-WoS) and quartile (Q).

in the way of assessing results, the concern about improving the lecture format remains a didactic setting that can be adapted to interactive and collaborative learning conditions.

Multidimensional classification of structural components

The articles cover three perspectives: COM (21), COG (13), and ACT (11). The plasticity of the didactic methods shown above facilitates a complex approach to the research object. The limits are imposed by the research object as defined by the researcher, although the frequency of the studies assigned to one category or another is not the most important thing. What we wish to highlight is that the three orientations are present in the analyzed sample.

The guiding-aim of the selected studies correspond to: INS (4), DIS (7) and REN (24) and in 10 articles two of these aims are combined: INS + REN (1), INS + DIS (1), and DIS + REN (8). One sign of lecture maintenance is precisely that responses are designed to the problems of performance, and to those arising from teacher conduct as being responsible for the teaching action involved in all the facilitating modes of interactive learning.

The methodologies used in the generation and analysis of data focus on the QUAN perspective (42), while the other two options are only present in three articles (QUAL, 1 and MM, 2). It is important to underline that the journals with a higher impact factor located in the quartiles Q1 and Q2 do not propose any methodological restrictions, whilst some of them even advocate openly for the inclusion of research that responds to the quantitative, qualitative and mixed-method approaches. These journals are: *Anatomical Sciences Education, British Journal of Educational Technology, Interactive Learning Environments, Computers & Education, International Review of Research in Open and Distributed Learning, Research in Learning Technology, Journal of Pragmatics, Advances in Medical Education and Practice.*

The applied results and/or recommendations from the studies influence the following aspects: FOR (20), CRI (7), TEC (6), and in 12 articles two of these are combined: FOR + TEC (11) and TEC + CRI (1). What stands out is the teaching value of the lecture and the incorporation of telematic computerized resources. It also provides a reflection about the role of technical aspects in pedagogic and didactic improvements.

The intersection of the conceptual perspective dimension with the methodological affiliation provides us with a new picture of the data when the nine resulting regroupings are considered (see **Table 5**).

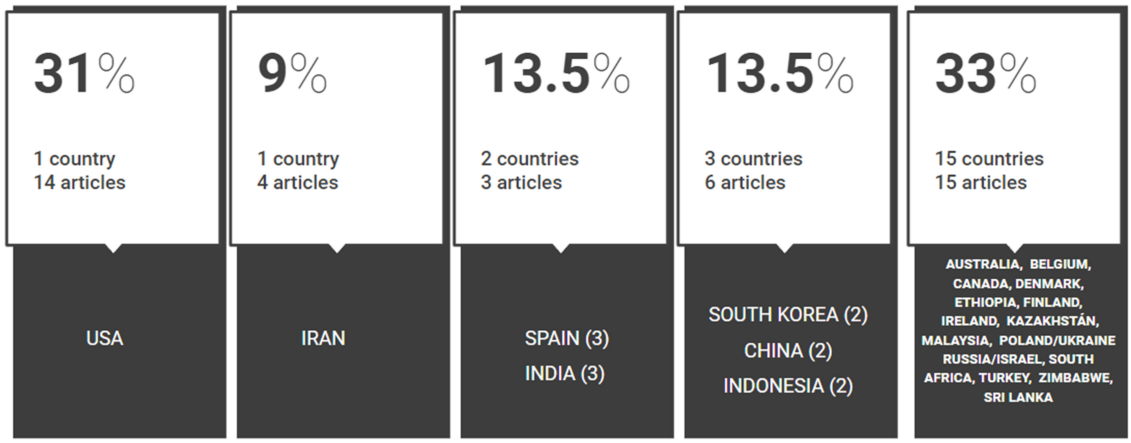


FIGURE 5
Geographical provenance ordered by number of articles.

TABLE 4 Fields of knowledge of the analyzed studies.

Field of knowledge	Number of articles
MED	11
SOC	11
CEX	9
HUM	5
EAD	4
DIV	4
BIO	1
Total	**45**

TABLE 5 Conceptual-methodological commitment (perspectives and methodologies).

Perspective conceptual/ Methodology	COM	COG	ACT	Total
QUAN	3, 6, 7, 8, 11, 16, 19, 20, 21, 22, 23, 25, 27, 31, 35, 36, 37, 39, 42	2, 4, 12, 14, 18, 24, 26, 28, 30, 32, 33, 40, 41	1, 9, 13, 15, 17, 29, 34, 38, 43, 45	**42**
QUAL			44	**1**
MM	5, 10			**2**
Total	**21**	**13**	**11**	**45**

An analytical view of the methodological commitment in each regrouping is presented as a whole in Table 6, characterized by:

Methodological commitment QUAN-COM, QUAN-COG, and QUAN-ACT

There are 19 QUAN-COM articles (42.2%) and they include 301 quotations (39.7%). Three articles have an impact factor (Q1, Q2 and Q3), and are from diverse fields of knowledge: SOC (6), CEX (4), DIV (3), EAD (2), MED (2); are from different continents: AME (4), ASI (8), EUR (5), AFR (1), and OCE (1). The aims are of type REN (9), DIS (4), INS (3) and DIS-REN (3), and the results of type FOR (7), TEC (4), CRI (3), TEC-CRI (1) and FOR-TEC (4).

There are 13 QUAN-COG articles (29%) and they include 220 quotations (29%). Six articles have an impact factor (Q1, Q2, and Q3), and are from diverse fields: MED (5), SOC (3), CEX (2), HUM (2), and EAD (1); and are from different continents: AME (6), ASI (5), and AFR (2). The aims are of type REN (9), DIS-REN (3), and INS-REN (1), and the results are of type FOR (6), FOR-TEC (5), TEC (1), and CRI (1).

There are 10 QUAN-ACT articles (22%) and they include 165 quotations (21.7%). Four have an impact factor (Q1, Q3, and Q4), are from diverse fields of knowledge: MED (4), CEX (2), SOC (2), and EAD (1), and are from three continents: AME (5), ASI (4) and EUR (1). The aims are of type REN (4), DIS (4), DIS-REN (2), and INS (1), and the results are of type FOR (5), CRI (3), and FOR-TEC (2).

TABLE 6 Analytical view of the methodological commitment in each regrouping.

Methodological commitment	No of articles	No of quotations	Impact factor					Field of knowledge							Continent					Aim				Result			
			Q1	Q2	Q3	Q4	SFI	SOC	CEX	MED	HUM	EAD	BIO	DIV	AME	ASI	EUR	AFR	OCE	REN	DIS	INS	COB	FOR	TEC	CRI	MIX
QUAN-COM	19 (42.2%)	301 (39.7%)	1	1	1	0	16	6	4	2	2	2	0	3	4	8	5	1	1	9	4	3	3	7	4	3	5
QUAN-COG	13 (29%)	220 (29%)	3	2	1	0	7	3	2	5	2	1	0	0	6	5	0	2	0	9	0	0	4	6	1	1	5
QUAN-ACT	10 (22%)	165 (21.7%)	2	0	1	1	6	2	2	4	0	1	0	0	5	4	1	0	0	4	3	1	2	5	0	3	2
QUAL-ACT	1 (2.3%)	1 (0.1%)	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0
MM-COM	2 (4.5%)	72 (9.5%)	0	0	0	0	2	0	1	0	0	0	1	1	0	1	1	0	0	1	0	0	1	1	1	0	0
Total	**45 (100%)**	**759 (100%)**	**6**	**3**	**3**	**1**	**32**	**11**	**9**	**11**	**5**	**4**	**1**	**4**	**15**	**18**	**8**	**3**	**1**	**24**	**7**	**4**	**10**	**20**	**6**	**7**	**12**

To summarize, the presence of the quantitative conceptual-methodological commitment in all the other structural aspects considered, and its prevalence in the most accredited scientific media, leads us to the conclusion that the web of quotations could provide interesting results for the subjects they cover.

Methodological commitment QUAL-ACT

One QUAL-ACT article was identified, with one quotation and no impact factor. It is from Ireland, and the field of knowledge is HUM. The aim is of type REN and the result is FOR.

In light of this datum and taking into account the comment about the methodological aperture of the journals with a JCR impact factor, it is worth underlining the numerous possibilities offered by qualitative methodology (Wertz et al., 2011).

Methodological commitment MM-COM

There are 2 MM-COM articles; they include 72 quotations (9.6%) and have no impact factor. The fields of knowledge are CEX and BIO, and they are from Europe and Asia. The aims are REN in one article and INS-DIS in the other, and the results are TEC and FOR.

Given that we believe the mixed-method conceptual-methodological commitment is suitable for the study of the innovative renovations of the lecture format (Tronchoni et al., 2021), and the scientific works with methods considered as inherent integrators of qualitative and quantitative data analysis (Anguera and Izquierdo, 2006; Bazeley, 2018; Izquierdo and Anguera, 2021), we can only hope that the MM commitment will be present in more studies. On the other hand, the products achieved with this conceptual-methodological commitment are appreciated, valued and recommended for publication in the best placed journals that cover the subjects of teaching and learning in higher education.

Discussion and conclusions

A descriptive analysis of the results obtained via the coding of the criteria dimensions in order to capture the differences produced by the systematically selected sample, produces two conclusions that provide answers to the two research questions posed in the introduction: the first conclusion responds to the question about the current scientific situation of the renovated lecture format subjected to empirical study. The second conclusive response places value on the identified processes and results, whilst at the same time demonstrating the need to articulate a proposal that incorporates an open and dialogued vision of the teaching system, whose continued renovation should be founded on empirical research in as far as this is possible and necessary.

Understanding the evolution of the relationships between the various components of the teaching system is vital if universities are to offer effective and efficient teaching. Given that we have found no other systematic reviews of the proposed key terms, it is not possible either to verify whether the description carried out and the conclusions we propose are in line with other reviews, nor to indicate in which aspects our findings differ from those provided by other systematic reviews of the lecture format.

Before entering into the argumentation thread that sustains the inferential and proactive path of this section, it is worth pointing out that our empirical conceptual-methodological approach connected with educational assessment (Tronchoni et al., 2021) has a point of contact—differences aside—with the *Direct Instruction* movement (Engelmann and Colvin, 2006). When it comes to discussing the coarse matter of the frequency distribution of the different criteria and the subtle silence of the vacuums that the analyzed results produce, we include as a contrast the idea of systematizing the way the acquisition of new knowledge is accessed in terms of direct instruction proposals, i.e., the teaching system of the lecture in our case. We believe that the vacuums or lowest scores can provide a certain generalization of interest about the need to construct a common base open to a plurality of viewpoints, but with a clear message about the need to systematize the lecture format without renouncing flexibility, plasticity, web connectivity, or sustainable effectiveness.

> Research Question 1. Is there an interest in renovating the lecture format among the international research community whose remit is university teaching methods?

Conclusion 1. The geographical channeling of institutional backing and the evaluation of the visibility and scientific reliability of the web of quotations is proof positive for tackling the internationalized challenge of the renovation of the lecture format. Unfortunately, the methodological commitment remains incomplete in not providing qualitative and mixed-method studies, and the databases consulted are not sensitive to the research carried out in South American countries.

This conclusion is based on the following evidence found in our research:

a) In the group of selected articles, the lecture is present in a wide diversity of journals. Some of them show a JCR impact factor, elaborated on the *Web of Science* platform (WoS). The visibility of this subject in the scientific-academic community interested in the renovation and innovation of teaching methods in general and the lecture in particular, is guaranteed in the period consulted. The five journals with the highest impact factor according to year of publication, in ascending order are:

Interactive Learning Environments (2012), *British Journal of Educational Technology* (2013), *International Review of Research in Open and Distributed Learning* (2019), *Anatomical Sciences Education* (2019), and *Computers & Education* (2018). The conceptual-methodological commitments QUAN-COM (3), QUAN-COG (6), and QUAN-ACT (3) were present in the articles with an impact factor. The articles conceived as QUAL-ACT (1) and MM-COM (2), were published in journals indexed on other platforms.

b) The expository lecture, with a greater or lesser scientific visibility, is a subject that raises interest among researchers in different countries on all continents. The geographical prevalence of the selected articles belongs to the scientific production of researchers in Asia (mainly Iran, India, South Korea, China and Indonesia) and the USA. Our database search did not produce selected articles from South America. On the other hand, Asia and North America cover the methodological commitments QUAN-COM (12), QUAN-COG (11), and QUAN-ACT (9); Europe shows an interest in the methodological commitment QUAN-COM (5) and Africa in QUAN-COG (2) and QUAN-COM (1). Once again, the options that structure the research object with a methodological commitment QUAL or MM were silenced.

c) The appearance of quotations in the scientific production of articles is an indicator of the appreciation and value given to the subject, and of its subsequent incorporation into new articles that promote the applied proposals in real situations within the field of the acquisition of curricular, declarative and procedural knowledge, and of that pertaining to the area of values, attitudes, and emotions. Quotations generate networks of interest through the mobilization of said advances, creating tendencies within studies. This suggests that networks of influence are being formed. The expository lecture is defined by a directly visible web potential that has set in motion 759 connections in the total of the 45 articles in our sample. Only one recently published article did not register any quotations. The web is structured into lesser groupings depending on the methodological commitments found from the QUAN commitment and these are thus distributed from greater to lesser number of quotations: QUAN-COM (301), QUAN-COG (220), and QUAN-ACT (165).

> Research Question 2. What improvements to the lecture format do the reviewed articles suggest, within the framework of the communicative matrix of interactive learning?

Conclusion 2. It is very difficult to know the characteristics of the lectures that have been subject to intervention and their relationship with other teaching and learning methods. However, the internationalized agenda of the subjects covered is sufficiently pertinent to give rise to partial improvements in the exploitation of technological opportunities (ICT) applied to the transmission of knowledge, the use of strategies and the inclusion of participative tasks and techniques. Unfortunately, the analyzed sample does not reflect the concept of a communicative matrix within the organizational and institutional context of the intervened lectures; whilst all the articles are in line with the ideal of promoting *interactive learning,* the consideration as to how teaching should be adjusted does not appear.

The reflection of the obtained data on the improvements achieved by the interventions carried out in lectures covers the following subjects:

a) *Information processing and performance.* A total of 17.7% of the higher education sample analyzed describes teaching based on improvements in lecture design, beginning with the problems raised by providing information to be effectively remembered (DIS-REN, 8). The subject of attention and memory functions in interactive learning is linked to good performance and an increase in learning potential in lectures (French and Kennedy, 2017). Another aspect to bear in mind when considering the effects of the informative approach proposed to the students is whether to promote learning based on investigative competencies or on repetitive production activities (Lundvall and Johnson, 2016). Since both forms of learning are complementary, there is room for the design of mixed trajectories. The shadow of unmonitored (by the teacher) repetitive learning grows longer when we consider that 53% of our sample places emphasis on performance (REN, 24). We do not know the diversified cultures of the universities that use the lecture method, nor do we know the relationship between this format and other methods applied in class sessions, but the emphatic concern about performance might be indicative of a more conventional (the exclusive performance of declarative content) rather than innovative feature of the researched teaching practice.

b) *Prevalent thematic resonance.* Two articles with greater resonance show the polarities of the professional and research interests of the studies carried out.

On the one hand, the article by Bailey et al. (2012) with IF (JCR) 0.70 (Q4) has 78 quotations (29% of the group of articles with an impact factor). This article was published in the journal *The International Union of Biochemistry and Molecular Biology,* by three professors from the Biochemistry Department of the University of Nebraska and the Chemistry Department of Seattle University (USA). The article deals with how to transform an expository lecture into a format centered on the students' learning. The professional nature of the proposal

connects with the pedagogical revolution of renovating the lecture by incorporating opportunities for participation (ACT), and developing skills (DIS) that are necessary for the learning of content that can be transmitted with an expository format (FOR).

On the other hand, the article by Hegeman (2015), of Missouri Western State University (USA), has 77 quotations (16% of the group of articles with no impact factor) and was published in the journal *Online Learning*. The article analyzes the results of learning facilitated by the use of multimedia material and note-taking. The students are given cards designed by the teacher in order to guide the study of pre-recorded expository algebra sessions. The article, conceived from a cognitive perspective (COG), promotes the implication of the teaching staff in the technical handling (TEC) of telematic opportunities and the construction of tutorial material, with the aim of facilitating a greater performance (REN) from the students.

c) *Communicative matrix of the educational system.* It is striking that just 13% of the sample gives a description of teaching that places the focus on the teaching format in relation to performance or student aptitude (INS-REN, 1; INS-DIS, 1), and on the appropriateness of the lecture for carrying out certain learning (INS 4). One silenced aspect derived from this fact is the concern about the influence of teacher-student learning relationships (Entwistle, 2018). The good design and good application of an educational system—which may be the improved lecture—should incorporate the complex loop of communicative interactions between the four identifiable levels of exchange in the educational system: (i) the political-economic and cultural-educational conditions both inside and outside the classroom, (ii) the academic and professional side of the disciplines, (iii) the physical and virtual meeting spaces, and (iv) the personalities of the students and teachers. This reading of the renovated lecture reinforces the ritual of roles (Goffman, 1967) that can promote dialogic interaction within the juncture of teaching with interactive learning. There are those who search for this juncture in valuing the role played in the design and execution of the lecture by the active, attentive student who is a synthesizer of retrievable and revisable information. Another line of juncture seeks to potentiate oral participation. In this context of considerations, 71% of the reviewed studies direct their concern about communication toward academic performance, with some mentions of student aptitude or the expository teaching of the content covered (COM-REN, 10; COM-COB, 4). If we take into account that 47% of the articles incorporate the COM perspective, then it would be important for the subject of academic performance to predominate in order to vouch for the virtues of the renovated lecture format. We suspect that the communicative focus does not respond to a holistic and dialogic approach, and that the concern about individual student aptitude is centered on the satisfaction that preferably cognitive experiences generate in being able to drive learning itself, with the aid of the teacher's exposition and the development of necessary strategies and abilities.

d) *The geographical and institutional extent of attempts to improve the lecture format does not provide the necessary contextualization of those innovations.* The internationalization of the aim of improving the lecture format, understood as an expert teaching format, is undoubtedly benefiting from the possibilities offered by ICT, multimedia options and the combined use of face-to-face and synchronous virtual sessions (O'Callaghan et al., 2017; Gegenfurtner and Ebner, 2019; Raes et al., 2020). These new possibilities, sensitive to the educational and economic conditions of each university and country, present the need to develop specific designs in terms of how to implement and drive instructional participative interaction that mobilizes reasons to cooperate, together with the strategic use of collaborative cognitive-emotional abilities implied in the process of acquiring academic knowledge.

The reviewed articles propose changes (COM, COG, ACT) but do not give contextual keys to understanding how to go about it and to conceptualize what is proposed as an intellectual and experiential result above and beyond a mere recuperation of information.

To summarize, from standpoint regarding the conceptualization of the innovation of the lecture format and the conceptual-methodological commitment adopted in the empirical research of this subject (Tronchoni et al., 2021), we believe that the review carried out alerts us to the lack of studies that provide an integrated response to the conditions and roles of interactive learning, together with tackling the production of emotional-intellectual experiences that reinforce the dialogic and collaborative links of all the participants. Whilst all the methodological options might be appropriate for structuring empirical studies about the improvements sought by good interactive design and a good execution of the renovated lecture, we would like to underline that the mixed-method approach of systematic observation (Anguera et al., 2017) fits well with the idea of being able to finalize reliable formative assessments contingent upon the diversity of people, the disciplines involved, space-time conditions, own and imported educational cultures, and the most distal influences. The focus of the observation centers on participative interaction, a mechanism responsible for the organization of exchanges and for controlling the means of producing academic knowledge, and for the emotional-intellectual experience.

Above and beyond the satisfaction produced by academic results, the emotional-intellectual experience that emerges from the social implication in the construction of knowledge can be considered a powerful resource for personal growth and collective wellbeing (Claxton, 1984).

Every systematic review has inherent limitations to its own profile—such as the proposal of selection criteria for primary documents—that inevitably have an influence on the results, both on those obtained, and on the vacuums detected. An example of this is the *culture of research* itself which may exist in relation to the expository lecture in Latin American countries, made invisible by opting for selection criteria that we feel to be suitable (such as the English language). Furthermore, another limit could be that the filter with the term *lecture* may hide diverse understandings of the lecture format within the specification of each study.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

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

Acknowledgments

We gratefully acknowledge the support of a Spanish Government subproject *Integration ways between qualitative and quantitative data, multiple case development, and synthesis review as main axis for an innovative future in physical activity and sports research* (PGC2018-098742-B-C31) (2019–2021) (Ministerio de Ciencia, Innovación y Universidades, Programa Estatal de Generación de Conocimiento y Fortalecimiento Científico y Tecnológico del Sistema I+D+i), that is part of the coordinated project *New approach of research in physical activity and sport from mixed methods perspective* (NARPAS_MM) (SPGC201800X098742CV0).

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amaniyan, S., Pouyesh, V., Bashiri, Y., Snelgrove, S., and Vaismoradi, M. (2020). Comparison of the conceptual map and traditional lecture methods on students' learning based on the VARK learning style model: A randomized controlled trial. *SON* 6, 1–9. doi: 10.1177/2377960820940550

Anguera, M. T., and Izquierdo, C. (2006). "Methodological approaches in human communication: From complexity of perceived 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., 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

Bailey, C. O., Minderhout, V., and Loertscher, J. (2012). Learning transferable skills in large lecture halls: Implementing a POGIL approach in biochemistry. *Biochem. Mol. Biol. Educ.* 40, 1–7. doi: 10.1002/bmb.20556

Barr, R. B., and Tagg, J. (1995). From teaching to learning: A new paradigm for undergraduate education. *Change* 27, 12–26. doi: 10.1080/00091383.1995.10544672

Bates, M., Curtis, S., and Dismore, H. (2017). Learning approaches and lecture attendance of medical students. *J. Furth. High. Educ.* 42, 248–258. doi: 10.1080/0309877X.2016.1261089

Bazeley, P. (2018). *Integrating analyses in mixed methods research*. Thousand Oaks, CA: Sage.

Buchanan, T., and Palmer, E. (2017). Student perceptions of the history lecture: Does this delivery mode have a future in the humanities? *J. Univ. Teach. Learn. Pract.* 14, 1–17. doi: 10.53761/1.14.2.4

Burnham, K. D., and Mascenik, J. (2018). Comparison of student performance and perceptions of a traditional lecture course versus an inverted classroom format for clinical microbiology. *J. Chiropr. Educ.* 32, 90–97. doi: 10.7899/JCE-17-21

Buzzanell, P. M. (2017). Rethinking lecture-learning from communicative lenses: A response to forum essays. *Commun. Educ.* 66, 250–252. doi: 10.1080/03634523.2017.1287412

Chimmalgi, M. (2019). Interactive lecture in the dissection hall: Transforming passive lecture into a dynamic learning experience. *Anat. Sci. Educ.* 12, 191–199. doi: 10.1002/ase.1803

Claxton, G. (1984). *Live and learn. An introduction to the psychology of growth and change in everyday life*. Manhattan, NY: Harper and Row.

Cronbach, L., and Snow, R. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington.

Dannels, D. P. (2016). Opening lines: Scholarly inquiry and learning outcomes in communication. *Commun. Educ.* 65, 480–483. doi: 10.1080/03634523.2016.1208260

Darling, A. (2017). The lecture and the learning paradigm. *Commun. Educ.* 66, 253–255. doi: 10.1080/03634523.2017.1285039

Darling-Hammond, L., Hyler, M. E., and Gardner, M. (2017). *Effective teacher professional development*. Palo Alto, CA: Learning Policy Institute.

Deroey, K. L., and Taverniers, M. (2011). A corpus-based study of lecture functions. *Mod. Spr.* 105, 1–22.

Dolnicar, S., Kaiser, S., Matus, K., and Vialle, W. (2009). Can Australian universities take measures to increase the lecture attendance of marketing students? *J. Mark. Educ.* 31, 203–211. doi: 10.1177/0273475309345202

Drouin, M. A. (2014). If you record it, some won't come: Using lecture capture in introductory psychology. *Teach. Psychol.* 41, 11–19. doi: 10.1177/0098628313514172

Engelmann, S., and Colvin, G. (2006). *Rubric for identifying authentic direct instruction programs*. Eugene, OR: Engelmann Foundation.

Entwistle, N. (2018). *Student learning and academic understanding. A research perspective with implication for teaching*. Cambridge, MA: Academia Press, Elsevier.

Fanguy, M., Costley, J., Baldwin, M., Lange, C., and Wang, H. (2019). Diversity in video lectures: Aid or hindrance? *IRRODL* 20, 44–62. doi: 10.19173/irrodl.v20i2.3838

Fortanet-Gómez, I., and Ruíz-Madrid, N. (2014). Multimodality for comprehensive communication in classroom: Questions in guest lectures. *Ibérica* 28, 203–224.

French, S., and Kennedy, G. (2017). Reassessing the value of university lectures. *Teach. High. Educ.* 22, 639–654. doi: 10.1080/13562517.2016.1273213

Gegenfurtner, A., and Ebner, C. (2019). Webinars in higher education and professional training: A meta-analysis and systematic review of randomized controlled trials. *Educ. Res. Rev.* 28:100293. doi: 10.1016/j.edurev.2019.100293

Giménez-Moreno, R. (2012). The interdependence of repetition and relevance in university lectures. *J. Pragmat.* 44, 744–755. doi: 10.1016/j.pragma.2012.02.013

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face interaction*. London: Aldine.

Hayden, E., and Chory, R. (2018). In defense of the lecture: Revisiting and reassessing its place within management pedagogy. *Organ. Manag. J.* 16, 350–362.

Hegeman, J. S. (2015). Using instructor-generated video lectures in online mathematics courses improves student learning. *Online Learn. Volume* 19, 70–87. doi: 10.24059/olj.v19i3.484

Izquierdo, C., and Anguera, M. T. (2021). The Analysis of interpersonal communication in sport from mixed methods strategy: The integration of qualitative-quantitative elements using systematic observation. *Front. Psychol.* 12:637304. doi: 10.3389/fpsyg.2021.637304

Kim, H. (2018). Impact of slide-based lectures on undergraduate students' learning: Mixed effects of accessibility to slides, differences in note-taking, and memory term. *Comput. Educ.* 123, 13–25. doi: 10.1016/j.compedu.2018.04.004

Kinnari-Korpela, H. (2014). Using short video lectures to enhance mathematics learning – experiences on differential and integral calculus course for engineering students. *Inform. Educ.* 14, 67–81. doi: 10.15388/infedu.2015.05

Kramer, M. W. (2017). Sage on the stage or bone at the board? *Commun. Educ.* 66, 245–247. doi: 10.1080/03634523.2016.1272129

Lundvall, B. Å, and Johnson, B. (2016). "The learning economy," in *The learning economy and the economics of hope*, ed. B. Å Lundvall (London: Arthem Press), 108–130.

Mallin, I. (2017). Lecture and active learning as a dialectical tension. *Commun. Educ.* 66, 242–243. doi: 10.1080/03634523.2016.1275720

Mathiasen, H. (2015). Digital voting systems and communication in classroom lectures. *J. Interact. Media Educ.* 1, 1–8. doi: 10.5334/jime.ah

Matthew, G. (2020). The effect of adding same-language subtitles to recorded lectures for non-native, english speakers in E-learning environments. *RLT* 28:2340. doi: 10.25304/rlt.v28.2340

Mazer, J. P., and Hess, J. A. (2017). Forum: The lecture and student learning. *Commun. Educ.* 66, 236–255. doi: 10.1080/03634523.2017.1287411

Meyer, K. R., and Hunt, S. K. (2017). The lost art of lecturing: Cultivating student listening and notetaking. *Commun. Educ.* 66, 239–241. doi: 10.1080/03634523.2016.1275719

Monk, J. M., and Newton, G. (2018). Use of a scaffolded case study assignment to enhance students' scientific literacy skills in undergraduate nutritional science education: Comparison between traditional lecture and distance education course formats. *Int. J. High. Educ.* 7, 95–106. doi: 10.5430/ijhe.v7n2p95

Navaz, A. M. M. (2013). A study on perception of lecturer-student interaction in english medium science lectures. *Novitas ROYAL Res. Youth Lang.* 7, 117–136.

Ní Riain, I., Dawson, C., and McCarthy, M. (2018). Role-play in literature lectures: The students' assessment of their learning. *Int. J. Scholarsh. Teach. Learn.* 12:8. doi: 10.20429/ijsotl.2018.120108

Nordin, N. R. M., Stapa, S. H., and Darus, S. (2013). Developing a specialized vocabulary word list in a composition culinary course through lecture notes. *Adv. Lang. Lit. Stud.* 4, 78–88. doi: 10.7575/aiac.alls.v.4n.1p.78

O'Callaghan, F. V., Neumann, D. L., Jones, L., and Creed, P. A. (2017). The use of lecture recordings in higher education: A review of institutional, student, and lecturer issues. *Educ. Inf. Technol.* 22, 399–415. doi: 10.1007/s10639-015-9451-z

Özcan, K. (2013). Student evaluation of lecture and teaching effectiveness in higher education. *Educ. Res. Rev.* 8, 378–389. doi: 10.5897/ERR2013.1154

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, G. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372:n71. doi: 10.1136/bmj.n71

Pérez-Llantada, C., and Ferguson, G. R. (2006). *English as a globalization phenomenon: From a linguistic microcosm*. Valencia: Publicacions de la Universitat de València.

Raes, A., Detienne, L., Windey, I., and Depaepe, F. (2020). A systematic literature review on synchronous hybrid learning: Gaps identified. *Learn. Environ. Res.* 23, 269–290. doi: 10.1007/s10984-019-09303-z

Ruesch, J., and Bateson, G. (1951). *Communication, the social matrix of psychiatry*. New York, NY: W. W. Norton & Company.

Samarasekera, D. D., Gwee, M. C. E., Long, A., and Lock, B. (2018). "Lectures and large groups," in *Understanding medical education: Evidence, theory, and practice*, 3rd Edn, eds T. Swanwick, K. Forrest, and B. C. O'Brien (Hoboken, NJ: Wiley Blackwell), 111–121.

Sciullo, N. J. (2017). The lecture's absent audience. *Commun. Educ.* 66, 237–239. doi: 10.1080/03634523.2016.1275722

Shabani, A., Mohammadi, A., Mojtahedzadeh, R., Hosseinic, A., Valadkhani, S., Sistanic, A., et al. (2020). Does the sequence of flipped and lecture-based classes affect the academic achievement and satisfaction of medical students? *J. E-Learn. Knowl. Soc.* 16, 86–93. doi: 10.20368/1971-8829/1135277

Solomon, Y. (2020). Comparison between problem-based learning and lecture-based learning: Effect on nursing students' immediate knowledge retention. *Adv. Med. Educ. Pract.* 7, 947–952. doi: 10.2147/AMEP.S269207

Stacy, J. (2009). The guide on the stage: In defense of good lecturing in the history classroom. *Soc. Educ.* 73, 275–278.

Starichenko, B. E., Egorov, A. N., and Yavich, R. (2013). Features of application of classroom response system at the lectures in Russia and Israel. *Int. J. High. Educ.* 2:2013. doi: 10.5430/ijhe.v2n3p23

Stearns, S. (2017). What is the place of lecture in student learning today? *Commun. Educ.* 66, 243–245. doi: 10.1080/03634523.2016.1275723

Steinert, Y., and Snell, L. S. (1999). Interactive lecturing: Strategies for increasing participation in large group presentations. *Med. Teach.* 21, 37–42. doi: 10.1080/01421599980011

Stockard, J., Wood, T., Coughlin, C., and Raspica, C. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Rev. Educ. Res.* 88, 479–507. doi: 10.3102/0034654317751919

Struyven, K., Dochy, F., and Janssens, S. (2012). Explaining students' appraisal of lectures and student-activating teaching: Perceived context and student characteristics. *Interact. Learn. Environ.* 20, 391–422.

Tanahoung, C., Chitaree, R., Soankwan, C., Sharma, M. D., and Johnston, I. D. (2009). The effect of interactive lecture demonstrations on students' understanding of heat and temperature: A study from Thailand. *Res. Sci. Technol. Educ.* 27, 61–74. doi: 10.1080/02635140802658909

The Bologna Declaration (1999). *Joint declaration of the european ministers of education about the european higher education area*. Available online at: http://www.mec.es/universidades/eees/files/Declaracion_Bolonia.pdf (accessed April 25, 2007).

Thwin, E. P. A., and Lwin, Z. (2018). Simple interactive lecturing strategies for fostering students'engagement and active participation. *Med. Sci. Educ.* 28, 203–209. doi: 10.1007/s40670-017-0492-3

Tronchoni, H. (2019). *Estudio observacional de la comunicación multimodal en el aula universitaria: Contextos y estructuras de participación discursiva*

en las sesiones magistrales. [Tesis Doctoral]. Barcelona: Universidad de Barcelona.

Tronchoni, H., Izquierdo, C., and Anguera, M. T. (2018). Participatory interaction in lectures: Theoretical framework and construction of an observational instrument. *Publicaciones* 48, 81–108. doi: 10.30827/publicaciones.v48i1.733

Tronchoni, H., Izquierdo, C., and Anguera, M. T. (2021). Regulation of participatory interaction in university lectures. Co-constructive training proposal based on observational methodology as mixed method strategy. *Publicaciones* 51, 89–110. doi: 10.30827/publicaciones.v51i2.20751

UNESCO Declaration (1998). *La educación superior en el siglo XXI*. Paris: Visión y Acción, UNESCO.

Vázquez, J. J., and Chiang, E. P. (2016). Preparing students for class: A clinical trial testing the efficacy between multimedia pre-lectures and textbooks in an economics course. *J. Coll. Read. Learn.* 13, 37–46. doi: 10.19030/tlc.v13i2.9631

Waldeck, J. H., and Weimer, M. (2017). Sound decision making about the lecture's role in the college. *Commun. Educ.* 66, 247–249. doi: 10.1080/03634523.2016.1275721

Werner, J. M., Scovotti, C., Cummings, R. G., and Bronson, J. W. (2018). Building a case for active learning: The use of lecture vs. other classroom activities at LMBC. *J. Learn. High. Educ.* 14, 7–15.

Wertz, F., Charmaz, K., McMullen, L., Josselson, R., Anderson, R., and McSpadden, E. (2011). *Five ways of doing qualitative analysis*. New York, NY: The Guilford Press.

Younis, H., and Elbanna, S. (2022). Teaching in times of crisis: The impact of the COVID-19 pandemic on higher education. *Innov. Educ. Teach. Int.* 1–12. doi: 10.1080/14703297.2022.2060850

Zarraonandia, T., Aedo, I., Diaz, P., and Montero, A. (2013). An augmented lecture feedback system to support learner and teacher communication. *BJET* 44, 616–628. doi: 10.1111/bjet.12047

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 24 November 2022
DOI 10.3389/fpsyg.2022.971126

OPEN ACCESS

EDITED BY
Fco. Pablo Holgado-Tello,
National University of Distance Education (UNED), Spain

REVIEWED BY
Tsuyoshi Ide,
IBM Research, United States
Adeline Leclercq Samson,
Université Grenoble Alpes, France

*CORRESPONDENCE
Jing Lu
luj282@nenu.edu.cn
Jiwei Zhang
zhangjw713@nenu.edu.cn

SPECIALTY SECTION
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

RECEIVED 16 June 2022
ACCEPTED 02 November 2022
PUBLISHED 24 November 2022

CITATION
Cui Y, Lu J, Zhang J, Shi N, Liu J and Meng X (2022) A stochastic approximation expectation maximization algorithm for estimating Ramsay-curve three-parameter normal ogive model with non-normal latent trait distributions. *Front. Psychol.* 13:971126. doi: 10.3389/fpsyg.2022.971126

A stochastic approximation expectation maximization algorithm for estimating Ramsay-curve three-parameter normal ogive model with non-normal latent trait distributions

Yuzheng Cui[1], Jing Lu[1]*, Jiwei Zhang[2]*, Ningzhong Shi[1], Jia Liu[1] and Xiangbin Meng[1]

[1]Key Laboratory of Applied Statistics of Ministry of Education, School of Mathematics and Statistics, Northeast Normal University, Changchun, China, [2]Faculty of Education, Northeast Normal University, Changchun, China

In the estimation of item response models, the normality of latent traits is frequently assumed. However, this assumption may be untenable in real testing. In contrast to the conventional three-parameter normal ogive (3PNO) model, a 3PNO model incorporating Ramsay-curve item response theory (RC-IRT), denoted as the RC-3PNO model, allows for flexible latent trait distributions. We propose a stochastic approximation expectation maximization (SAEM) algorithm to estimate the RC-3PNO model with non-normal latent trait distributions. The simulation studies of this work reveal that the SAEM algorithm produces more accurate item parameters for the RC-3PNO model than those of the 3PNO model, especially when the latent density is not normal, such as in the cases of a skewed or bimodal distribution. Three model selection criteria are used to select the optimal number of knots and the degree of the B-spline functions in the RC-3PNO model. A real data set from the PISA 2018 test is used to demonstrate the application of the proposed algorithm.

KEYWORDS

item response theory, Ramsay curve, 3PNO model, marginal maximum likelihood estimation, stochastic approximation EM algorithm (SAEM), density estimation

1. Introduction

A premise of item response theory (IRT) is that observed item responses are indicators of one or more latent traits. For item parameter estimation, the latent variable is usually presumed to be normally distributed. However, many psychological constructs, such as ambition, dysthymia, and borderline personality disorder, as well as other latent traits in sociology such as drug abuse, are unlikely to be normally distributed in a general population.

For example, a psychiatric disorder trait is typically positively skewed in a general population because most people are located at the non-pathological end of the trait, whereas a small group of individuals is spread out along the mild, moderate, and severe end of the disorder (Woods, 2006; Woods and Thissen, 2006; Wall et al., 2015; Wang et al., 2018). In addition, variables representing symptoms of pathology that are rare in the general population may be skewed because they exist at low levels for most people and at high levels for a few individuals. Therefore, the assumption of the normal distribution of latent traits leads to biased parameter estimates when the true latent trait distribution $g(\theta)$ is non-normal (Woods, 2006, 2007; Woods and Thissen, 2006; Woods and Lin, 2009; Azevedo et al., 2011; DeMars, 2012; Molenaar et al., 2012; Reise and Revicki, 2014; Wall et al., 2015; Reise et al., 2018).

Various studies have developed approaches to dealing with non-normal distributions of latent traits. In particular, the empirical histogram (EH) approach (Bock and Aitkin, 1981; Reise and Revicki, 2014) has been proposed to estimate the height of the latent trait density $g(\theta)$ at each quadrature point (i.e., the values on the θ continuum, usually the number of quadrature points is large) instead of computing the heights based on the normal density. This is more flexible than the expectation maximization (EM) algorithm (Bock and Lieberman, 1970; Dempster et al., 1977; Bock and Aitkin, 1981), which is restricted to the normal assumption of latent traits. However, the EH method is sensitive to the user-specified rectangular quadrature scheme. In addition, the graphical representations from the EH method are usually "choppy" or jagged, which makes it difficult to use them to clarify the characterizations of latent traits.

To address this issue, several methods have been proposed to approximate the latent trait density more precisely, including log-linear smoothing (LLS; Casabianca and Lewis, 2015), the Davidian curve (Woods and Lin, 2009), and the Ramsay curve (Woods, 2006; Woods and Thissen, 2006). When incorporated with IRT, these methods simultaneously estimate the latent trait density and the item parameters, but they use different approaches to estimate the latent trait density $g(\theta)$. Specifically, LLS matches M moments of the original distribution to create a smoothed distribution of latent traits while making fewer assumptions about its form and maintaining parsimony. Davidian-curve IRT (DC-IRT; Woods and Lin, 2009) provides a smooth representation of $g(\theta)$ using a unidimensional Davidian curve, as described by Zhang and Davidian (2001). Ramsay-curve IRT (RC-IRT; Woods, 2006; Woods and Thissen, 2006) estimates the latent trait density using Ramsay curves based on B-spline functions.

This paper focuses on RC-IRT because it is one of the most flexible and easy-to-apply methods for estimating the latent-ability density. Woods and Thissen (2006) first introduced RC-IRT to detect and correct for the non-normality of $g(\theta)$, and they estimated the item parameters of the two-parameter logistic (2PL; Birnbaum, 1968) model using the marginal maximum likelihood estimation with the EM (MML-EM; Baker and Kim, 2004) method. Newton–Raphson iteration was used to update the shape parameter $\boldsymbol{\eta}$. Woods (2008) extended this approach to estimate an RC-IRT model for the three-parameter logistic (3PL; Birnbaum, 1968) model. Subsequently, Monroe and Cai (2014) proposed using a Metropolis–Hastings–Robbins–Monro (MHRM; Cai, 2010) algorithm to estimate an RC-IRT approach for the unidimensional graded response model (GRM; Samejima, 1969). The major advantage of this approach is that the covariance matrix estimates can be easily obtained as a byproduct. However, in their study, the GRM was limited to the logistic version.

To date, there have been no studies examining RC-IRT in the context of the normal ogive model. The reason for this is that the EM algorithm usually involves numerical integration calculations, which is intractable for the normal ogive model itself because it already includes an integral term. To fill this knowledge gap, we propose to estimate the RC-3PNO model using a stochastic approximation EM (SAEM; Delyon et al., 1999) algorithm. Specifically, the integral calculation in the E-step is replaced by a stochastic approximation method, which greatly simplifies the calculation. After introducing latent variables, the complete data likelihood can be transformed into an exponential family distribution; thus, sufficient statistics can be used to easily update the estimates of item parameters in the M-step. This avoids the calculation of derivatives in the original M-step in the EM algorithm and further improves the computation efficiency. The latent density distribution $g(\theta)$ is estimated using traditional Newton–Raphson iteration for the proposed SAEM algorithm.

Note that for the estimation of shape parameter $\boldsymbol{\eta}$ and item parameters, the Bayesian maximum a posteriori (MAP; Greig et al., 1989) estimate is used (this will be interpreted further in Section 3). The posterior simulation methods make the posterior distributions easier to obtain; that is, the algorithms for posterior simulation can be used to obtain approximates of posterior moments. Various Bayesian estimates can be obtained based on the posterior samples obtained from the posterior distributions. In this study, two Bayesian estimates are used. One method is MAP estimate. In fact, the MAP estimate is an estimate of an unknown quantity, which is equal to the mode of the posterior distribution. The MAP can be used to obtain point estimates of an unobserved quantities based on empirical data. It is closely related to the maximum likelihood estimate, but employs an augmented optimization objective that incorporates a prior distribution (additional information available by quantifying the prior knowledge of the interested events). Therefore, MAP estimate can be seen as a regularization of maximum likelihood estimate. Another method is marginal maximum a posteriori (MMAP; Mislevy, 1986; Lee and Gauvain, 1996; Baker and Kim, 2004) estimate. MMAP estimate can be seen as an extension of the MAP estimate by integrating out the latent variables as the

nuisance parameters and then obtaining MAP estimates of the interested parameters. More details for estimation forms of our model can be found in Sections 3 and 4.

The remainder of this article is organized as follows. The second section presents the 3PNO model incorporating a Ramsay curve (denoted as the RC-3PNO model). The third section gives the marginal maximum a posteriori (MMAP) estimation of all the parameters to be estimated in the RC-3PNO model. The fourth section introduces the SAEM procedure for estimating the RC-3PNO model (hereafter referred to as the RC-SAEM algorithm), which is the main contribution of this study. The fifth section presents two simulation studies: one to select the optimal number of knots and the appropriate degree of the B-spline functions for the Ramsay curve and another to assess the performance of the proposed algorithm. A real data set from the PISA 2018 test is then used to demonstrate the application of the proposed algorithm in the sixth section. Finally, conclusions and directions for future research are provided.

2. The RC-3PNO model

Let U_{ij} (with realization u_{ij}) denote the dichotomous response variable of examinee i ($i = 1, \cdots, N$) to item j ($j = 1, \cdots, J$); $U_{ij} = 1$ denotes the correct response, and $U_{ij} = 0$ otherwise. The 3PNO model is defined as

$$P_j(\theta_i) = P(U_{ij} = 1|\boldsymbol{\Omega_j}, \theta_i) = c_j + (1 - c_j)\Phi(a_j\theta_i + b_j), \quad (1)$$

where: $\Phi(\cdot)$ is the cumulative function of the standard normal distribution; $\boldsymbol{\Omega}_j = (a_j, b_j, c_j)$ denotes the characteristic parameters of item j, in which $a_j \in [0, +\infty)$ is the discrimination parameter, $b_j \in (-\infty, +\infty)$ is the intercept parameter, and $c_j \in [0, 1]$ is the guessing parameter; and $\theta_i \in (-\infty, +\infty)$ is the latent-ability parameter of examinee i.

The latent trait distribution should be given in advance for the use of the MML and MMAP estimations in IRT models. In general, it is assumed that $\theta_i \sim N(0, 1)$, which is convenient for statistical computation. However, the normal-distribution assumption is not likely to be tenable, and this will decrease the accuracy of estimates from the MML and MMAP estimations. To address this issue, Woods and Thissen (2006) proposed the Ramsay-curve IRT, which is based on B-spline functions, to describe the latent trait distribution. This provides greater flexibility than the standard normal distribution of latent traits.

Following Woods and Thissen (2006), the latent trait distribution is modeled by

$$g(x_q|\boldsymbol{\eta}) = \frac{\exp[\boldsymbol{B}^*(x_q)\boldsymbol{\eta}]}{C}, \quad (2)$$

where

$$C = \sum_{q=1}^{Q} \exp[\boldsymbol{B}^*(x_q)] \quad (3)$$

is the normalization constant ensuring that $g(x_q|\boldsymbol{\eta})$ ($q = 1, 2, \cdots, Q$) sums to 1. In this study, x_q represents the 121 fixed points with equidistant distance along the interval [-6,6]. To avoid ambiguity with the latent variable with continuous support, the discrete x_q is used here.

In RC-IRT, the density of latent ability, $g(\theta)$, needs to be estimated simultaneously with the item parameters. The shape of the latent-ability density curve is determined by a shape parameter $\boldsymbol{\eta}$, which is a vector whose dimensionality is controlled by the choice of *knots* and *degree*. The dimension of $\boldsymbol{\eta}$ is $m = degree + knots - 1$. The support of the Ramsay-curve density can be numerically represented over a set of discrete points $\{x_q : q = 1, 2, \cdots, Q\}$ along the real number line. In this study, the discrete points are fixed at 121 equidistant values from -6 to 6 separated by steps of 0.1. The interval $[-6, 6]$ is a range of latent traits often used in RC-IRT (Woods and Thissen, 2006; Monroe and Cai, 2014), and this can contain the great majority of the latent abilities being tested. The *knots* are values at which the B-spline functions are connected to each other. Typically, the *knots* are evenly distributed over the range of θ. The number of *knots* in RC-IRT is usually selected from the range 2 to 6 (Woods and Thissen, 2006; Monroe and Cai, 2014). The parameter *degree* refers to the degree of the basic B-spline function. To some extent, although a larger number of *knots* may create a more flexible estimated density curve and a higher *degree* value could accommodate a sharper curve. However, sometimes the increase of *knots* or *degree* will make the Ramsay curve become overfitted, resulting in a more complex model than the appropriate model. And this will increase the parameter to be estimated in the model, which may deteriorate the estimation.

Given *knots*, *degree*, and the discrete points x_q, the corresponding basic B-spline functions $\boldsymbol{B}^*(x_q)$ are determined. The definition and derivation of B-spline function are beyond the scope of this study, and the interested readers can consult De Boor (1978) for details. Gathering Q discrete points together, the basic B-spline functions can be expressed as a $Q \times m$ matrix $\boldsymbol{B}^*$. The matrix $\boldsymbol{B}^*$ is assumed to be known here, and the only parameter that needs to be estimated is $\boldsymbol{\eta}$. Equations (1) and (2) construct the RC-3PNO model, which incorporates the Ramsay curve into the 3PNO model.

3. MMAP estimation for the RC-3PNO model

We denote $\Theta = (\theta_1, \theta_2, \cdots, \theta_N)$ and let $\boldsymbol{\Omega} = (\mathbf{a}, \mathbf{b}, \mathbf{c})$, where $(\mathbf{a}, \mathbf{b}, \mathbf{c}) = (a_j, b_j, c_j)$ $(j = 1, 2, \cdots, J)$. The parameters of the RC-3PNO model that need to be estimated are $\boldsymbol{\Omega}$ and $\boldsymbol{\eta}$, denoted as $\boldsymbol{\zeta} = (\boldsymbol{\Omega}, \boldsymbol{\eta})$.

The conditional distribution of U_{ij} given $\boldsymbol{\Omega}$ and Θ has a binomial form:

$$f(u_{ij}|\theta_i, \boldsymbol{\Omega}_j) = P_j(\theta_i)^{u_{ij}}[1 - P_j(\theta_i)]^{1-u_{ij}}, \quad (4)$$

where $P_j(\theta_i)$ is equal to Equation (1). Based on the local conditional independence assumption (Birnbaum, 1968), the probability of examinee i's conditional response pattern is

$$f(\boldsymbol{u}_i|\theta_i, \boldsymbol{\Omega}_j) = \prod_{j=1}^{J} f(u_{ij}|\theta_i, \boldsymbol{\Omega}_j). \quad (5)$$

The observed data is response matrix $\mathbf{U}$, the person parameter $\Theta = \{\theta_1, \theta_2, \cdots, \theta_N\}$ is viewed as the missing data, and thus the complete data is $(\mathbf{U}, \Theta)$. Conditional independence of item responses is assumed, as well as the independence of respondents, in accordance with practice. Thus, the complete data likelihood of $(\mathbf{U}, \Theta)$ is

$$L(\boldsymbol{\zeta}|\mathbf{U}, \Theta) = \prod_{i=1}^{N}\prod_{j=1}^{J} P_j(\theta_i)^{u_{ij}}[1 - P_j(\theta_i)]^{1-u_{ij}} g(\theta_i|\boldsymbol{\eta}), \quad (6)$$

where $P_j(\theta_i)$ is equal to Equation (1). Taking the natural logarithm of Equation (6), the log-likelihood of $(\mathbf{U}, \Theta)$ can be expressed as

$$\log L(\boldsymbol{\zeta}|\mathbf{U}, \Theta) = \log L(\boldsymbol{\Omega}|\mathbf{U}, \Theta) + \log L(\boldsymbol{\eta}|\Theta), \quad (7)$$

where

$$\log L(\boldsymbol{\Omega}|\mathbf{U}, \Theta) = \sum_{i=1}^{N}\sum_{j=1}^{J}\{u_{ij} \log \frac{P_j(\theta_i)}{1 - P_j(\theta_i)} + \log[1 - P_j(\theta_i)]\} \quad (8)$$

and

$$\log L(\boldsymbol{\eta}|\Theta) = \sum_{i=1}^{N} \log g(\theta_i|\boldsymbol{\eta}). \quad (9)$$

The complete data log-likelihood in Equation (7) can be seen as the sum of two independent parts: the logarithm of the likelihood of item parameters $\boldsymbol{\Omega}$ and the logarithm of the likelihood of $\boldsymbol{\eta}$. Thus, the processes of estimating $\boldsymbol{\Omega}$ and $\boldsymbol{\eta}$ can be conducted separately (Monroe and Cai, 2014), which improves the computational efficiency. Here, the MMAP estimation of $\boldsymbol{\zeta} = (\boldsymbol{\Omega}, \boldsymbol{\eta})$ is given in detail. The priors of $\boldsymbol{\Omega}$ are given below. The prior distribution for (a_j, b_j) is specified as

$$(a_j, b_j)' \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})I(a_j > 0) \quad (10)$$

for $j = 1, 2, \cdots, J$, where $N_2(\cdot)$ denotes a bivariate normal distribution. The prior distribution for c_j is chosen to be,

$$c_j \sim \text{Beta}(\alpha, \beta) \quad (11)$$

for $j = 1, 2, \cdots, J$. According to previous methods of estimating $\boldsymbol{\eta}$ in RC-IRT (Woods and Thissen, 2006; Monroe and Cai, 2014), a diffuse prior density of $\boldsymbol{\eta}$ is assumed:

$$\boldsymbol{\eta} \sim \text{MVN}(\boldsymbol{\mu}_{\boldsymbol{\eta}}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}}). \quad (12)$$

The estimation of $\boldsymbol{\eta}$ will be introduced in detail later.

From Equation (6), the marginalized likelihood of $\boldsymbol{\zeta}$ is

$$L(\boldsymbol{\zeta}|\mathbf{U}) = \prod_{i=1}^{N}\left\{\int \prod_{j=1}^{J} P_j(\theta_i)^{u_{ij}}[1 - P_j(\theta_i)]^{1-u_{ij}} g(\theta_i|\boldsymbol{\eta}) d\theta_i\right\} \quad (13)$$

Based on the priors in Equations (10)–(12) and the marginalized likelihood in Equation (13), the marginalized posterior distribution of $\boldsymbol{\zeta} = (\boldsymbol{\Omega}, \boldsymbol{\eta})$ is

$$\begin{aligned} f(\boldsymbol{\zeta}|\mathbf{U}) &= f(\boldsymbol{\Omega}, \boldsymbol{\eta}|\mathbf{U}) \quad (14)\\ &= \frac{L(\boldsymbol{\Omega}, \boldsymbol{\eta}|\mathbf{U})f(\boldsymbol{\Omega})f(\boldsymbol{\eta})}{\int\int L(\boldsymbol{\Omega}, \boldsymbol{\eta}|\mathbf{U})f(\boldsymbol{\Omega})f(\boldsymbol{\eta})d\boldsymbol{\Omega}d\boldsymbol{\eta}} \propto L(\boldsymbol{\Omega}, \boldsymbol{\eta}|\mathbf{U})f(\boldsymbol{\Omega})f(\boldsymbol{\eta}), \end{aligned}$$

where

$$f(\boldsymbol{\Omega}) = \prod_{j=1}^{J} f(a_j, b_j)f(c_j)I(a_j > 0) \quad (15)$$

is the joint prior density function of $\boldsymbol{\Omega}$, and $f(\boldsymbol{\eta})$ denotes the density function of a multivariate normal distribution. Thus, the MMAP estimation of $\boldsymbol{\zeta} = (\boldsymbol{\Omega}, \boldsymbol{\eta})$ is

$$\begin{aligned} \hat{\boldsymbol{\zeta}} &= \arg\max_{\boldsymbol{\zeta}\in\boldsymbol{\Theta}_{\boldsymbol{\zeta}}} \log f(\boldsymbol{\zeta}|\mathbf{U})] = \arg\max[\log L(\boldsymbol{\zeta}|\mathbf{U})\\ &+ \log f(\boldsymbol{\Omega}) + \log f(\boldsymbol{\eta})]. \quad (16) \end{aligned}$$

In the next section, an SAEM algorithm is developed to compute the MMAP estimate of $\boldsymbol{\zeta}$ in Equation (16), which is the main contribution of this study. Note that the SAEM algorithm includes a stochastic approximation step instead of the integral step in the EM algorithm (please see the next section for details); in other words, the SAEM algorithm does not need to marginalize the latent abilities, and the estimate of $\boldsymbol{\Omega}$ in the SAEM algorithm thus belongs to the maximum a posteriori (MAP) estimate. In addition, we use Newton–Raphson iteration to obtain the MAP estimate of the shape parameter $\boldsymbol{\eta}$, which was also adopted by Woods and Thissen (2006).

4. SAEM algorithm for the RC-3PNO model

In this section, an SAEM algorithm for the RC-3PNO model is developed to compute the MAP estimates for the shape parameters of the Ramsay curve and the item parameters. First, the relationship between the standard EM algorithm and

the SAEM algorithm is given. Second, a data-augmentation scheme for the 3PNO model is introduced, which means that the complete data likelihood formulation of the 3PNO model belongs to an exponential family form (Camilli and Geis, 2019; Geis, 2019). Sufficient statistics of the item parameters are also computed. Third, the estimation of the density curve of latent ability is depicted. Fourth, the estimation procedure of the SAEM algorithm is given for the RC-3PNO model.

4.1. Relationship between the EM algorithm and SAEM algorithm

The EM algorithm is briefly reviewed first. It is widely used in maximum likelihood or maximum a posteriori estimation for the incomplete data. The EM algorithm uses an expectation step (E-step) and a maximization step (M-step) to iteratively maximize the conditional expectation of the complete log-likelihood. In the E-step, the conditional expectation of the logarithmic complete data likelihood is adopted considering the observed data and the parameter values obtained in the previous step. In the M-step, the MAP estimates of the parameters are calculated based on the updated expectations in the E-step. The procedure alternates between E-step and M-step until convergence. In the case of exponential family distribution, the E-step and M-step can be simplified to update the expectation of the sufficient statistic and calculate the MAP estimate using the updated sufficient statistic, respectively.

However, in some cases, the EM algorithm is not applicable when either E-step or M-step is intractable or even cannot be performed in closed form. A possible solution for the complex computation of M-step is to replace the global optimization with a simpler conditional maximization chain, leading to the so-called ECM algorithm (Meng and Rubin, 1993). In IRT, the number of quadrature points in the numerical integral grows exponentially with the increasement of latent trait dimension. Therefore, the E-step will become very time-consuming as the latent trait dimension increases. Wei and Tanner (1990) proposed Monte Carlo EM (MCEM) to deal with this problem. The basic idea is to compute the expectation in the E-step by the Monte Carlo method. Geyer (1994) proved the convergence of the Monte Carlo maximum likelihood calculations, which provided the theoretical basis for MCEM. Delyon et al. (1999) proposed the SAEM algorithm as an alternative to the MCEM algorithm, which replaces the E-step of the EM algorithm with one iteration of the stochastic approximation procedure. Thereafter, the SAEM algorithm was widely used for its efficiency and convenience.

4.2. Data-augmentation scheme for the 3PNO model

The 3PNO model in Equation (1) can be rewritten as

$$P(U_{ij}=1|a_j,b_j,c_j,\theta_i)=c_j(1-\Phi(a_j\theta_i+b_j))+\Phi(a_j\theta_i+b_j), \quad (17)$$

and its form can be seen as a mixture of two Bernoulli distributions with the categorical probability $\Phi(a_j\theta_i+b_j)$. A dichotomous latent variable W_{ij} is defined, and $W_{ij}=1$ denotes examinee i knowing the answer of item j, and $W_{ij}=0$ otherwise. Because the 3PNO model does not contain a slipping parameter, examinees can answer the item correctly with probability 1 if they know the answer. The 3PNO model also includes a guessing parameter, and examinees can guess the correct answer with probability c_j if they do not know it. Thus, the following two equations hold:

$$P(U_{ij}|W_{ij}=1)=1^{U_{ij}}0^{1-U_{ij}}, \quad (18)$$

$$P(U_{ij}|W_{ij}=0)=c_j^{U_{ij}}(1-c_j)^{1-U_{ij}}. \quad (19)$$

A continuous latent variable Z_{ij} is then introduced:

$$W_{ij}=I(Z_{ij}>0), \quad (20)$$

where $I(\cdot)$ denotes the indicator function. The following conditional distribution holds:

$$Z_{ij}|\theta_i,a_j,b_j\sim N(a_j\theta_i+b_j,1). \quad (21)$$

The conditional probability of W_{ij} is

$$P(W_{ij}=1|\theta_i,a_j,b_j)=P(Z_{ij}>0|\theta_i,a_j,b_j)=\Phi(a_j\theta_i+b_j). \quad (22)$$

According to the total probability formula, the marginal probability of $U_{ij}=1$ is

$$P(U_{ij}=1|\boldsymbol{\Omega_j})=\Phi(a_j\theta_i+b_j)+[1-\Phi(a_j\theta_i+b_j)]\times c_j, \quad (23)$$

which is exactly the 3PNO model in Equation (1).

4.2.1. Complete data likelihood of the 3PNO model

According to Equations (18)–(21), the joint distribution of $(\boldsymbol{U},\boldsymbol{W},\boldsymbol{Z})$ is

$$P(U_{ij},W_{ij},Z_{ij}|\theta_i,\boldsymbol{\Omega_j})=P(U_{ij}|W_{ij},\theta_i,\boldsymbol{\Omega_j})\cdot P(W_{ij},Z_{ij}|\theta_i,\boldsymbol{\Omega_j}), \quad (24)$$

where

$$P(U_{ij}|W_{ij},\theta_i,\boldsymbol{\Omega_j})=[1^{W_{ij}}c_j^{1-W_{ij}}]^{U_{ij}}\cdot[0^{W_{ij}}(1-c_j)^{1-W_{ij}}]^{1-U_{ij}}, \quad (25)$$

$$P(W_{ij}, Z_{ij}|\theta_i, \boldsymbol{\Omega_j}) = [I(W_{ij}=1)I(Z_{ij}>0) + I(W_{ij}=0)I(Z_{ij}<0)] \times \phi(z_{ij} - a_j\theta_i - b_j). \quad (26)$$

Let $\mathbf{z}$ and $\mathbf{w}$ denote the observations of augmented variables $\boldsymbol{Z}$ and $\boldsymbol{W}$, respectively. $\boldsymbol{\Theta}$ denotes the data set of Θ sampled in the S1-step of SAEM (more details are given in a later subsection). Note that $\boldsymbol{\Theta}$ is also the augmented data sets in SAEM, and the estimate of $\boldsymbol{\Omega}$ is an MAP estimate. The augmented data sets are $\Psi = (\boldsymbol{\Theta}, \mathbf{w}, \mathbf{z})$, and the observed responses are $\mathbf{U}$. The complete data likelihood of Ω can then be expressed as

$$L(\boldsymbol{\Omega}|\mathbf{U}, \Psi) = \prod_{i=1}^{N}\prod_{j=1}^{J} P(U_{ij}, W_{ij}, Z_{ij}|\theta_i, \boldsymbol{\Omega_j})g(\theta_i), \quad (27)$$

According to Equations (24)–(26), we have

$$L(\boldsymbol{\Omega}|\mathbf{U}, \Psi) \propto \prod_{j=1}^{J}\left\{c_j^{\sum_{i=1}^{N}(1-w_{ij})u_{ij}}(1-c_j)^{\sum_{i=1}^{N}(1-w_{ij})(1-u_{ij})}\right\} \times \prod_{i=1}^{N}\prod_{j=1}^{J}\left\{\phi(z_{ij} - a_j\theta_i - b_j)g(\theta_i)\right\}. \quad (28)$$

It can be proved that $L(\boldsymbol{\Omega}|\mathbf{U}, \Psi)$ has a form of an exponential family distribution.

Two advantages need to be mentioned here. First, after introducing the augmented latent variables, the complete data likelihood has a form of exponential family distribution. Second, given complete data, the MAP estimates of item parameter $\boldsymbol{\Omega}$ can be expressed as several functions that are only concerned with sufficient statistics. In this case, we can directly implement the computation of the MAP estimates of the item parameters that only need to update the sufficient statistics, which greatly reduces the computational complexity and improves the computational efficiency. In addition, the SAEM algorithm converges to the local maximum (Delyon et al., 1999). Note that, due to the data-augmentation scheme, Equation (16) is not the objective function to be optimized in the SAEM algorithm. Instead, $\boldsymbol{\Theta}$ is viewed as augmented data in the SAEM algorithm and is thus included in the objective function to be optimized [see Equation (28)]. The augmented data of $\boldsymbol{\Theta}$ can be used in the estimation of $\boldsymbol{\eta}$ (this is elaborated later).

4.3. Estimation of density curve of latent ability

The estimation of the non-normal ability density curve relies on the computation of the shape parameter $\boldsymbol{\eta}$, which involves an optimization algorithm using either Newton–Raphson iterations (Woods and Thissen, 2006) or an MHRM algorithm (Monroe and Cai, 2014). Once the estimates of $\boldsymbol{\eta}$ have been obtained, these estimates can be used in Equation (2) to calculate $g(\theta_i|\boldsymbol{\eta})$ for a particular examinee or to construct the entire Ramsay-curve density. In this study, the Newton–Raphson iteration method was used to estimate $\boldsymbol{\eta}$, and the log-likelihood of the Ramsay curve is the objective function to be optimized. Its form is

$$\log L(\boldsymbol{\eta}|\Theta) = \sum_{i=1}^{N}\log(g(\theta_i|\boldsymbol{\eta})). \quad (29)$$

In practice, there may be some regions of the latent trait scale over which little or no information about the RC parameters $\boldsymbol{\eta}$ is available. As a result, the corresponding spline coefficients may become empirically underidentified (Woods and Thissen, 2006; Monroe and Cai, 2014). When this happens, the estimation of the entire set of coefficients will fail. To prevent such an estimation failure, a diffuse prior density is often assumed on $\boldsymbol{\eta}$ (Woods and Thissen, 2006; Monroe and Cai, 2014). The Bayesian MAP estimation can be used, and the Ramsay-curve posterior (RCP) density is then the product of the Ramsay-curve likelihood and an m-variate normal prior (Woods and Thissen, 2006; Monroe and Cai, 2014), where m is the number of coefficients. Since the normalization constant of the posterior is omitted, the logarithm RCP of $\boldsymbol{\eta}$ is given by

$$l_R(\boldsymbol{\eta}|\Theta) \propto \log L(\boldsymbol{\eta}|\Theta) - \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\mu_\eta})'\boldsymbol{\Sigma}_{\boldsymbol{\eta}}^{-1}(\boldsymbol{\eta} - \boldsymbol{\mu_\eta}), \quad (30)$$

where $\boldsymbol{\mu_\eta}$ and $\boldsymbol{\Sigma_\eta}$ are the prior mean vector and the covariance matrix of $\boldsymbol{\eta}$, respectively.

4.4. SAEM algorithm for estimating $\boldsymbol{\Omega}$ and $\boldsymbol{\eta}$

The SAEM algorithm was first proposed by Delyon et al. (1999), and it replaces the integral calculation with a stochastic approximation in the E-step, which significantly improves computational efficiency, especially for exponential family distributions. We assume that: the iteration has updated to step k; $\boldsymbol{\Omega}^{(k)} = (\boldsymbol{a}^{(k)}, \boldsymbol{b}^{(k)}, \boldsymbol{c}^{(k)})$ and $\boldsymbol{\eta}^{(k)}$ are the current estimates of the item parameters and the shape parameter of the Ramsay curve, respectively; and $\Psi^{(k)} = (\boldsymbol{\Theta}^{(k)}, \mathbf{w}^{(k)}, \mathbf{z}^{(k)})$ is the current augmented data. The detailed estimation steps of the SAEM algorithm incorporating the Ramsay curve (the RC-SAEM algorithm) at step $k+1$ are given as follows.

Simulation step (S-step). Sample $\boldsymbol{\Psi^{(k+1)}}$:

S1-step: Sample $\theta_i^{(k+1)}$

$$P(\theta_i^{(k+1)} = x_q|\mathbf{U}_i, \Omega^{(k)}, \boldsymbol{\eta}^{(k)}) = \frac{\prod_{j=1}^{J}\left\{P_j(x_q|\Omega^{(k)})^{u_{ij}}[1-P_j(x_q|\Omega^{(k)})]^{1-u_{ij}}g(x_q|\boldsymbol{\eta}^{(k)})\right\}}{\sum_{q=1}^{Q}\prod_{j=1}^{J}\left\{P_j(x_q|\Omega^{(k)})^{u_{ij}}[1-P_j(x_q|\Omega^{(k)})]^{1-u_{ij}}g(x_q|\boldsymbol{\eta}^{(k)})\right\}}, \quad (31)$$

S2-step: Sample $w_{ij}^{(k+1)}$

$$w_{ij}^{(k+1)}|\Omega^{(k)}, \mathbf{U}, \mathbf{\Theta}^{(k+1)} \sim \begin{cases} \text{Bernoulli}\left(\frac{\Phi(a_j^{(k)}\theta_i^{(k+1)}+b_j^{(k)})}{c_j^{(k)}+(1-c_j^{(k)})\Phi(a_j^{(k)}\theta_i^{(k+1)}+b_j^{(k)})}\right), & u_{ij}=1 \\ \text{Bernoulli}(0). & u_{ij}=0 \end{cases} \quad (32)$$

S3-step: Sample $z_{ij}^{(k+1)}$

$$z_{ij}^{(k+1)}|\Omega^{(k)}, \mathbf{\Theta}^{(k)}, \mathbf{w}^{(k+1)} \sim N(a_j^k\theta_i^{(k+1)} + b_j^{(k)}, 1)[I(z_{ij}^{(k+1)} > 0)w_{ij}^{(k+1)} + I(z_{ij}^{(k+1)} \leq 0)(1 - w_{ij}^{(k+1)})], \quad (33)$$

Stochastic approximation step (SA-step). Update sufficient statistics $S_j^{(k+1)}$ ($j = 1, \cdots, J$):

Based on the factorization theorem, the sufficient statistics of the item parameters $\mathbf{\Omega}$ are

$$S_j(\mathbf{U}, \Psi^{(k+1)}) = (S_{j1}^{(k+1)}, S_{j2}^{(k+1)}, S_{j3}^{(k+1)}, S_{j4}^{(k+1)}) \quad (j = 1, 2, \cdots, J), \quad (34)$$

and

$$S_{j1}^{(k+1)} = \mathbf{S}^{*\prime(k+1)}\mathbf{S}^{*(k+1)}, \quad (35)$$

$$S_{j2}^{(k+1)} = \mathbf{S}^{*\prime(k+1)}\mathbf{z}_j^{(k+1)}, \quad (36)$$

$$S_{j3}^{(k+1)} = \sum_{i=1}^{N}(1 - w_{ij}^{(k+1)}), \quad (37)$$

$$S_{j4}^{(k+1)} = \sum_{i=1}^{N}(1 - w_{ij}^{(k+1)})u_{ij}, \quad (38)$$

where $\mathbf{S}^{*(k+1)} = (\mathbf{\Theta}^{(k+1)}, \mathbf{1}_N)$, $\mathbf{1}_N$ is a unit column vector with dimension N, and the vector $\mathbf{z}_j^{(k+1)}$ is the jth column of augmented data set $\mathbf{z}^{(k+1)}$ in the $k+1$th iteration.

Thus, the stochastic approximation step is:

$$S_j^{(k+1)} = S_j^{(k)} + \gamma_k[S_j(\mathbf{U}, \mathbf{\Psi}^{(k+1)}) - S_j^{(k)}], \quad (39)$$

where $\Psi^{(k+1)} = (\mathbf{\Theta}^{(k+1)}, \mathbf{w}^{(k+1)}, \mathbf{z}^{(k+1)})$ is the augmented data sets that are simulated from the S-step, and $\{\gamma_k, k = 1, 2, \cdots\}$ is a decreasing sequence of gain constants, as defined by Robbins and Monro (1951), which satisfies

$$\gamma_k \epsilon (0, 1], \sum_{k=1}^{\infty}\gamma_k = \infty, \sum_{k=1}^{\infty}\gamma_k^2 < \infty.$$

Maximization step (M-step). Update $\mathbf{\Omega}^{(k+1)}$ and $\boldsymbol{\eta}^{(k+1)}$:

M1-step. Update $\mathbf{\Omega}^{(k+1)}$

Based on $L(\mathbf{\Omega}|\mathbf{U}, \Psi)$ and the prior distributions in Equations (10) and (11), the posterior distributions of $\mathbf{\Omega_j} = (a_j, b_j, c_j)$ are

$$(a_j, b_j)'|\mathbf{U}, \Psi^{(k+1)} \sim N_2(\boldsymbol{\mu}^{*(k+1)}, \mathbf{\Sigma}^{*(k+1)})I(a_j > 0) \quad (40)$$

and

$$c_j|\mathbf{U}, \Psi^{(k+1)} \sim \text{Beta}(\alpha^{*(k+1)}, \beta^{*(k+1)}), \quad (41)$$

where

$$\boldsymbol{\mu}^{*(k+1)} = (S_{j1}^{(k+1)} + \mathbf{\Sigma})^{-1}(S_{j2}^{(k+1)} + \mathbf{\Sigma}^{-1}\boldsymbol{\mu}),$$
$$\mathbf{\Sigma}^{*(k+1)} = (S_{j1}^{(k+1)} + \mathbf{\Sigma}^{-1})^{-1},$$
$$\alpha^{*(k+1)} = \alpha + S_{j4}^{(k+1)},$$
$$\beta^{*(k+1)} = \beta + S_{j3}^{(k+1)} - S_{j4}^{(k+1)}.$$

$\boldsymbol{\mu}, \mathbf{\Sigma}, \alpha, \beta$ are the hyper-parameters in prior distributions of $\boldsymbol{a, b, c}$, please refer to Equations (10) and (11).

Thus, the MAP estimates of a_j, b_j, and c_j for the $k+1$th iteration are

$$\hat{a}_j^{(k+1)} = \boldsymbol{\mu}^{*(k+1)}[1]I(\boldsymbol{\mu}^{*(k+1)}[1] > 0) + \delta \times I(\boldsymbol{\mu}^{*(k+1)}[1] \leq 0), \quad (42)$$

$$\hat{b}_j^{(k+1)} = \boldsymbol{\mu}^{*(k+1)}[2], \quad (43)$$

$$\hat{c}_j^{(k+1)} = \frac{\alpha^{*(k+1)} + S_{j4}^{(k+1)} - 1}{\alpha^{*(k+1)} + \beta^{*(k+1)} + S_{j3}^{(k+1)} - 2}, \quad (44)$$

where δ is a tiny positive number to satisfy $a_j > 0$, $\boldsymbol{\mu}^{*(k+1)}[1]$ denotes the first element of $\boldsymbol{\mu}^{*(k+1)}$ in the $k+1$th iteration, and $\boldsymbol{\mu}^{*(k+1)}[2]$ denotes the second element of $\boldsymbol{\mu}^{*(k+1)}$ in the $k+1$th iteration.

M2-step. Update $\boldsymbol{\eta}^{(k+1)}$:

Update $\boldsymbol{\eta}^{(k+1)}$ according to the Newton–Raphson iteration, which satisfies

$$\boldsymbol{\eta}^{(t+1)} = \boldsymbol{\eta}^{(t)} - \left(\frac{\partial^2 l_R}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'}\right)^{-1}\frac{\partial l_R}{\partial\boldsymbol{\eta}}, \quad (45)$$

where l_R is the logarithm RCP of $\boldsymbol{\eta}$ in Equation (30). Note that t is the number of iterations in the process of implementing the Newton-Raphson iteration algorithm. Let $\boldsymbol{\eta}^{(k+1)} = \boldsymbol{\eta}^{(t+1)}$ when Newton-Raphson iteration algorithm reaches convergence after executing $t+1$ iterations of inner loop and continue the computations of SAEM algorithm.

Repeat the S-step, SA-step, and M-step until the convergence criteria are satisfied. Here, the SAEM algorithm is considered

to have converged when the maximum absolute difference of the MAP estimates between two adjacent iterations (i.e., max $|\dot{\boldsymbol{\zeta}}^{(k)} - \dot{\boldsymbol{\zeta}}^{(k+1)}|$) is less than 10^{-4} or the maximum number of iterations (selected as 2500) is reached.

Note that the augmented data sets in the S-step can be simulated m_k sets in the original SAEM algorithm (Delyon et al., 1999), that is, $\Psi_p^{(k+1)} = (\Theta_p^{(k+1)}, \mathbf{w}_p^{(k+1)}, \mathbf{z}_p^{(k+1)})$ $(p = 1, 2, \cdots, m_k)$. In this case, $S_j(\mathbf{U}, \boldsymbol{\Psi}^{(k+1)})$ in Equation (39) can be replaced by the average value of the m_k updated sufficient statistics computed from these augmented data sets, that is, $\frac{\sum_{p=1}^{m_k} S_j(\mathbf{U}, \Psi_p^{(k+1)})}{m_k}$. According to previous studies of the SAEM algorithm, the number of simulations $m_k = 1$ is suggested to be set for all the iterations (Delyon et al., 1999; Kuhn and Lavielle, 2004), which makes the M-step straightforward to implement and increases the computational efficiency. In most cases, the increasing of m_k will not improve the accuracy of the algorithm. For the Robbins–Monro gain coefficient, let $\gamma_k = (\frac{1}{k})^{\alpha}$, where $\alpha \geqslant 0$. A larger step size (that is, $\alpha = 0$) can accelerate the rate of convergence, but this will result in inflation of the Monte Carlo error introduced when approximating the integral by the average of a set of simulations in the SA-step (Jank, 2006). A smaller step size (that is, $\alpha = 1$) may allow the sequence of estimates to approach the neighborhood of the solution with a small Monte Carlo error, but it will also slow down the convergence rate (Jank, 2006; Geis, 2019). In this work, the step size γ_k was chosen to be 1 in the first 1,000 iterations to ensure that enough steps were used when quickly approaching the neighborhood of the solution, but this also inflates the Monte Carlo error at the same time (Jank, 2006). Then, we let $\gamma_k = \frac{1}{k-1000}$ when $k > 1,000$ to rapidly reduce the Monte Carlo error of the estimates, though this slows down the convergence rate (Gu and Zhu, 2001; Kuhn and Lavielle, 2004; Jank, 2006; Geis, 2019).

4.5. Evaluation criteria

In fact, researchers have conducted in-depth studies on model selection methods based on evidence function (log-likelihood function), such as likelihood ratio test (LR-test) and chi-square difference test. However, these methods actually have some drawbacks and limitations. As depicted in Woods (2006), LR-test is not an ideal evaluation criterion for RC-IRT for two reasons. One is its tendency to select large models. It tends to select the largest model that is significantly better than the true model. Another limitation is that, like all chi-square difference tests, it requires the larger model to fit the data in an absolute sense, which is difficult to establish. A chi-square test of absolute fit is usually not appropriate in IRT because the number of possible response patterns is large, and the probability of any one of the patterns is small; thus, statistics like Pearson's are not chi-square distributed (Maydeu-Olivares and Cai, 2006). For these reasons, LR tests alone should not be relied upon for model selection. Thus, Woods (2006) considered the following three model selection criteria.

Three model selection criteria—Akaike's information criterion (AIC; Akaike, 1973), Bayesian information criterion (BIC; Schwarz, 1978), and Hannan–Quinn information criterion (HQIC; Hannan, 1987; Woods, 2007, 2008)—are considered:

$$\mathrm{AIC} = -2\log L + \log(N), \tag{46}$$

$$\mathrm{BIC} = -2\log L + n\log(N), \tag{47}$$

$$\mathrm{HQIC} = -2\log L + 2n\log(\log(N)), \tag{48}$$

where $\log L$ is the log-likelihood of all the parameters Ω and $\boldsymbol{\eta}$, n is the number of parameters, and N is the sample size.

As the number of *konts* and *degree* increases, the number of free parameters increases and the goodness of fitting is improved. AIC (BIC and HQIC) encourages the goodness of data fitting (information provided by the evidence function) but tries to avoid overfitting (prevent the cases of too many free parameters). The purpose of information criterion is to find the balance between model fit and model complexity. The preferred model should be the one with the lowest AIC (BIC and HQIC) values.

To evaluate the accuracy of the item parameter recoveries, the bias and root-mean-square error (RMSE) are calculated. Supposing R is the number of replications, the bias of parameter ω is

$$\mathrm{Bias} = \frac{1}{R}\sum_{r=1}^{R}(\hat{\omega}_r - \omega), \tag{49}$$

and the RMSE of ω is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\omega}_r - \omega)^2}, \tag{50}$$

where $\hat{\omega}_r$ is the parameter estimate at the r*th* replication and ω is the true value of the parameter.

Since the scales of the true RC parameters are complicated and difficult to handle, the bias and RMSE are less appropriate measures of recovery accuracy. Instead, the integrated square error (ISE),

$$\mathrm{ISE}(\hat{g}) = \int \left[g(\theta|\hat{\boldsymbol{\eta}}) - g(\theta|\boldsymbol{\eta})\right]^2 d\theta, \tag{51}$$

is used to measure the discrepancy between the true and estimated RCs, as used by Woods and Lin (2009) and Monroe and Cai (2014). The ISE was multiplied by 1,000 to facilitate comparison. The values of ISE were computed across all replications.

TABLE 1 Model selection results under different combinations of *knots* and *degree* when θ is normal, skewed, or bimodal.

	Normal			Skewed			Bimodal		
	AIC	BIC	HQIC	AIC	BIC	HQIC	AIC	BIC	HQIC
2knots-degree3	34121.1	34582.5	34296.5	33954.3	34415.6	34129.6	32576.4	33037.7	32751.7
2knots-degree4	34099.8	34566.1	34277.0	33968.3	34434.5	34145.5	32639.4	33105.6	32816.6
3knots-degree3	34061.1	34531.7	34238.3	33815.8	34282.0	33993.0	32531.3	32997.6	32708.5
3knots-degree4	34089.8	34561.0	34268.9	33913.9	34385.0	34092.9	32499.8	32970.9	32678.8
4knots-degree3	34117.4	34588.5	34296.4	33762.6	34233.7	33941.7	32381.9	32853.0	32561.0
4knots-degree4	**34055.6**	**34527.4**	**34236.5**	33745.7	34221.8	33926.7	32456.2	32932.3	32637.2
5knots-degree3	34080.2	34556.3	34261.1	33715.4	34191.8	33896.4	**32214.9**	**32690.9**	**32395.8**
5knots-degree4	34119.7	34600.6	34302.5	33786.8	34267.8	33969.6	32268.4	32749.3	32451.1
6knots-degree3	34113.2	34594.1	34296.0	**33710.9**	**34191.5**	**33893.7**	32241.5	32722.5	32424.3
6knots-degree4	34068.5	34554.4	34253.2	33741.0	34226.9	33925.7	32317.9	32803.7	32502.5

The boldfaced values indicate the smallest AIC, BIC, and HQIC values in each column.

5. Simulation studies

5.1. Simulation study 1

The first simulation study was performed to select the optimal numbers of *knots* and the *degree* of the B-spline functions for the RC-3PNO model based on three model selection criteria—the AIC, BIC, and HQIC—as well as to show the item parameter recoveries when the true ability density is normal, skewed, or bimodal.

5.1.1. Design

The true latent-ability densities were represented by rectangular quadrature points, ranging from -6 to 6 in steps of 0.1. For the skewed and bimodal cases, the true ability density was generated by mixing two normal densities, that is, $p_1 N(\mu_1, \sigma_1^2) + p_2 N(\mu_2, \sigma_2^2)$, in which $p_1 + p_2 = 1$. For the skewed density, the generating parameters were: $\mu_1 = -2.7$, ${\sigma_1}^2 = 0.2$, $\mu_2 = 1.1$, and ${\sigma_2}^2 = 1.1$. The skewness and kurtosis of θ were 2.46 and 8.45, respectively. For the bimodal density, the generating parameters were: $\mu_1 = -2$, ${\sigma_1}^2 = 0.25$, $\mu_2 = 2.5$, and ${\sigma_2}^2 = 0.5$. In this case, the skewness and kurtosis of θ were 1.45 and 4.21, respectively. The true item parameters were set to be consistent with common practice in IRT. The discrimination parameters a were sampled from $U(1, 2.5)$, the intercept parameters b were simulated from $N(0, 1)$, and the guessing parameters c were generated from Beta(5, 17).

For each of the three true latent-ability densities, 10 models with different combinations of *knots* and *degree* were fitted to the generated data. The *degree* values of the B-spline functions were either 3 or 4, and the number of *knots* was chosen to be between 2 and 6; these are the typical choices when estimating $\boldsymbol{\eta}$ in RC-IRT (Woods and Thissen, 2006; Monroe and Cai, 2014). The sample size was set to be 1,000, and the test length was fixed at 30. Therefore, there were 30 simulation conditions, and each simulation condition was conducted 100 times.

5.1.2. Results

Table 1 presents the model selection results when the true latent-ability densities are normal, skewed, and bimodal. The bold-faced values indicate the smallest AIC, BIC, and HQIC values in each column. In addition, under each of the three shapes of ability density, the AIC, BIC, and HQIC are consistent to choose one common combination of the *knots* and *degree*.

When the true ability density is normal, all three model selection criteria result in the RC-3PNO model with 4 *knots* and a *degree* of 4 (denoted as the 4-4 RC-3PNO model) being chosen as the best-fitting model. However, the differences in the model selection results under each condition are slight overall. When the true ability density is skewed, the values of AIC, BIC, and HQIC are the smallest for the RC-3PNO model with 6 *knots* and a *degree* of 3 (denoted as the 6-3 RC-3PNO model); the best model chosen in the bimodal case is the RC-3PNO model with 5 *knots* and a *degree* of 3 (denoted as the 5-3 RC-3PNO model). In addition, in the skewed and bimodal cases, the values of AIC, BIC, and HQIC have obvious discrepancies across different combinations of *knots* and *degrees* for the RC-3PNO model. Specifically, for the skewed case, the model selection results for the RC-3PNO model with *knots* and *degree* combinations 2-3, 2-4, and 3-4 show relatively large values compared with the best-fitting model, i.e., the 6-3 RC-3PNO model, with discrepancies over 200. The reason for this may be that a *knots* value of 2 or 3 is not sufficient to describe a skewed ability density. The RC-3PNO models with *knots* and *degree* combinations of 4-3, 4-4, 5-3, 5-4, 6-3, and 6-4 result in very little differences in the AIC, BIC, and HQIC values when the true ability density

TABLE 2 Bias and RMSE of item parameter estimates under different combinations of knots and degree in normal case.

	a		*b*		*c*	
	Bias	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
2knots-degree3	0.101	0.354	−0.023	0.366	−0.006	0.069
2knots-degree4	0.091	0.329	−0.004	0.357	−0.009	0.070
3knots-degree3	0.086	0.299	−0.006	0.333	−0.008	0.068
3knots-degree4	0.084	0.325	−0.012	0.342	−0.007	0.071
4knots-degree3	0.081	0.310	−0.025	0.346	−0.025	0.068
4knots-degree4	0.079	0.308	0.005	0.334	−0.011	0.067
5knots-degree3	0.094	0.349	−0.032	0.367	−0.009	0.067
5knots-degree4	0.081	0.321	−0.026	0.354	−0.008	0.068
6knots-degree3	0.083	0.299	−0.016	0.335	−0.009	0.067
6knots-degree4	0.078	0.298	−0.014	0.337	−0.008	0.068

TABLE 3 Bias and RMSE of item parameter estimates under different combinations of knots and degree in skewed case.

	a		*b*		*c*	
	Bias	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
2knots-degree3	0.046	0.276	−0.065	0.352	−0.007	0.080
2knots-degree4	0.039	0.261	−0.004	0.330	−0.015	0.078
3knots-degree3	0.041	0.251	0.034	0.329	−0.018	0.082
3knots-degree4	0.042	0.268	−0.046	0.339	−0.008	0.079
4knots-degree3	0.047	0.262	−0.028	0.363	−0.008	0.082
4knots-degree4	0.056	0.282	−0.024	0.370	−0.008	0.083
5knots-degree3	0.043	0.266	−0.017	0.381	−0.010	0.082
5knots-degree4	0.052	0.280	−0.042	0.367	−0.007	0.082
6knots-degree3	0.055	0.285	−0.043	0.377	−0.007	0.080
6knots-degree4	0.051	0.276	−0.060	0.393	−0.005	0.084

is skewed. For the bimodal case, the discrepancies in model selection results between the RC-3PNO models with *knots* and *degree* combinations of 2-3, 2-4, 3-3, 3-4, and 4-4 and the best-fitting model, i.e., the 5-3 RC-3PNO model, are greater than 200. The differences in model selection results between the models with the combinations 5-4 and 6-3 and the best-fitting model (5-3) are extremely small.

Tables 2–4 show the item parameter estimation results. There are no distinct differences in the bias and RMSE values of item parameters *a*, *b*, and *c* for the fitted models across all the conditions; however, some subtle variations still exist. Thus, the choices of *knots* and *degree* in this simulation study had no noticeable influences on the bias and RMSE values of the item parameters. The standard errors of item parameters are presented in Figures 1–3. In the majority of cases, the standard errors of item parameters *a* and *b* are below 0.08, and the standard errors of parameter *c* are below 0.02, which are within the tolerable ranges. It indicates that the RC-SAEM algorithm performs well in estimation stability.

5.2. Simulation study 2

This simulation study was conducted to compare the performance of the proposed RC-SAEM algorithm, the original SAEM, and the MCMC algorithm in estimating the item parameters of the 3PNO model.

5.2.1. Design

The shapes of the true ability density, $g(\theta)$, were set to be normal, skewed, and bimodal distributions (Woods and Thissen, 2006; Monroe and Cai, 2014). For the skewed case, the skewness and kurtosis of θ were 1.72 and 9.16, respectively; in the bimodal case, the skewness and kurtosis of θ were 0.95 and 2.74, respectively. The numbers of examinees were set to be 500, 1,000, and 2,000 to represent small, medium, and large sample sizes, respectively. The test lengths were set as 15 and 30. Therefore, a total of 18 simulation conditions

TABLE 4 Bias and RMSE of item parameter estimates under different combinations of knots and degree in bimodal case.

	a		*b*		*c*	
	Bias	**RMSE**	**Bias**	**RMSE**	**Bias**	**RMSE**
2knots-degree3	0.078	0.304	−0.039	0.369	−0.011	0.080
2knots-degree4	0.084	0.311	−0.016	0.367	−0.015	0.081
3knots-degree3	0.095	0.324	0.048	0.363	−0.020	0.084
3knots-degree4	0.075	0.299	−0.016	0.374	−0.012	0.079
4knots-degree3	0.087	0.324	−0.039	0.429	−0.008	0.087
4knots-degree4	0.087	0.319	0.038	0.367	−0.017	0.084
5knots-degree3	0.071	0.303	−0.029	0.420	−0.011	0.087
5knots-degree4	0.076	0.304	−0.024	0.423	−0.008	0.086
6knots-degree3	0.073	0.310	−0.047	0.489	−0.010	0.087
6knots-degree4	0.082	0.310	−0.012	0.410	−0.013	0.084

were manipulated. Each simulation condition was replicated 100 times.

The data-generating model and the fitted model were the same, that is, the 6-3 RC-3PNO model (6 *knots* with a *degree* of 3). As noted in the description of the model, 121 discrete points [i.e., x_q ($q = 1, 2, \cdots, Q$), and $Q = 121$] from −6 to 6 in steps of 0.1 were used to describe the true ability density. After the true value of $\boldsymbol{\eta}$ was selected, the true ability of θ_i was manipulated according to grid sampling, similar to the S1 step of the RC-SAEM algorithm. That is, first, a true $\boldsymbol{\eta}$ value was chosen corresponding to the given shape of $g(x_q|\boldsymbol{\eta})$. Then, the probabilities of $\theta_i = x_q$ in the grid sampling were set to $g(x_q|\boldsymbol{\eta})$ ($q = 1, 2, \cdots, Q$). The values of θ_i ($i = 1, 2, \cdots, n$) were standardized to have a mean of 0 and a standard deviation of 1. The values after standardization were chosen to be true values of Θ. The true item parameters were the same as those in simulation study 1.

To avoid the effects of the choice of prior distribution on the estimation results, the priors for *a* and *b* were chosen to be non-informative priors. The prior for the *c* parameter was set to be Beta(5, 17) (the mean is $\frac{5}{5+17} = 0.227$), which is consistent with the common prior choice in IRT (Harwell and Baker, 1991; Béguin and Glas, 2001) because the nominal guessing probability is around 0.25 for multiple-choice items with four options. These priors for the item parameters were adopted in the SAEM, RC-SAEM, and MCMC algorithms.

5.2.2. Results

Tables 5, 6 show the bias and RMSE values of the item parameter estimates under different simulation conditions. For simplicity, the average values of the bias and RMSE across *J* items are presented. For the normal distribution of the ability density, the bias and RMSE values of parameters *a* and *b* from the RC-SAEM algorithm are smaller than those from the SAEM algorithm. The bias values of the *a* parameter from the MCMC algorithm are larger than those from the other two algorithms under all of the three sample sizes. We can see that the estimation results of the MCMC algorithm are not very satisfactory when the sample size is 500 and the test length is 15; this indicates that a sample size of 500 is not large enough for precise estimation of the 3PNO model using MCMC with non-informative priors on *a* and *b*. The poor performance of the MCMC algorithm for the normal distribution case may be due to the choice of non-informative priors on the parameters *a* and *b*. Therefore, the RC-SAEM algorithm performs best when the true θ density is normal. In addition, the RMSEs of parameters *a* and *b* show an approximately decreasing trend as the sample size increases.

In the cases of the skewed and bimodal distributions, the bias and RMSE values of *a* and *b* from the RC-SAEM algorithm are noticeably lower than those from the SAEM and MCMC algorithms, indicating that the proposed RC-SAEM algorithm is effective for skewed and bimodal densities. It is worth noting that the bias values of *a* from the MCMC algorithm are larger than those from the RC-SAEM and SAEM algorithms, while the corresponding RMSEs are smaller than those from the SAEM algorithm in a few conditions, which demonstrates that the estimation of *a* parameter under the MCMC algorithm in skewed and bimodal cases is less accurate than the other two algorithms. Although the RMSEs of *a* parameter from the MCMC algorithm are slightly smaller than those from the RC-SAEM algorithm in bimodal cases when the test length is 15 in Table 6, the bias of *a* under RC-SAEM still has an obvious advantage over that of MCMC.

The advantage of RC-SAEM is most evident when considering the bias and RMSE values of parameter *b* when the true ability density is skewed or bimodal. Specifically, for the skewed ability density, the bias and RMSE values of the *b* parameter from the RC-SAEM algorithm are markedly less than those from the SAEM and MCMC algorithms. In addition,

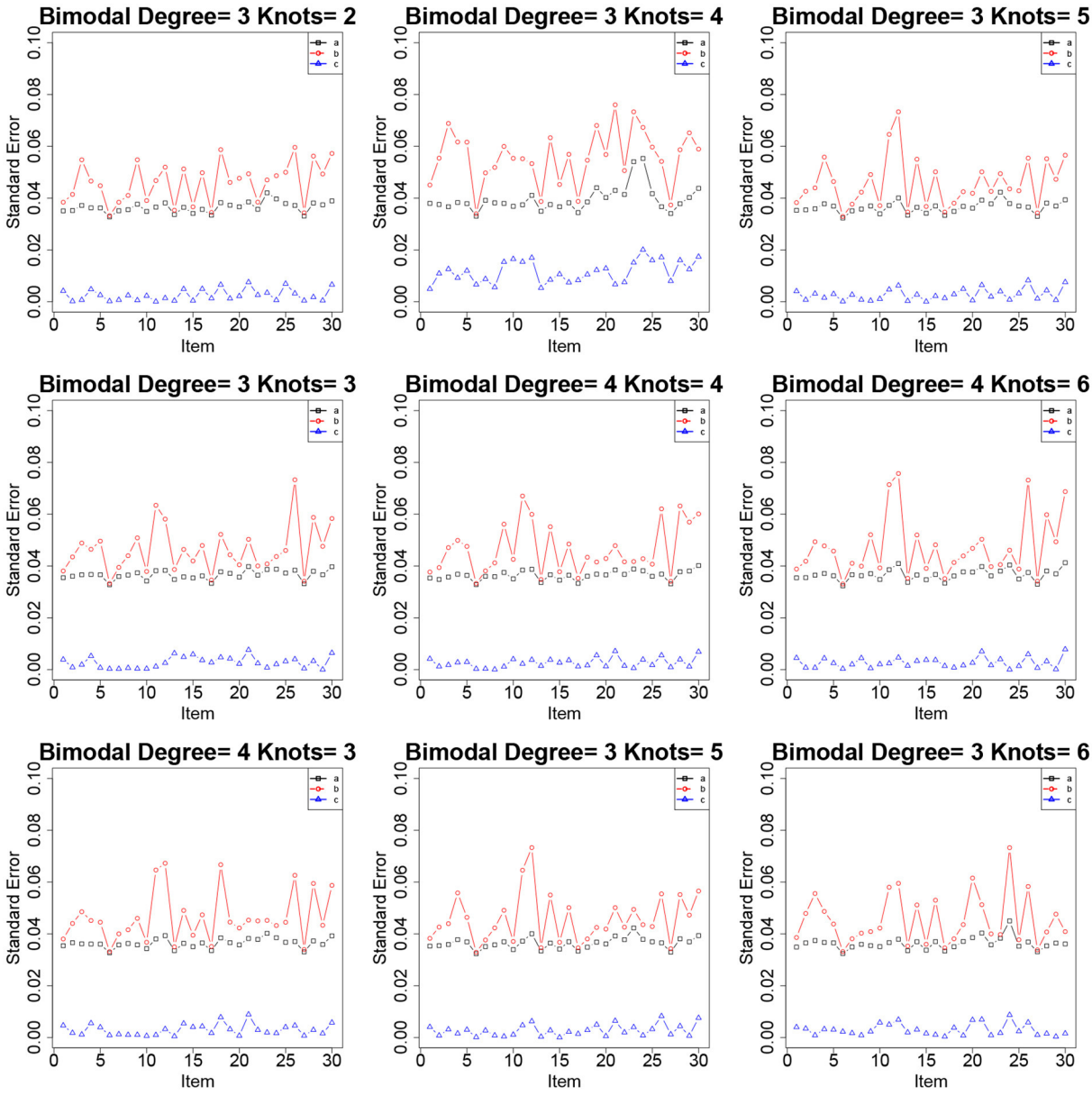


FIGURE 1
The standard errors of item parameters using the RC-SAEM algorithm when the latent trait is bimodal in simulation study 1.

for the bimodal latent-ability density, although the RMSE of *b* has no marked differences from the SAEM and RC-SAEM algorithms, the bias of *b* from the RC-SAEM algorithm is still lower than those from the other two. As can be seen from Table 5, in the case of the skewed ability density with a sample size 1,000, the SAEM and MCMC algorithms show obvious biased values on *b* with the absolute values over 0.3; in contrast, the RC-SAEM algorithm shows precise estimates of *b*, with a bias of 0.001 and an RMSE of 0.187. In general, the proposed RC-SAEM algorithm has a distinct advantage over the SAEM and MCMC algorithms in terms of the biases of *a* and *b*. The differences in the bias and RMSE of *c* from the RC-SAEM and SAEM algorithms are very small. The RMSE of *c* from the MCMC algorithm is slightly larger than those from the other two algorithms in the skewed and bimodal cases.

Figures 4, 5 show the true and estimated latent-ability density curves when the true ability densities are normal, skewed, or bimodal. The true and estimated latent-ability density

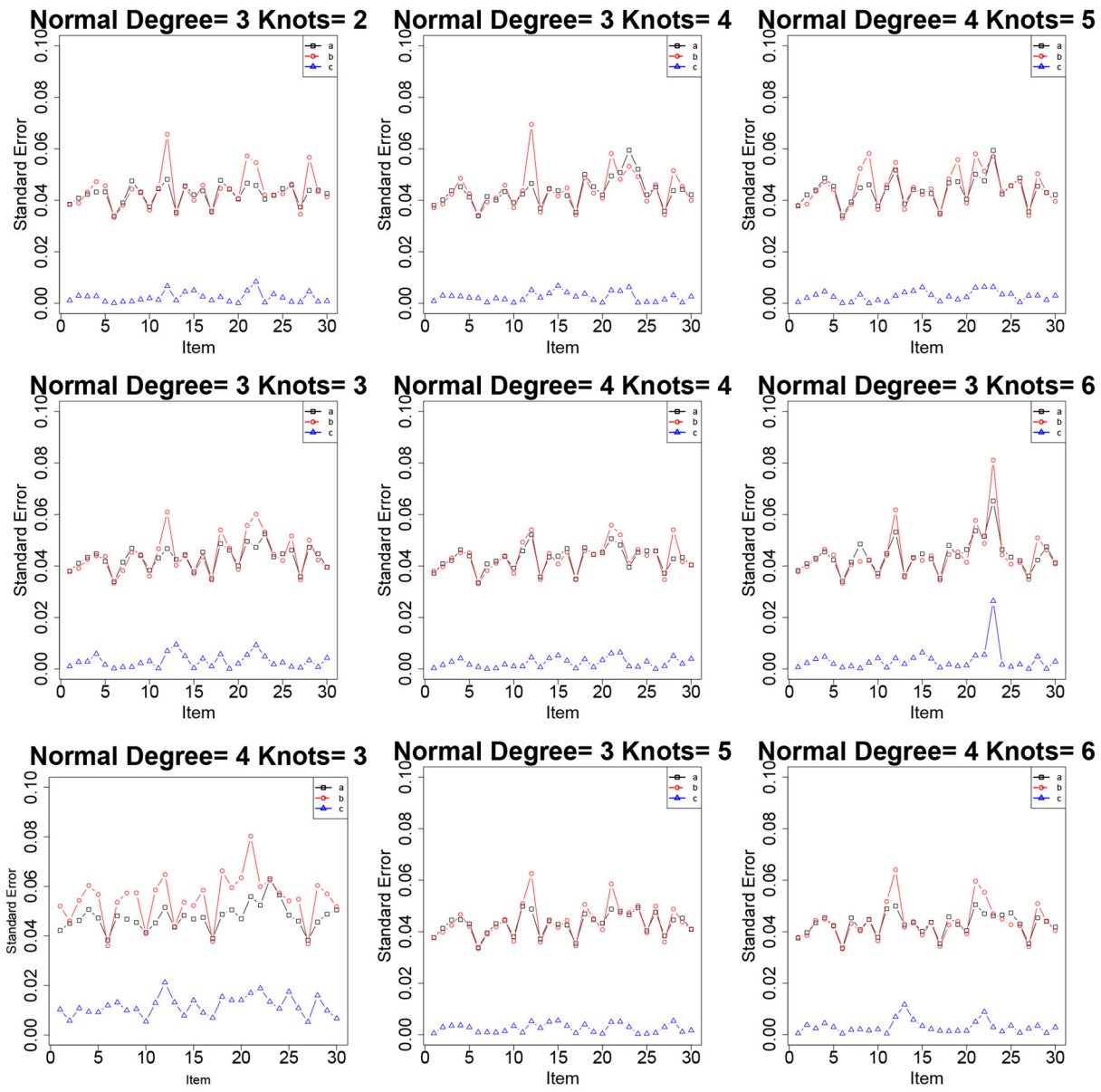


FIGURE 2
The standard errors of item parameters using the RC-SAEM algorithm when the latent trait is normal in simulation study 1.

curves are almost coincident when the true ability density is skewed, and the two curves show only slight differences when the true latent-ability density is bimodal. In addition, the estimation results of item parameters and the ISE values for 15 items in Table 6 are generally similar to those of 30 items in Table 5. This shows that the accuracy and precision of the RC parameters from the RC-SAEM algorithm are not markedly influenced by the number of items. This result indicates that 15 items is sufficient for the proposed RC-SAEM algorithm to provide satisfactory estimation of the RC parameters for the RC-3PNO model. In contrast to the previous methods in RC-IRT (Woods, 2008), our proposed RC-SAEM algorithm has obvious advantages in terms of estimating RC parameters with relatively short test lengths. For the space limitation, the standard errors of item parameters are presented in Figures A1–A4 of the Appendix. As can be seen from the figures, the standard errors of item parameters show a decreasing trend as the sample size increases. Moreover, the standard errors of the item parameters under all conditions are within reasonable ranges.

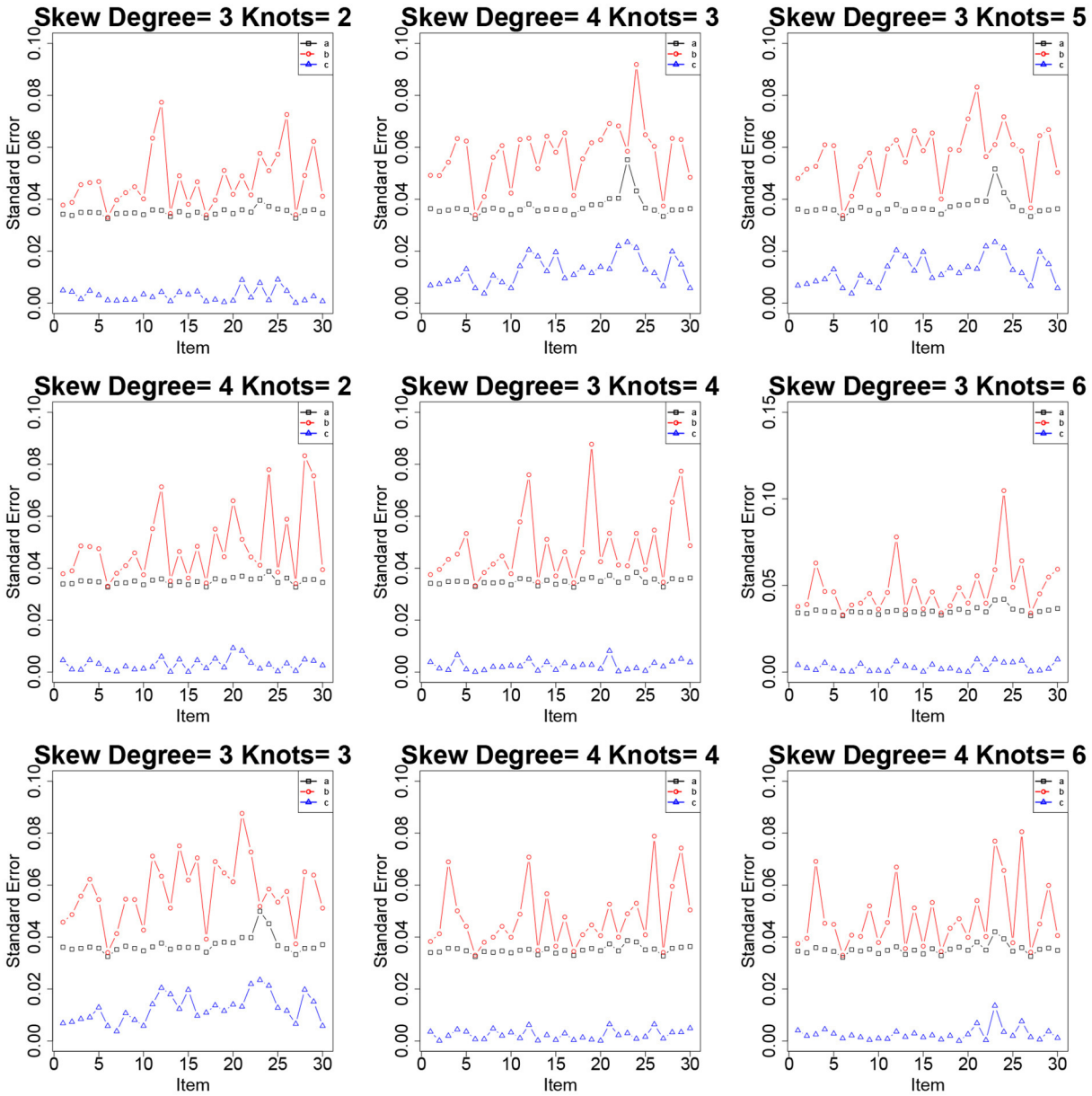


FIGURE 3
The standard errors of item parameters using the RC-SAEM algorithm when the latent trait is skew in simulation study 1.

6. Empirical study

A real data set from the computer-based mathematics assessment of the Programme for International Student Assessment (PISA 2018; OECD, 2019) in China was analyzed. Binary responses from 872 subjects on 11 items in one test form were selected. The SAEM algorithm was fitted to the RC-3PNO and 3PNO models for the real data set. The specifications of the priors of the model parameters were the same as those used in simulation study 2. The same three model selection criteria (AIC, BIC, and HQIC) were used to assess the fit of the model to the real data.

Among the RC-3PNO models with different *knots* and *degree* values (i.e., the 10 combinations used in simulation study 1), that with 2 *knots* and a *degree* of 3 yielded the lowest AIC, BIC, and HQIC values. Thus, this model was selected as the best-fitting model in the subsequent analysis. Table 7 shows the model selection results for this 2-3 RC-3PNO model and the 3PNO model. As shown, the values of all three model selection criteria are smaller for the RC-3PNO model than the 3PNO

TABLE 5 Bias and RMSE of the item parameter estimates and the values of ISE statistic under different ability densities when the test length is 30.

			a		*b*		*c*		ISE
			Bias	RMSE	Bias	RMSE	Bias	RMSE	
$N = 500$	Normal	SAEM	0.044	0.322	-0.059	0.305	-0.010	0.069	-
		RC-SAEM	0.031	0.261	−0.008	0.236	−0.014	0.067	0.061
		MCMC	0.051	0.265	−0.082	0.294	0.012	0.057	-
	Skewed	SAEM	0.096	0.508	-0.387	0.584	0.016	0.084	-
		RC-SAEM	0.030	0.261	−0.021	0.239	−0.017	0.071	−0.006
		MCMC	0.112	0.433	−0.399	0.592	0.034	0.071	-
	Bimodal	SAEM	0.094	0.350	-0.059	0.405	−0.012	0.079	–
		RC-SAEM	0.097	0.376	−0.024	0.471	-0.012	0.083	−0.143
		MCMC	0.154	0.370	-0.191	0.383	0.038	0.070	–
$N = 1,000$	Normal	SAEM	0.017	0.202	−0.044	0.218	−0.011	0.060	-
		RC-SAEM	0.015	0.186	0.001	0.185	−0.015	0.059	0.027
		MCMC	0.023	0.170	-0.002	0.174	0.003	0.047	-
	Skewed	SAEM	0.043	0.334	−0.323	0.436	0.022	0.083	–
		RC-SAEM	0.014	0.190	0.001	0.187	−0.018	0.072	−0.039
		MCMC	0.056	0.286	−0.313	0.415	0.034	0.065	-
	Bimodal	SAEM	0.067	0.275	−0.072	0.342	-0.005	0.072	-
		RC-SAEM	0.074	0.281	−0.025	0.385	−0.010	0.075	−0.035
		MCMC	0.157	0.316	-0.240	0.370	0.047	0.072	–
$N = 2,000$	Normal	SAEM	0.012	0.152	−0.018	0.155	−0.012	0.057	-
		RC-SAEM	0.009	0.136	0.012	0.139	-0.014	0.055	0.008
		MCMC	0.015	0.137	−0.038	0.154	0.004	0.041	-
	Skewed	SAEM	0.024	0.256	-0.294	0.362	0.032	0.078	-
		RC-SAEM	0.009	0.148	0.000	0.154	−0.015	0.066	−0.007
		MCMC	0.026	0.203	-0.261	0.318	0.039	0.059	-
	Bimodal	SAEM	0.060	0.249	−0.075	0.314	−0.002	0.065	-
		RC-SAEM	0.056	0.237	0.001	0.324	-0.006	0.073	0.087
		MCMC	0.223	0.343	−0.315	0.460	0.067	0.080	-

N denotes the sample size, normal, skewed, and bimodal refers to the shape of the ability density. The SAEM refers to estimating 3PNO model using SAEM, RC-SAEM refers to estimating RC-3PNO model using SAEM, and the MCMC refers to estimating 3PNO model using MCMC algorithm.

model. This indicates that the RC-3PNO model gives a better model fit than the 3PNO model for this real data set.

The estimated latent-ability density curve of the RC-3PNO model is presented in Figure 6. It can be seen that the estimated latent trait density curve for this model has an obviously negatively skewed trend, which indicates that the math ability of these subjects is above the mean of the population. Therefore, the conventional methods assuming a normal distribution of latent abilities may result in negatively biased ability estimates. Table 8 shows the item parameter estimates for the 11 items of the PISA test form.

7. Discussion

In real testing, the assumption of a normal distribution of latent abilities in IRT may be violated. For example, non-normality could result from the sampling of one or more distinct populations such as those with or without a "disorder." In this case, the use of traditional algorithms, such as MML-EM, SAEM, and MCMC, in which the latent-ability distribution is restricted to normality, leads to severely biased parameter estimates (Woods, 2006, 2007; Azevedo et al., 2011; DeMars, 2012; Molenaar et al., 2012; Wall et al., 2015; Reise et al., 2018). Several methods are used to relax the normal assumption of latent trait distribution. For example, the EH method, log-linear smoothing, Davidian-curve IRT, and Ramsay-curve IRT, have been proposed to estimate the distribution of latent ability simultaneously with the item parameters. The Ramsay curve is flexible in that it can describe non-normal latent-ability distributions (Ramsay, 2000). To date, several RC-IRT models have been developed. For instance, the RC-2PL model (Woods and Thissen, 2006), RC-3PL model (Woods, 2008), and the logistic GRM incorporating a Ramsay curve (Monroe and Cai, 2014). However, the normal ogive model incorporating a

TABLE 6 Bias and RMSE of the item parameter estimates and the values of ISE statistic under different ability densities when the test length is 15.

			a		*b*		*c*		ISE
			Bias	RMSE	Bias	RMSE	Bias	RMSE	
$N = 500$	Normal	SAEM	0.111	0.586	−0.106	0.362	−0.007	0.074	–
		RC-SAEM	0.052	0.385	-0.054	0.286	−0.008	0.068	0.061
		MCMC	0.153	0.439	−0.110	0.414	0.011	0.066	-
	Skewed	SAEM	0.134	0.740	-0.423	0.652	0.017	0.093	–
		RC-SAEM	0.038	0.345	−0.066	0.303	-0.006	0.071	-0.006
		MCMC	0.191	0.473	−0.411	0.585	0.028	0.073	-
	Bimodal	SAEM	0.068	0.362	−0.104	0.292	−0.023	0.066	–
		RC-SAEM	0.058	0.363	0.015	0.334	−0.034	0.073	−0.143
		MCMC	0.131	0.350	−0.060	0.326	0.011	0.063	-
$N = 1,000$	Normal	SAEM	0.036	0.287	-0.053	0.259	−0.008	0.067	–
		RC-SAEM	0.050	0.277	−0.013	0.227	−0.008	0.068	0.027
		MCMC	0.106	0.299	0.011	0.229	0.001	0.056	-
	Skewed	SAEM	0.057	0.453	−0.366	0.502	0.014	0.083	–
		RC-SAEM	0.020	0.247	−0.037	0.239	−0.011	0.072	−0.011
		MCMC	0.134	0.327	−0.308	0.432	0.026	0.065	–
	Bimodal	SAEM	0.053	0.289	−0.128	0.255	−0.012	0.059	-
		RC-SAEM	0.049	0.332	0.010	0.276	−0.032	0.076	-0.035
		MCMC	0.107	0.284	−0.058	0.292	0.014	0.059	–
$N = 2,000$	Normal	SAEM	0.034	0.244	−0.027	0.207	−0.009	0.065	–
		RC-SAEM	0.029	0.191	0.021	0.176	-0.009	0.062	0.008
		MCMC	0.088	0.210	−0.043	0.179	0.003	0.049	-
	Skewed	SAEM	0.025	0.360	−0.306	0.400	0.025	0.082	-
		RC-SAEM	0.010	0.188	−0.004	0.187	−0.013	0.071	−0.007
		MCMC	0.108	0.272	−0.257	0.363	0.032	0.061	-
	Bimodal	SAEM	0.048	0.269	-0.161	0.249	-0.007	0.062	-
		RC-SAEM	0.038	0.284	0.017	0.226	−0.030	0.075	0.087
		MCMC	0.095	0.275	−0.081	0.314	0.021	0.061	–

N denotes the sample size, normal, skewed, and bimodal refers to the shape of the ability density. The SAEM refers to estimating 3PNO model using SAEM, RC-SAEM refers to estimating RC-3PNO model using SAEM, and the MCMC refers to estimating 3PNO model using MCMC algorithm.

Ramsay curve is rarely used due to the constraints that the normal ogive model itself requires for the integral calculation.

To fill the gap of estimating the normal ogive models in RC-IRT, we propose here an SAEM algorithm to estimate the RC-3PNO model with non-normal latent-ability distributions. In contrast to the traditional EM algorithm, the stochastic approximation step of the SAEM algorithm avoids the need for complex integral computation, and the M-step is straightforward to execute due to obtaining sufficient statistics of the MAP estimates for the item parameters in exponential family distributions, which greatly simplifies the computation and improves its efficiency. Compared with the MHRM algorithm, for exponential family distributions, the SAEM algorithm does not need differential calculations when the standard errors of estimates are not required; thus, the calculations of the SAEM algorithm are easier to execute. The estimates of the item parameters and the shape of the latent-ability density can be simultaneously obtained from this new algorithm. By introducing a Ramsay curve into the SAEM procedure, the new algorithm can be applied not only to a normal distribution of latent abilities but also to non-normal scenarios such as skewed and bimodal distributions.

Simulation study 1 investigated three model selection criteria to select the optimal *knots* and *degree* values in the B-spline functions of Ramsay curves. The choice of *knots* and *degree* had no noticeable influence on the bias and RMSE values of the item parameters. The results of simulation study 2 indicated that the proposed RC-SAEM algorithm generally performs better than the SAEM and MCMC algorithms when the true ability density is skewed or bimodal. Specifically, the RC-SAEM algorithm is obviously superior to the SAEM and MCMC algorithms according to the bias of item parameters in the RC-3PNO model when the true θ is skewed or bimodal. Although the RMSE values of item parameter estimates from the

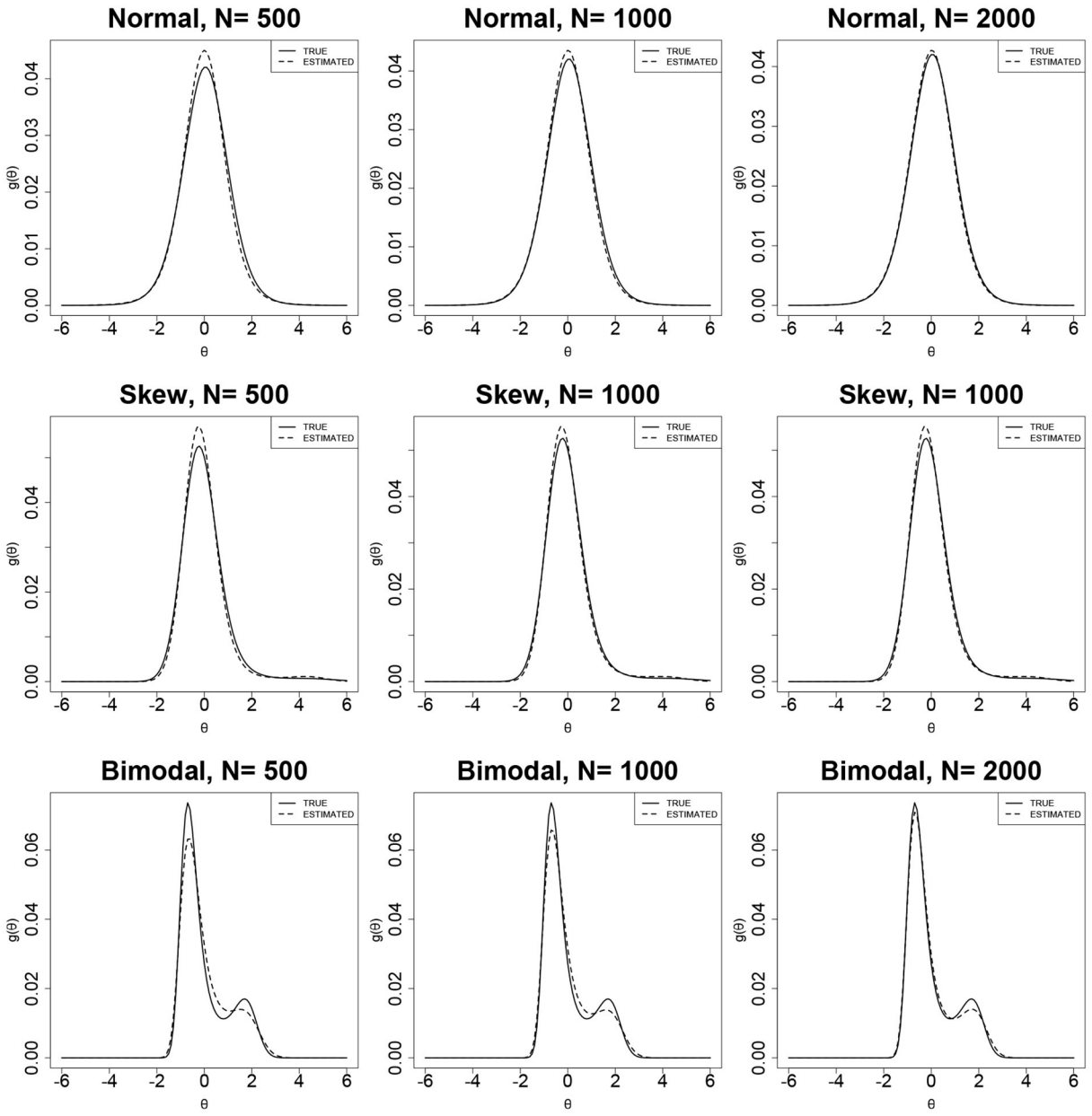


FIGURE 4
True and estimated density curves using the RC-SAEM algorithm when test length is 30.

RC-SAEM algorithm are sometimes slightly larger than those of the SAEM in bimodal cases, especially for the *b* parameter, they are still within the acceptable range. Compared with the suggested sample size of 1,000 for the 3PL model used in RC-IRT (Woods, 2008), a sample size of 500 is large enough for the estimation of parameters in RC-3PNO with a test length of 15.

For the empirical example, according to the model selection criteria (AIC, BIC, and HQIC), the RC-3PNO model gives a better model fit than the 3PNO model. The shape of the estimated Ramsay curve indicates that the latent abilities of these examinees are mainly distributed at the higher level of the latent-ability continuum. In real testing, for binary responses influenced by guessing, although both the RC-3PL and RC-3PNO models serve as possible alternatives, the RC-3PNO model is suggested because the proposed SAEM algorithm avoids the need for calculations of the integral in the E-step and the derivatives in the M-step of the original EM algorithm, which greatly simplifies the computation. We suggest that the RC-3PNO model can be used to detect a non-normal shape in a latent trait distribution. In this case, our proposed RC-SAEM

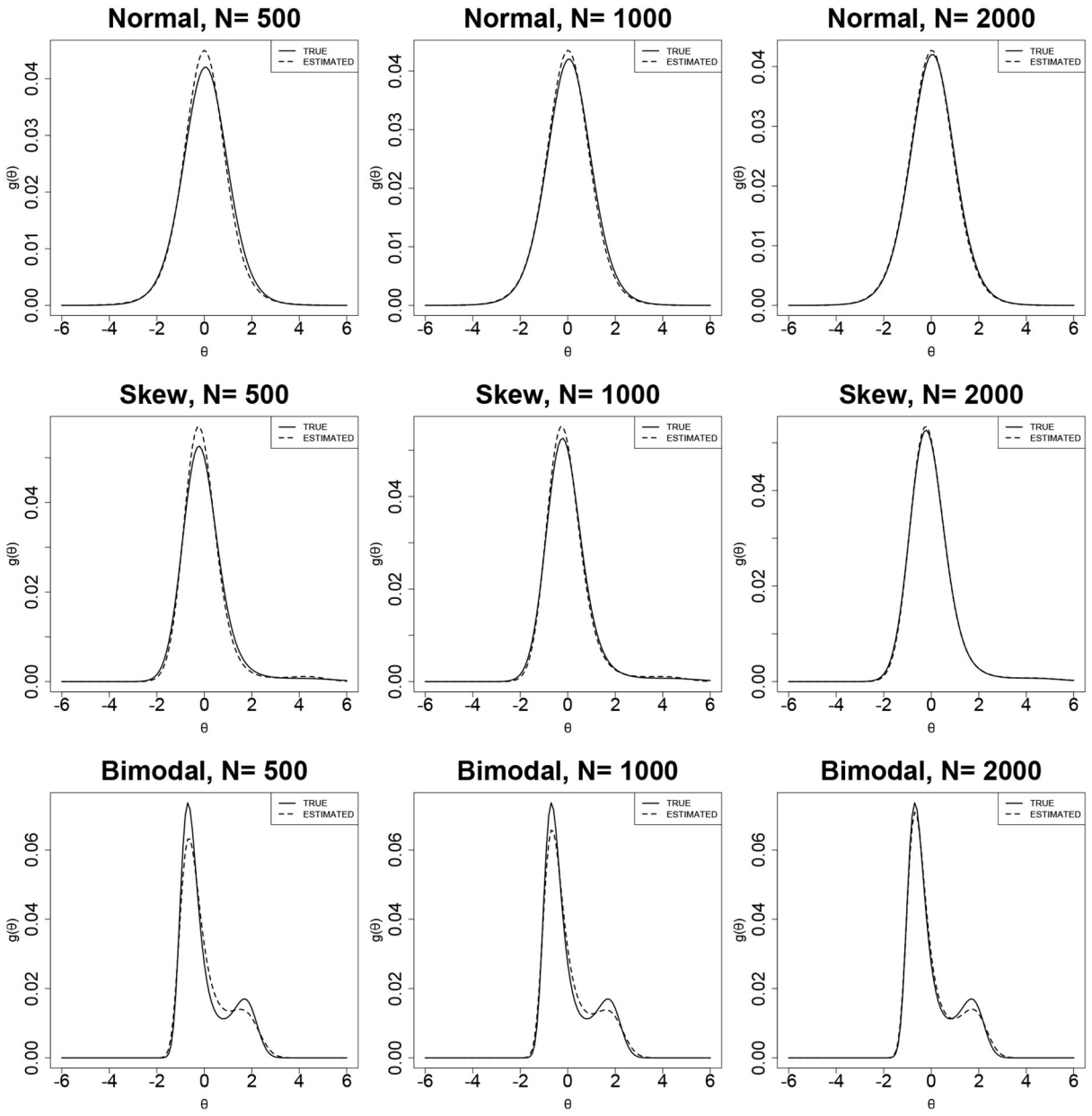


FIGURE 5
True and estimated density curves using the RC-SAEM algorithm when test length is 15.

TABLE 7 Model selection results for the real data set.

	Model parameters	$-2\log L$	AIC	BIC	HQIC
RC-3PNO	34	9664.3	9738.3	9914.9	9805.9
3PNO	30	10196.1	10262.1	10419.5	10322.3

algorithm can also be adopted to simultaneously estimate the item parameters and the latent-ability density.

Several limitations and extensions of the proposed RC-SAEM algorithm need to be mentioned. First, the proposed RC-SAEM algorithm can be extended to other models, such as the GRM and the four-parameter normal ogive (4PNO) model (Culpepper, 2016). Second, a notable fact is that a multidimensional generalization of RC-IRT has not yet been developed. When such a development occurs, the proposed RC-SAEM algorithm can be extended to multidimensional models. Third, future research could compare the proposed RC-SAEM algorithm with other algorithms involving methods that relax the normality assumption of latent traits, such as DC-IRT and LLS. Fourth, the proposed RC-3PNO model together with the SAEM estimation algorithm could be investigated in

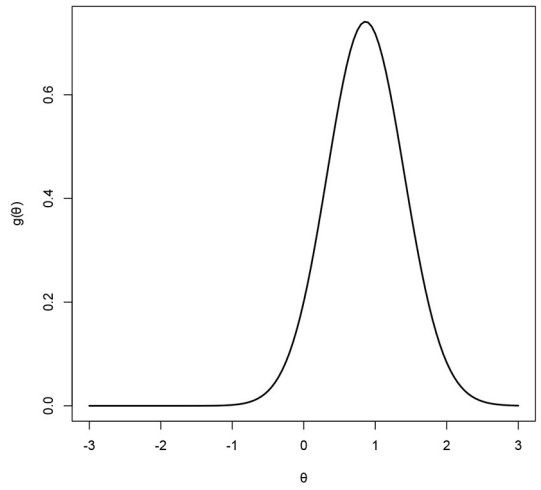


FIGURE 6
Estimated Ramsay-curves for the real data set.

TABLE 8 Item parameter estimates under RC-3PNO model for the real data set.

Item	*a*	*b*	*c*
1	0.23	0.00	0.12
2	1.19	0.53	0.45
3	0.88	−0.83	0.66
4	0.39	0.76	0.92
5	1.22	−0.09	0.55
6	1.64	0.60	0.44
7	1.89	−0.72	0.12
8	1.17	−0.85	0.28
9	1.56	−0.04	0.37
10	1.24	−0.49	0.60
11	1.26	−0.72	0.62

other application domains, such as psychopathology measures involving evidently non-normal latent traits (e.g., borderline personality disorder and dark-triad traits) or medical fields (e.g., drug abuse). Finally, Kuhn and Lavielle (2004) have shown that the SAEM algorithm can also be used for estimating the asymptotic covariance matrix of the maximum-likelihood estimate, and this could be adopted in the RC-SAEM algorithm in the future.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://www.oecd.org/pisa/.

Author contributions

YC and JLu completed the writing of the article and original thoughts. JLi, YC, and JLu provided key technical support. YC provided the data. JZ, JLu, YC, and JLi completed the article revisions. NS and XM provided technical support. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 12001091), China Postdoctoral Science Foundations (Grant Nos. 2021M690587 and 2021T140108), and the Fundamental Research Funds for the Central Universities of China (Grant No. 2412020QD025).

Conflict of interest

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.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.971126/full#supplementary-material

References

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* 60, 255–265. doi: 10.1093/biomet/60.2.255

Azevedo, C. L., Bolfarine, H., and Andrade, D. F. (2011). Bayesian inference for a skew-normal IRT model under the centred parameterization. *Comput. Stat. Data Anal.* 55, 353–365. doi: 10.1016/j.csda.2010.05.003

Baker, F. B., and Kim, S. H. (2004). *Item Response Theory: Parameter Estimation Techniques*, eds S. G. Pantula (North Carolina State University; CRC Press).

Béguin, A. A., and Glas, C. A. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika* 66, 541–561. doi: 10.1007/BF02296195

Birnbaum, A. (1968). "Some latent trait models and their use in inferring an examinee's ability," in *Statistical Theories of Mental Test Scores*, eds F. M. Lord and M. R. Novick (Reading: MIT Press), 397–479.

Bock, R. D., and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika* 46, 443–459. doi: 10.1007/BF02293801

Bock, R. D., and Lieberman, M. (1970). Fitting a response model for dichotomously scored items. *Psychometrika* 35, 179–197. doi: 10.1007/BF02291262

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika* 75, 33–57. doi: 10.1007/s11336-009-9136-x

Camilli, G., and Geis, E. (2019). Stochastic approximation EM for large-scale exploratory IRT factor analysis. *Stat. Med.* 38, 3997–4012. doi: 10.1002/sim.8217

Casabianca, J. M., and Lewis, C. (2015). IRT item parameter recovery with marginal maximum likelihood estimation using loglinear smoothing models. *J. Educ. Behav. Stat.* 40, 547–578. doi: 10.3102/1076998615606112

Culpepper, S. A. (2016). Revisiting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika* 81, 1142–1163. doi: 10.1007/s11336-015-9477-6

De Boor, C. (1978). *A Practical Guide to Splines*. New York, NY: Springer-Verlag. doi: 10.1007/978-1-4612-6333-3

Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Stat.* 94–128. doi: 10.1214/aos/1018031103

DeMars, C. E. (2012). A comparison of limited-information and full-information methods in Mplus for estimating item response theory parameters for nonnormal populations. *Struct. Equat. Model. Multidiscipl. J.* 19, 610–632. doi: 10.1080/10705511.2012.713272

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B Stat Methodol.* 39, 1–22. doi: 10.1111/j.2517-6161.1977.tb01600.x

Geis, E. (2019). *Stochastic approximation EM for exploratory item factor analysis* (Ph.D. thesis) Rutgers The State University of New Jersey, School of Graduate Studies, New Brunswick, NJ, United States.

Geyer, C. J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 56, 261–274. doi: 10.1111/j.2517-6161.1994.tb01976.x

Greig, D. M., Porteous, B. T., and Seheult, A. H. (1989). Exact maximum a posteriori estimation for binary images. *J. R. Stat. Soc. Ser. B Methodol.* 51, 271–279. doi: 10.1111/j.2517-6161.1989.tb01764.x

Gu, M. G., and Zhu, H. T. (2001). Maximum likelihood estimation for spatial models by Markov chain Monte Carlo stochastic approximation. *J. R. Stat. Soc. Ser. B Methodol.* 63, 339–355. doi: 10.1111/1467-9868.00289

Hannan, E. J. (1987). Rational transfer function approximation. *Stat. Sci.* 2, 135–151. doi: 10.1214/ss/1177013343

Harwell, M. R., and Baker, F. B. (1991). The use of prior distributions in marginalized Bayesian item parameter estimation: a didactic. *Appl. Psychol. Meas.* 15, 375–389. doi: 10.1177/014662169101500409

Jank, W. (2006). Implementing and diagnosing the stochastic approximation EM algorithm. *J. Comput. Graph. Stat.* 15, 803–829. doi: 10.1198/106186006X157469

Kuhn, E., and Lavielle, M. (2004). Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM Probabil. Stat.* 8, 115–131. doi: 10.1051/ps:2004007

Lee, C. H., and Gauvain, J. L. (1996). "Bayesian adaptive learning and MAP estimation of HMM, in *Automatic Speech and Speaker Recognition*, eds C. H. Lee, F. K. Soong, and K. K. Paliwal (Boston, MA: Springer), 83–107. doi: 10.1007/978-1-4613-1367-0_4

Maydeu-Olivares, A., and Cai, L. (2006). A cautionary note on using G^2(dif) to assess relative model fit in categorical data analysis. *Multivariate Behav. Res.* 41, 55–64. doi: 10.1207/s15327906mbr4101_4

Meng, X. L., and Rubin, D. B. (1993). Maximum likelihood via the ECM algorithm: a general framework. *Biometrika* 80, 267–278. doi: 10.1093/biomet/80.2.267

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika* 51, 177–195. doi: 10.1007/BF02293979

Molenaar, D., Dolan, C. V., and De Boeck, P. (2012). The heteroscedastic graded response model with a skewed latent trait: testing statistical and substantive hypotheses related to skewed item category functions. *Psychometrika* 77, 455–478. doi: 10.1007/s11336-012-9273-5

Monroe, S., and Cai, L. (2014). Estimation of a Ramsay-Curve item response theory model by the Metropolis-Hastings Robbins-Monro algorithm. *Educ. Psychol. Meas.* 74, 343–369. doi: 10.1177/0013164413499344

OECD. (2019). *PISA 2018 Assessment and Analytical Framework*. Paris: OECD Publishing. doi: 10.1787/b25efab8-en

Ramsay, J. O. (2000). Differential equation models for statistical functions. *Can. J. Stat.* 28, 225–240. doi: 10.2307/3315975

Reise, S. P., and Revicki, D. A. (2014). *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*. New York, NY: Taylor and Francis. doi: 10.4324/9781315736013

Reise, S. P., Rodriguez, A., Spritzer, K. L., and Hays, R. D. (2018). Alternative approaches to addressing non-normal distributions in the application of IRT models to personality measures. *J. Pers. Assess.* 39, 363–374. doi: 10.1080/00223891.2017.1381969

Robbins, H., and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* 22, 400–407. doi: 10.1214/aoms/1177729586

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika* 34 (Suppl. 1), 1–97. doi: 10.1007/BF03372160

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136

Wall, M. M., Park, J. Y., and Moustaki, I. (2015). IRT modeling in the presence of zero-inflation with application to psychiatric disorder severity. *Appl. Psychol. Meas.* 39, 583–597. doi: 10.1177/0146621615588184

Wang, C., Su, S., and Weiss, D. J. (2018). Robustness of parameter estimation to assumptions of normality in the multidimensional graded response model. *Multivariate Behav. Res.* 53, 403–418. doi: 10.1080/00273171.2018.1455572

Wei, G. C. G., and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Am. Stat. Assoc.* 85, 699–704. doi: 10.1080/01621459.1990.10474930

Woods, C. M. (2006). Ramsay-curve item response theory (RC-IRT) to detect and correct for nonnormal latent variables. *Psychol. Methods* 11, 253–270. doi: 10.1037/1082-989X.11.3.253

Woods, C. M. (2007). Ramsay-curve IRT for Likert-type data. *Appl. Psychol. Meas.* 31, 195–212. doi: 10.1177/0146621606291567

Woods, C. M. (2008). Ramsay-curve item response theory for the three-parameter logistic item response model. *Appl. Psychol. Meas.* 32, 447–465. doi: 10.1177/0146621607308014

Woods, C. M., and Lin, N. (2009). Item response theory with estimation of the latent density using Davidian curves. *Appl. Psychol. Measure.* 33, 102–117. doi: 10.1177/0146621608319512

Woods, C. M., and Thissen, D. (2006). Item response theory with estimation of the latent population distribution using spline-based densities. *Psychometrika* 71, 281–301. doi: 10.1007/s11336-004-1175-8

Zhang, D., and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* 57, 795–802. doi: 10.1111/j.0006-341X.2001.00795.x

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 25 November 2022
DOI 10.3389/fpsyg.2022.996609

OPEN ACCESS

EDITED BY
Jorge L. Puga,
University of Granada,
Spain

REVIEWED BY
Isabella Fioravante,
Pontificia Universidad Católica de Chile,
Chile
Seyedmohammad Mirhosseini,
Shahroud University of Medical Sciences,
Iran
Fidel Rebón,
Vice President of data at Giunti Nesplora,
Spain

*CORRESPONDENCE
Jose D. García-Franco
PhD@garcia-franco.net

SPECIALTY SECTION
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

RECEIVED 18 July 2022
ACCEPTED 21 October 2022
PUBLISHED 25 November 2022

CITATION
García-Franco JD, Díez FJ and Carrasco MÁ (2022) Probabilistic graphical model for the evaluation of the emotional and dramatic personality disorders. *Front. Psychol.* 13:996609. doi: 10.3389/fpsyg.2022.996609

Probabilistic graphical model for the evaluation of the emotional and dramatic personality disorders

Jose D. García-Franco[1]*, Francisco J. Díez[1] and Miguel Á. Carrasco[2]

[1]Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain, [2]Department of Psychology of Personality, Evaluation and Treatment. Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain

Personality disorders are psychological ailments with a major negative impact on patients, their families, and society in general, especially those of the dramatic and emotional type. Despite all the research, there is still no consensus on the best way to assess and treat them. Traditional assessment of personality disorders has focused on a limited number of psychological constructs or behaviors using structured interviews and questionnaires, without an integrated and holistic approach. We present a novel methodology for the study and assessment of personality disorders consisting in the development of a Bayesian network, whose parameters have been obtained by the Delphi method of consensus from a group of experts in the diagnosis and treatment of personality disorders. The result is a probabilistic graphical model that represents the psychological variables related to the personality disorders along with their relations and conditional probabilities, which allow identifying the symptoms with the highest diagnostic potential. This model can be used, among other applications, as a decision support system for the assessment and treatment of personality disorders of the dramatic or emotional cluster. In this paper, we discuss the need to validate this model in the clinical population along with its strengths and limitations.

KEYWORDS

personality disorder, probabilistic graphical model, Bayesian network, Delphi, artificial intelligence, knowledge engineering, decision support system

Introduction

We can define personality as the set of traits and qualities that shape a person's way of being and differentiate him or her from others. According to DSM-5, personality disorders can be identified as an enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual's culture. This pattern tends to be stable and of long duration; its onset can be traced back at least to adolescence or early adulthood and affect at least two areas of life (i.e., cognition, affectivity, interpersonal functioning, or impulse control) in an enduring, inflexible, pervasive way across a broad range of personal and social situations, which leads to clinically significant distress or impairment in social,

occupational, or other important areas of functioning (American Psychiatric Association, 2013). While there exist uncountable different configurations that make the individual unique, some of them are more adaptive to the environment and society, while others can be considered dysfunctional, leading to significant psychological distress. Some maladaptive configurations are more prevalent than others and are often seen together; they are termed "personality disorders."

The diagnosis and treatment of personality disorders have several challenges, such as the difficulty of diagnosing many of the maladaptive personality configurations under the current diagnostic approach, or the lack of consensus in the assessments due to evaluator biases. These difficulties are further analyzed in Section Evaluation of Personality Disorders.

The goal of this study is to develop a framework for the research and assessment of personality disorders in the emotional and dramatic cluster, which encompasses the antisocial (ATS), borderline (BDL), narcissistic (NAR), histrionic (HST), and passive-aggressive (PAG) disorders.

We apply artificial intelligence (AI) techniques to integrate different paradigms for the evaluation of personality disorders, which will provide clinicians with a more holistic and accurate tool that will allow them to assess relevant maladaptive psychological variables and psychological distress. This way, clinicians will have a more integral view of the relevant maladaptive psychological variables contributing to psychological distress, which could help reduce the clinical judgment biases derived from the differing backgrounds and profiles of the evaluators. Furthermore, it has been shown that diagnostic accuracy improves when the clinicians have the opportunity to reflect on their diagnosis assisted with the feedback and explanations offered by a decision support system (Oniśko, 2001).

The result of our work is a Bayesian network that models the most relevant psychological constructs related to the emotional and dramatic personality disorders. It contains a number of nodes representing those psychological constructs, a structure representing the relations of probabilistic dependence and independence among these constructs, and a set of conditional probabilities that allows us to draw inferences. These probabilities lead to some metrics, such as the likelihood ratio, which allows us to increase the diagnostic utility of screening and diagnostic tools.

This model allows us to infer the most probable diagnosis given a set of symptoms and find out the sources of psychological distress, which would make good therapeutic targets.

The burden of personality disorders

Some studies indicate that the prevalence of personality disorder lies between 4.4 and 13.0% for the general population (Samuels et al., 2002; Coid, 2003; Lenzenweger et al., 2007; Huang et al., 2009), and can reach as high as 45% among psychiatric outpatients (Zimmerman et al., 2005). This variability can best be seen in Torgersen (2014) work.

Previous research suggests that, although some personality disorders may be considered ego-syntonic, the negative consequences for both the individual and his or her close relatives are significant, ranging from a decrease in both, quality of life (Torgersen, 2014), and life expectancy due to self-harming behaviors (Pompili et al., 2004; Krysinska et al., 2006; Zaheer et al., 2008), to problems with the law due to domestic violence (Whisman and Schonbrun, 2009) or criminal behavior (de Barros and de Pádua Serafim, 2008; Samuels, 2011). Personality disorders also impose a high cost on society as a whole due to the increased use of public health services (Chiesa et al., 2002) and absenteeism from work (Soeteman et al., 2008).

Evaluation of personality disorders

Personality disorders are traditionally assessed by self-report questionnaires, rating scales, interviews, or projective techniques, with significant sources of variance (i.e., information, observation, interpretation, criterion). Many of these tools have not been constructed from an accurate psychometric perspective and have relied exclusively on clinical judgment, rather than an actuarial method, to arrive at a diagnosis (Westen and Shedler, 1999a). Even when some of the most popular and psychometrically well-founded tests (e.g., the Millon Clinical Multiaxial Inventory, MCMI; or the Minnesota Multiphasic Personality Inventory, MMPI) or structured interviews (e.g., Personality Disorder Interview–IV PDI–IV or the Structured Clinical Interview SCID–II) are used to make a diagnosis, they are often time-consuming and always have to be conducted by experienced or well-trained professionals. Moreover, these traditional procedures have focused mainly on the symptoms described in the DSM (Westen and Shedler, 1999a; Widiger and Lowe, 2011), which, in spite of being considered the "gold standard," do not examine personality disorders from an integrated and holistic approach. As a result, the most frequently diagnosed personality disorder is the "Not Otherwise Specified" (Clark et al., 1997; Verheul and Widiger, 2004; Livesley, 2012) and 60% of patients in need of clinical psychotherapeutic attention due to a personality pathology are currently undiagnosable on DSM Axis II (Westen and Arkowitz-Westen, 1998).

Furthermore, the pressure imposed in successive revisions of the DSM to improve its internal and external validity, keeping at the same time a manageable number of symptoms (currently less than 10), helps explain the high comorbidity between personality disorders as well as the additional relations between symptoms and disorders beyond those described in the DSM (Westen and Shedler, 1999b). However, in real life, maladaptive personality is multifactorial and it is not conceivable that every patient fits neatly into a single personality disorder.

Due to these limitations, according to Westen and Shedler (1999a), most clinicians rely, primarily, on inferences drawn from the patient narrative of their lives and relations. This approach, while helping address the limitations previously discussed, is

time-consuming and likely to induce a bias in the clinical judgment, which is known to reduce the diagnostic accuracy. Meehl (1954) proved that statistical judgment is up to 13% more accurate than clinical judgment (Ægisdóttir et al., 2006).

However, the biggest shortcoming and one of the main reasons that led scientists to push forward the research on personality disorders is the inadequate coverage of their different expressions (Widiger, 2007) and the lack of comprehensiveness (Westen and Shedler, 2000).

Given that the DSM has not yet provided an optimal solution for the evaluation of personality disorders, scientists have pursued other directions. Research has led to alternative frameworks that relate other psychological constructs to both general and individual personality disorders, such as the five-factor model (Lynam and Widiger, 2001; Widiger et al., 2002; Samuel and Widiger, 2004; Bagby et al., 2005), defense mechanisms (Berman and McCann, 1995; Cramer, 1999; Bowins, 2010), and Millon's biosocial model (Piersma et al., 2002; Mullins-Sweatt and Widiger, 2007; Millon, 2011).

These alternative frameworks, which have the potential to discriminate those persons with an adaptive personality from those with a disordered personality, and also between different personality disorders, are not generally used, *per se*, for the diagnosis of personality disorders, even though these frameworks are supported by empirical research or by a solid theoretical basis.

Most assessment tools are based on the DSM criteria (Widiger and Lowe, 2011), so these limitations apply, to more or less an extent, to the usual evaluation questionnaires used nowadays by clinical psychologists; hence, the need to incorporate these alternative frameworks into the evaluation of personality disorders. The advantages of a unified framework that increases coverage of symptoms by including all the psychological constructs related to personality disorders justify our research, as nowadays the treatment of personality disorders is individualized, aiming at the person's symptoms rather than at the disorder itself (Millon and Grossman, 2007; Millon and Grossman, 2007a,b). Furthermore, a more comprehensive measurement tool could allow us to reduce biases, both those induced by the person being evaluated, since we would have more information on which to make a decision, as well as those of the evaluator since it could enhance his/her clinical judgment with a statistical/probabilistic tool.

Decision support systems in psychology

One of the main applications of AI is the development of expert systems which are software programs able to mimic the human decision process (Saibene et al., 2021). Many expert systems have been built for different medical domains, but very few for psychology. Saibene et al. (2021), in a five-year review of the literature, identified 43 studies regarding the application of expert systems in healthcare; only 2 were related to psychology, and none of them to personality or its disorders although Luxton (2014) had identified several areas of psychology where the use of AI technology could make a difference.

From 2015 onward there has been, according to Graham et al. (2019), a steep increase in the number of publications about AI for mental health. However, our database search (Scopus, Web of Science, Science Direct, PubMed, IEEE Xplore) with the terms "expert system," "decision support system," or "artificial intelligence" on the one hand, and "personality disorders" or any of the individual disorders on the other, only returned tangential research (Singh et al., 2020; Ellouze et al., 2021; Khazbak et al., 2021), proposals (Tuena et al., 2020; Sulistiani et al., 2021; Szalai, 2021), or proofs of concept (Nunes et al., 2009; Casado-Lumbreras et al., 2012; Randa and Permanasari, 2014; Laijawala et al., 2020).

We conjecture that this scarcity of decision support systems in the field of personality disorders may be, in part, because psychological diagnosis is based on phenomenology. Thus, it can be highly subjective as it depends on the experiences of a person with psychological problems. Conversely, medical diagnosis is often helped by laboratory results and other objective quantitative measures, in addition to clinical signs (Fernando et al., 2011). However, an application of Bayesian methods that is gaining importance nowadays is the analysis of networks in which, through a directed acyclic graph and machine learning techniques, an attempt is made to determine the causal relations between the nodes in the network (Briganti et al., 2020; Černis et al., 2021).

Furthermore, there are two trends to build expert systems. One consists in eliciting and encoding the knowledge of human experts; the other, in applying machine learning algorithms to a large dataset (Constantinou et al., 2016). The latter has the problem that curated medical data regarding psychiatric disorders is generally unavailable (Suhasini et al., 2011). In the case of knowledge-based systems, the problem is that the causal mechanism that drives the relations among variables is either poorly understood or mediated by a large number of hidden variables, which makes it very difficult to elicit expert knowledge; additionally, obtaining the numerical parameters for these systems is even more difficult. Moreover, many AI classification techniques, such as neural networks and support vector machines (SVMs) only work with large data sets and not with expert knowledge.

To achieve the proposed goals, we present in Section 2 the methodology used, and in Section 3 the structure of the resulting model, the raw probabilities obtained, and the likelihood ratios for the symptoms of personality disorders. We conclude the presentation with a discussion of the model and its applications in clinical and research settings (Section 4).

Materials and methods

Participants

We recruited two groups of psychologists with academic and/or clinical expertise in the diagnosis and treatment of personality disorders.

The first group ($n = 5$), which has several years of clinical experience ($M = 12; SD = 7$), was tasked with validating the psychological variables, identified through a literature search, and the structure of the model.

The second group ($n = 7$), also having several years of experience ($M = 20; SD = 15$), was responsible for obtaining the conditional probability tables used as parameters in the model.

Instruments

For the development of the model, a set of questionnaires was used to define the structure of the model and another set to obtain the conditional probabilities. These questionnaires were custom-made and tailored to obtain the causal links among nodes and the probabilities of the symptoms conditioned on the disorders.

All the questionnaires were completed using forms embedded within PDF files, which could be received, answered, and sent back electronically, thus facilitating the participants' engagement.

For the identification of the causal relations between personality disorders and symptoms, the experts were provided with a questionnaire with several tables, one for each psychological framework. For each table, every row corresponds to one of the symptoms, and every column to one of the five personality disorders. The questionnaire consisted of checkboxes (one per cell on each table), which allowed entering a yes/no answer indicating whether the symptom is related to the personality disorder.

Symptoms and dependency links were previously established through a literature review and the study of different psychological measurement instruments for personality disorders. The relations cited as relevant in the literature had previously been checked. Participants were instructed to unmark the checkbox should they consider that a relationship is not sufficiently relevant (if it was previously checked) or leave it blank (if it was not). Similarly, if the experts considered that a symptom was related to a particular personality disorder, they were instructed to mark the checkbox if it was not already marked, or leave it checked if it already was, thus validating the previous literature search.

To standardize the interpretation of symptoms, we briefly described them in the questionnaire. Furthermore, at the end of the form, there was a free-text field so that the experts could add any missing psychological constructs and their relations with the disorders.

To obtain the parameters of the model, the second group of experts was given a set of questionnaires classified by personality disorder.

Again, the rows corresponded to the symptoms but, in this case, through the columns, we sought the probability that the symptom defined in the row would be present when: (*a*) the personality disorder was also present, (*b*) when the personality disorder was absent (control group) and (*c*) the probability that the symptom may cause significant psychological distress.

The scale for data input consisted of a rating scale from 0 to 100. This scale was conceptually divided into four intervals, which were assigned four probability categories: 0–25 "not probable," 25–50 "improbable," 50–75 "probable," and 75–100 "very probable." A graph depicting this division was printed on the header of each page and served as a guide for the psychologist, who is usually more familiar with Likert scales, to elicit the probabilities. The answers were recorded on numerical text fields in each cell, which allowed entering a value between 0 and 100.

Following the Delphi method, the first questionnaire was common to all the participants. This form included, as items, all the parameters that we would need for the construction of the model.

In the next round, a personalized form was used for each participant. For those items in which there was no consensus, defined as those answers that were more than one standard deviation away from the mean, his/her previous response, as well as aggregated data about the responses of other experts, were included. The participant had the chance to modify the previous answer or to keep it. For those items for which there was consensus, it was not allowed to modify the previous answer.

Procedure

The participants in this research received by e-mail a letter of introduction and an invitation to participate in the project. No expert ever knew the identity of the others. All questionnaires included instructions for their correct completion and a demographic data form.

Regarding the structure of the model, the dependency relations finally included were those for which there was consensus (simple majority) among the first group of experts. We anticipated that those relations for which there was no clear consensus would not be sufficiently relevant to significantly affect the accuracy of the model, given that probabilities would be assigned based on the strength of that relation.

The probabilities for the model were obtained using the Delphi method, with at least two rounds. After the first round, the experts were provided with aggregated data (mean and standard deviation) of the answers given in the previous round by all the participants. Each expert could keep his/her previous response or modify it. The process ended when a consensus had been reached or when no further progress was obtained after successive rounds.

According to Hsu and Sandford (2007), the key factor for the success of the Delphi technique is the choice of experts. The number of participants should be enough to obtain a representative sample of expert opinions (Latif et al., 2016), but an excessive number would slow down the process without a substantial improvement in accuracy (Hsu and Sandford, 2007).

In a systematic review of consensus-building methods, Waggoner et al. (2016) suggest having 6 to 11 participants. As previously mentioned, we involved 7 experts in this phase.

The number of rounds required in the methodology is not established. Waggoner et al. (2016) propose a minimum of two

rounds, which is the minimum required to obtain at least one feedback from their colleagues. However, although no maximum number of rounds is established, other authors, like Hasson et al. (2000) and Woudenberg (1991), argue that two rounds are usually sufficient, as this is when maximum accuracy is reached. We have used two rounds in this research since, after analyzing the results of the second one, we saw an obvious risk of a regression to the mean, thus reducing the diversity of responses.

Although the use of the Delphi methodology to obtain conditional probability tables seems promising, we have only found two studies using it (Chen and Huang, 2018; Wu et al., 2018). However, the details of the implementation of the method are not described in those papers, so we have relied on a general approach (Hasson et al., 2000; Waggoner et al., 2016) and adapted it to our research.

The value finally selected for each probability was the average of the responses in the last round.

Development of the probabilistic graphical model

A probabilistic graphical model (PGM) is an encoded probability distribution in which the variables are represented as nodes and the dependence relations as edges between nodes.

A Bayesian network (BN) is a type of PGM consisting of an acyclic directed graph and a conditional probability table for each node given its parents,

$$P\left(X_i \middle| pa\left(X_i\right)\right).$$

The joint probability implicitly represented by a BN is:

$$P\left(X_1, X_2 \ldots X_n\right) = \prod_i P\left(X_i \middle| pa\left(X_i\right)\right),$$

where $pa\left(X_i\right)$ is the set of parents of node X_i in the graph.

A *finding* determines with certainty the state of a variable; for example, the value "true" or "high." The set of all the findings available at a point in time is called *evidence*.

Probabilistic reasoning consists in calculating the posterior probabilities of variables of interest that are not in the evidence.

One advantage of BN is the ease of integrating statistical data with expert knowledge. Another one is the possibility of working with missing data. Furthermore, BN have good accuracy even with small data sets with the use of canonical models (Oniśko et al., 2001) or when probabilities are not overly precise (Uusitalo, 2007).

The most common sources of information to build Bayesian networks are statistical data, scientific literature, and human experts (Druzdzel and van der Gaag, 2000). In this research, we have combined a search of the scientific literature and knowledge elicitation from human experts.

The construction of a probabilistic graphical model for a given domain has three phases; identifying the variables, defining the structure of the model and obtaining the conditional probabilities (Druzdzel and van der Gaag, 2000). We have carried out them using the graphical user interface of OpenMarkov, an open-source tool (Arias et al., 2011) and then exported the model to the academic version of GeNIE (Druzdzel, 1999) to take advantage of its graphing capabilities.

We should note that, although OpenMarkov is very useful for building Bayesian networks, we can benefit from customized software development that acts as an interface between the user and the model. Such an interface, which we developed in conjunction with the Bayesian network throughout this research, improves the usability of the system and allows a clinician to interact with the model without the need to know about Bayesian networks or their building tools.

Identification of the relevant variables, the type of variable (continuous or discrete) and the number of different states

The variables included in the model should cover as broadly as possible the psychological spectrum related to the personality disorders that we want to assess, but without including duplicated or highly correlated variables.

These psychological constructs should be easily measurable and, if possible, familiar to the clinical psychologists who will make use of the decision support system. Therefore, the selection of those variables was performed using the "snowball" method of literature review, taking as starting points papers about commonly used questionnaires for the diagnosis of personality disorders.

Included in the model as nodes are all the symptoms of the classical DSM diagnostic method. None of the specific constructs from the alternative dimensional diagnostic method published in the latest version of the DSM were considered due to the small amount of research on this new model and the absence of some personality disorders (i.e., narcissistic, histrionic and passive-aggressive personality disorders). However, since this dimensional model is an adaptation of the older five-factor model, its exclusion will not have a negative impact because the same psychological constructs are covered by the five-factor model which, additionally, has been extensively used as a personality measurement instrument and in relation to personality disorders (Costa and Widiger, 2002; Widiger and Costa, 2013).

Regarding the five-factor model, we have included in our model all the traits from the domains of neuroticism, extraversion, and agreeableness and all the traits of openness and conscientiousness, except the traits of aesthetics, ideas, values, and achievement-striving, which are the ones that, according to the majority of the studies reviewed (Lynam and Widiger, 2001; Widiger et al., 2002; Samuel and Widiger, 2004; Bagby et al., 2005) did not have a strong relation with personality disorders of the dramatic or emotional type.

The psychological constructs of the DSM-5 new diagnostic method that capture the severity of the personality disorder

(Hutsebaut et al., 2016) has been included. These variables, namely identity, empathy, intimacy, and self-direction, correspond to the general factors common to all the personality disorders and match the four scales of the level of personality functioning (LPFS; Hopwood et al., 2018).

In addition to the variables related to the diagnosis of personality and its disorders, other variables that facilitate the differential diagnosis have been included in the model, such as defense mechanisms (acting out, idealization, denial, dissociation, devaluation, projection, projective identification, splitting, displacement, and passive aggression; American Psychiatric Association, 2000) and the six polarities (pleasure, pain, active, passive, self, other) from the Millon's biosocial theory related to the maladaptive configurations of the individual's styles of adaptation to the environment (Millon, 2011).

Along with the variables we have just described, which correspond to the symptoms, we have also included in the model five nodes corresponding to the personality disorders, as well as other nodes (14 in total) that we use to measure the psychological distress that cluster of symptoms may produce in the patient.

Although the measurements for the psychological variables and even the personality disorders are continuous in nature, we have discretized all the variables. This is a common approach, as there are no efficient algorithms to deal with Bayesian networks that include continuous variables, either for inference or learning, even for very simple models.

Furthermore, given that the computational complexity increases very fast with the number of states, we have only used binary variables (yes/no, present/absent) for the DSM framework and for the defense mechanisms. The nodes representing the personality disorders themselves and the psychological distress have been also modeled as binary variables.

Variables from the level of personal functioning, the five-factor, and the biosocial models have been discretized into three states: low, medium, and high. However, for the five-factor and the biosocial models, the medium state not only indicates a point between the other extreme values, but also it implies that the score obtained is not significant and that it falls within the population mean.

Identifying and representing the causal relations

We have modeled the network assuming that personality disorders cause the symptoms. This way we limit the number of ancestor nodes and reduce the overall complexity of the model. Therefore, a node will only have as many ancestors as the number of personality disorders that may cause it.

An overview of the model structure is presented in Figure 1.

The first two levels of that figure correspond to a BN2O model, which is widely used in medical expert systems (Heckerman, 1990). It consists of an upper level whose nodes represent possible diagnostics, and a lower level (the middle level in our figure), containing the symptoms, observations, medical tests, etc.

The third level in the figure is an extension to the model, first introduced in this research. When introducing evidence about the symptoms, those that are absent may cancel the impact of those that are present, leading to a false-negative diagnosis. The third level in the model alleviates the problem by allowing us to detect clusters of maladaptive symptoms even when the diagnosis is negative. These nodes, which represent the psychological distress in the individual, are also used to perform a sensitivity analysis and to indicate the best therapeutic targets for treatment.

We can observe in the figure that there are no dependency links between diagnoses, which would indicate comorbidity, or between symptoms, which would indicate some kind of correlation among them. The absence of relations between symptoms is deliberate, motivated by the need to reduce the complexity of the model. On the one hand, we have avoided introducing highly correlated symptoms, as it would be redundant, and, on the other, weak dependencies are usually removed given that they do not significantly change the results in classification tasks (Kjærulff, 1994). Furthermore, the inclusion of these relations would not affect the diagnosis given that, when we make a node deterministic by introducing a finding, its state is not affected by the probabilities given its ancestor nodes. As for comorbidity between diagnoses, while it is documented between personality disorders, we model this comorbidity through the common symptoms that these disorders have; hence, the lack of direct links among disorders.

The initial list of dependency links between symptoms and personality disorders for the probabilistic graphical model was obtained from the same literature review used to identify the relevant psychological constructs, and then peer-reviewed by the team of experts, as explained above, using the questionnaire designed for this purpose.

Obtaining the conditional probabilities

Probabilistic graphical models allow for the combination of experimental data with expert knowledge. Since a sufficient amount of suitable data is rarely available in the field of mental health (Suhasini et al., 2011), the probabilities associated with the nodes were elicited from a group of experts. However, a person's experience may be biased by his/her professional experience; we overcome this drawback by using the Delphi methodology for obtaining a consensus, as explained in Section 2.3.

One of the advantages of this method, in addition to the elimination of outlier answers, is that it encourages the participants to reflect on their answers, thus reducing idiosyncratic biases or a tendency to answer too quickly due to fatigue and the large number of items.

The results obtained through the questionnaires are the raw probabilities that indicate the chance that the symptom is present when a single personality disorder is also present (or absent). To obtain the conditional probability tables for the model, it is

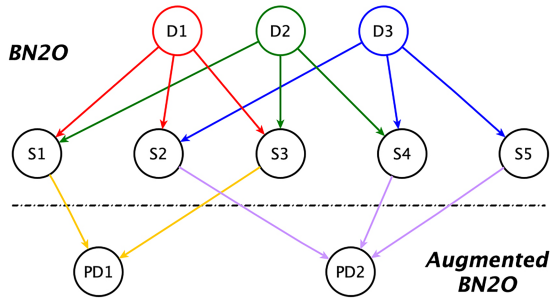


FIGURE 1
Augmented BN2O model. D_x=Disorder; S_y=Symptom; PD_z=Psychological distress.

necessary to first carry out a transformation, due to the difficulty of eliciting from the experts the probabilities of the symptoms when we have to take into account the joint presence or absence of several personality disorders simultaneously.

Moreover, the presence of a large number of ancestor nodes causes an exponential increase in computational complexity (an instance of "the curse of dimensionality"), which we have solved by using canonical models (Diez and Druzdzel, 2006) and taking advantage of the "independence of causal influence" property. This property assumes that the impact of a single cause on the effect does not depend on other causes that may exist, their order, or their interaction (Heckerman and Breese, 1994). Furthermore, canonical models allow complexity to grow linearly with the number of ancestor nodes. So, despite obtaining an approximation to the true values, we actually may gain accuracy by simplifying the elicitation of expert knowledge.

Regarding our model, for two-state variables, we used a "leaky OR" model, and for those three-state variables whose "neutral" state—understood as the absence of disorder or anomaly—is the lowest, we used a "leaky MAX." For an in-depth review of these and other canonical models, see (Diez and Druzdzel, 2006).

However, the above-mentioned canonical models are not adequate for modeling all of the three-state nodes because: (*a*) some nodes behave as inhibitors themselves, that is, they reduce the probability that the symptom is present when a given disorder is also present; and (*b*) for these three-state variables, the default state is not its lowest.

To deal with these variables, we have developed a novel canonical model that allows us to work with multi-state variables without the limitations described above. Its rationale is that there are causes that count as evidence in favor of a given effect. The more evidence we have, either because given the cause the effect is very likely, or because there are several causes supporting the effect, the greater the probability that said effect is present. Conversely, the more evidence against the effect, the less likely it is to be present. We assume that, as in clinical diagnosis by professionals, the probability of the effect (a symptom) depends on the weighting of the evidence for and against, taking into account that not all findings have the same diagnostic potential.

The raw probabilities we obtained using the Delphi method, besides being necessary for generating the conditional probability tables for the model, allow us, for each symptom, to calculate the likelihood ratio with respect to each personality disorder, which is a widely used metric in clinical settings for measuring diagnostic strength.

The positive likelihood ratio for a test result indicates the magnitude of the increase in the probability of a given disorder when the test is positive. Conversely, the negative likelihood ratio for a test result indicates the decreased likelihood of a given disorder when the test is negative (Hayden and Brown, 1999; Grimes and Schulz, 2005).

By identifying symptoms with a higher positive likelihood ratio, we can develop a reduced measurement instrument to confirm the presence of personality disorders of the dramatic and emotional type in a clinical setting. Conversely, by identifying symptoms with a lower negative likelihood ratio we can design a screening instrument to rule out the presence of those personality disorders in the general population.

Results

Raw probabilities obtained with the Delphi methodology

The results presented in the following tables are the probabilities that each symptom is present when the personality disorder (ATS, BDL, NAR, HST, or PAG) is also present, the probability that the symptom is present in the absence of any personality disorder (Norm.) and the psychological distress the symptom may provoke (PD).

For ease of reading, the results have been split into different tables and classified by diagnostic framework: DSM (Table 1), defense mechanism (Table 2), level of personality functioning (Table 3), five-factor model (Table 4), and Millon's biosocial model framework (Table 5). The prevalence of personality disorders is shown in Table 6 for both the clinical and the general population.

Most of the symptoms described here are maladaptive, i.e., they have a positive correlation with the personality disorder (which is also maladaptive). However, for the five-factor model (Table 4) and Millon's biosocial model (Table 5), the presence of a symptom may imply an increase in probabilities with one disorder but a decrease in probabilities with another disorder. A direct relation is represented by an upward pointing arrow and an inverse relation by a downward arrow.

The results obtained correspond to the average of the probabilities provided by the experts in the final round of the Delphi method. However, it is interesting to mention that the consensus degree of the experts in the first round was, on average, similar for all the personality disorders ($66.43\% \pm 12.10\%$).

10.3389/fpsyg.2022.996609

In the second round, the experts modified a considerable number of responses that fell outside the range of consensus by the experts ($79.63\% \pm 25.80\%$), but the consensus degree raised only slightly ($72.21\% \pm 10.76\%$). The average probability for the presence of a symptom in the presence of the corresponding personality disorders was $71.92\% \pm 11.08\%$. Alternatively, the average probability of the presence of a symptom in the absence of any personality disorder was $25.05\% \pm 9.00\%$.

As for the clinically significant psychological distress that the symptoms described in the model are capable of producing, we obtained a mean probability of $47.63\% \pm 19.03\%$.

TABLE 1 Probabilities (%) of DSM symptoms for cluster-B personality disorders.

DSM symptom	Personality disorders					Norm.	PD
	ATS	BDL	NAR	HST	PAG		
DSM-ATS-01	76.4	—	—	—	—	11.4	46.4
DSM-ATS-02	81.4	—	—	—	—	27.9	28.6
DSM-ATS-03	64.3	75.0	—	—	—	36.4	52.1
DSM-ATS-04	77.1	70.7	—	—	—	35.0	60.7
DSM-ATS-05	65.7	66.4	—	—	—	25.7	41.4
DSM-ATS-06	81.4	—	—	—	—	22.9	36.4
DSM-ATS-07	80.7	—	73.6	—	—	11.4	27.1
DSM-BDL-01	—	81.4	—	64.3	—	26.4	69.3
DSM-BDL-02	—	86.4	—	65.0	—	17.9	67.1
DSM-BDL-03	—	88.6	—	—	—	11.4	76.4
DSM-BDL-04	—	85.7	—	—	—	17.1	78.6
DSM-BDL-05	—	76.4	—	—	—	15.7	78.6
DSM-BDL-06	—	85.7	—	72.1	—	17.9	79.3
DSM-BDL-07	—	82.1	—	—	—	16.4	79.3
DSM-BDL-08	75.7	80.7	—	—	—	22.9	72.9
DSM-BDL-09	—	63.6	—	40.7	—	10.0	75.7
DSM-NAR-01	—	—	85.7	—	—	23.6	14.3
DSM-NAR-02	—	—	85.7	—	—	22.9	16.4
DSM-NAR-03	—	—	91.4	—	—	25.0	19.3
DSM-NAR-04	—	—	90.0	80.0	—	22.1	26.4
DSM-NAR-05	—	—	84.3	—	—	23.6	14.3
DSM-NAR-06	—	—	85.7	—	—	29.3	25.0
DSM-NAR-07	79.3	—	77.1	—	—	16.4	22.1
DSM-NAR-08	—	—	77.1	—	77.9	32.1	23.6
DSM-NAR-09	—	—	86.4	—	—	24.3	19.3
DSM-HST-01	—	—	—	87.9	—	16.4	48.6
DSM-HST-02	—	—	—	81.4	—	19.3	45.0
DSM-HST-03	—	—	—	78.6	—	21.4	55.7
DSM-HST-04	—	—	—	81.4	—	22.1	35.0
DSM-HST-05	—	—	—	77.9	—	22.1	27.1
DSM-HST-06	—	—	—	87.9	—	15.7	42.1
DSM-HST-07	—	63.6	—	82.1	—	25.0	35.7
DSM-HST-08	—	62.1	—	80.7	—	17.1	44.3
DSM-PAG-01	67.1	—	—	—	82.9	22.1	57.1
DSM-PAG-02	—	—	—	61.4	77.9	17.1	57.9
DSM-PAG-03	72.9	—	—	—	77.1	22.1	67.9
DSM-PAG-04	75.0	—	—	—	76.4	22.9	57.9
DSM-PAG-05	—	—	65.0	—	74.3	22.9	52.9
DSM-PAG-06	—	—	—	—	76.4	24.3	57.9
DSM-PAG-07	—	—	—	—	86.4	19.3	64.3

ATS = antisocial; BDL = borderline; NAR = narcissistic; HST = histrionic; PAG = passive-aggressive; Norm. = normative (no personality disorder); PD = psychological distress.

TABLE 2 Probabilities (%) of defense mechanisms for cluster-B personality disorders.

Defense mechanism	Personality disorders					Norm.	PD
	ATS	BDL	NAR	HST	PAG		
Acting Out	85.7	84.3	—	70.0	—	27.9	60.0
Idealization	—	67.1	—	—	—	27.1	44.3
Denial	75.7	78.6	80.0	77.1	—	38.6	28.6
Dissociation	47.1	—	55.0	72.1	—	15.0	55.0
Devaluation	—	85.0	44.3	—	—	17.9	69.3
Projection	76.4	—	70.0	—	—	42.1	34.3
Projective identification	—	—	—	—	77.9	21.4	62.9
Splitting	—	87.9	—	72.1	—	22.9	64.3
Displacement	—	—	—	—	70.0	24.3	54.3
Passive aggression	—	71.4	—	58.6	88.6	24.3	48.6

ATS = antisocial; BDL = borderline; NAR = narcissistic; HST = histrionic; PAG = passive-aggressive; Norm. = normative (no personality disorder); PD = psychological distress.

TABLE 3 Probabilities (%) of level of personality functioning (LPF) scales for cluster-B personality disorders.

LPF scale	Personality disorders					Norm.	PD
	ATS	BDL	NAR	HST	PAG		
Identity	69.3	87.9	65.7	77.9	67.1	15.0	57.9
Self-direction	62.1	80.0	51.4	65.0	70.0	22.1	49.3
Empathy	85.0	75.7	65.0	70.0	78.6	15.0	27.1
Intimacy	80.0	79.3	43.6	75.7	69.3	12.9	45.7

ATS = antisocial; BDL = borderline; NAR = narcissistic; HST = histrionic; PAG = passive-aggressive; Norm. = normative (no personality disorder); PD = psychological distress.

Probabilistic graphical model

Given the structure of the model validated by the first group of experts and the raw probabilities obtained from the second group of experts, we built the Bayesian network.

Nodes of the model

The nodes of the model correspond to all the psychological variables and symptoms listed in the first column of the aforementioned tables. Additionally, it should be added the five nodes corresponding to the five personality disorders we are evaluating and the 14 nodes related to the psychological distress caused by each symptom grouping.

These 14 nodes are distributed as follows: one for each personality disorder in the DSM model (5 in total), 4 for each domain in the FFM model (all except for openness), 3 for the personal functioning scale, one for the defense mechanisms, and

10.3389/fpsyg.2022.996609

TABLE 4 Probabilities (%) of five-factor model (FFM) traits for cluster-B personality disorders.

FFM trait	Personality disorders					Norm.	PD
	ATS	BDL	NAR	HST	PAG		
Anxiety	↓ 57.9	↑ 77.9	—	—	—	44.3	70.7
Angry hostility	↑ 77.1	↑ 80.7	↑ 62.9	—	↑ 77.1	35.7	52.1
Depression	—	↑ 77.1	—	↑ 47.9	—	46.4	77.9
Self-consciousness	↓ 67.9	—	—	—	—	34.3	71.4
Impulsiveness	↑ 83.6	↑ 83.6	—	—	—	37.1	55.7
Vulnerability	—	↑ 80.0	—	↑ 68.6	—	32.9	75.0
Warmth	↓ 63.6	↓ 48.6	↓ 63.6	—	—	32.9	34.3
Gregariousness	↓ 54.3	↓ 38.6	—	↑ 75.0	—	24.3	38.6
Assertiveness	—	—	↑ 62.9	—	↓ 77.1	33.6	61.4
Activity	—	—	—	↑ 57.9	—	47.9	25.7
Excitement seeking	↑ 65.0	—	↑ 49.3	↑ 65.7	—	41.4	30.0
Positive emotions	—	—	—	↑ 54.3	—	27.9	70.7
Fantasy	—	↑ 60.0	↑ 79.3	↑ 77.9	—	35.0	N/A
Feelings	—	—	—	↑ 57.9	—	25.7	N/A
Actions	—	↑ 43.6	—	↑ 65.7	—	33.6	N/A
Trust	↓ 75.0	↓ 65.0	↓ 56.4	↑ 59.3	↓ 73.6	38.6	45.7
Straightforwardness	↓ 84.3	↓ 62.1	↓ 73.6	—	↓ 75.0	35.7	24.3
Altruism	↓ 86.4	—	↓ 76.4	—	—	33.6	18.6
Compliance	↓ 86.4	↓ 70.0	↓ 75.7	—	↓ 75.7	27.1	46.4
Modesty	↓ 65.0	—	↓ 87.1	—	—	38.6	24.3
Tender-mindedness	↓ 80.7	—	↓ 75.0	—	—	24.3	17.1
Competence	—	↓ 75.7	↑ 76.4	—	↓ 70.7	25.0	69.3
Order	—	↓ 54.3	—	—	—	36.4	36.4
Dutifulness	↓ 80.7	—	—	—	↓ 70.0	32.1	28.6
Self-discipline	↓ 68.6	—	—	—	↓ 64.3	40.0	45.7
Deliberation	↓ 74.3	↓ 82.1	—	↓ 70.0	—	32.9	45.7

ATS = antisocial; BDL = borderline; NAR = narcissistic; HST = histrionic; PAG = passive-aggressive; Norm. = normative (no personality disorder); PD = psychological distress; N/A = not applicable. Upward arrow = direct relation between symptom and disorder; downward arrow = inverse relation.

TABLE 5 Probabilities (%) of polarities for cluster-B personality disorders.

Polarity	Personality disorders					Norm.	PD
	ATS	BDL	NAR	HST	PAG		
Pleasure	—	↓ 72.9%	↑ 77.1%	↑ 58.6%	↓ 57.1%	↑ 40.0% / ↓ 22.5%	N/A
Pain	—	↑ 67.9%	—	↓ 44.3%	↑ 72.1%	↑ 30.0% / ↓ 20.0%	N/A
Active	—	—	↑ 74.3%	↑ 55.0%	—	↑ 47.5%	N/A
Passive	—	↑ 56.4%	—	↓ 63.6%	↑ 59.3%	↑ 25.0% / ↓ 22.5%	N/A
Self	↑ 82.1%	—	↑ 85.7%	↓ 41.4%	—	↑ 30.0% / ↓ 15.0%	N/A
Other	—	—	—	↑ 20.7%	—	↑ 20.0%	N/A

ATS = antisocial; BDL = borderline; NAR = narcissistic; HST = histrionic; PAG = passive-aggressive; Norm. = normative (no personality disorder); PD = psychological distress; N/A = not applicable. Upward arrow = direct relation between symptom and disorder; downward arrow = inverse relation.

a final one that measures the general psychological distress caused by personality disorders.

Structure of the model

The structure of the model can be determined based on the tables themselves, taking into account that the existence of a probability between symptom and disorder, as seen in the aforementioned tables, implies an arc in the graphical representation.

Furthermore, each of the 14 nodes that account for the psychological distress is linked with the nodes that represent the symptoms or the personality disorders causing that psychological distress.

TABLE 6 Prevalence (%) of dramatic and emotional personality disorders and psychological distress.

Personality disorder	Prevalence		PD
	Clinical population	General population	
Antisocial	12.4	2.4	70.0
Borderline	19.3	3.5	87.1
Narcissistic	11.9	4.3	61.4
Histrionic	13.3	3.6	72.9
Passive-aggressive	9.1	3.0	62.1

PD = psychological distress.

Parameters of the model

For the nodes corresponding to the psychological variables listed under the DSM (Table 1) and the defense mechanisms (Table 2) frameworks, the conditional probabilities were obtained by using the probabilities directly if the node has only one ancestor node, or with the help of a canonical model "leaky OR" otherwise (Diez and Druzdzel, 2006).

For the level of personality functioning paradigm (Table 3), the conditional probability tables are obtained using the canonical "leaky MAX" model (Diez and Druzdzel, 2006).

For the five-factor model (Table 4) and Millon's biosocial model framework (Table 5), we have used a logistic-Gaussian canonical model specifically designed for this research, which allows us to overcome some of the limitations of other canonical models and to take into account the differing prevalence of each symptom, trait, or scale in the population.

For those nodes that have no ancestors, i.e., for each of the five personality disorders, the conditional probability coincides with the prevalence (obtained as well by the Delphi method), which is shown in Table 6 for both the clinical and the general population.

Figure 2 presents a schematic overview of the variables and relations included in the model, and Figure 3 shows a screenshot of the model described above before entering any finding in OpenMarkov's inference mode. In addition, we include a map of the model's variables to facilitate its understanding. However, given its length, it is published as supplementary material.

A working model stored in the format of OpenMarkov or Genie will be supplied upon request.

Likelihood ratio for the improvement of diagnostic efficiency

From the probabilities elicited using knowledge engineering techniques, we have not only been able to obtain the conditional probability tables for the model but also very relevant information on the ranking and relative importance of each symptom with respect to the personality disorders studied.

Through the likelihood ratio, we can identify those symptoms that can most efficiently confirm or rule out the presence of personality disorders.

Tables 7 and 8 show the symptoms that have a positive likelihood ratio greater than 5 or a negative likelihood ratio smaller than 0.2 respectively, which will cause a moderate change in the post-test probabilities with respect to the pre-test probabilities.

Probing the model for content validity: Sensitivity analysis and strength of influence

Except for the graphical representation of the structure of the model or its usefulness in a practical application, it is difficult to ascertain the validity of the model by merely studying the parameters.

One way to solve this problem is by studying the strength influence for the links and the sensitivity analysis of the nodes. This allows us to assess the correctness of the conditional probability tables.

The model has been exported from OpenMarkov to the academic version of GeNIE (Druzdzel, 1999) to take advantage of its graphing capabilities. In Figures 4–6, we can see a sensitivity analysis and the strength of influence for, respectively, the DSM antisocial symptoms, the DSM borderline symptoms, and the LPF scales.

In these images, the nodes in the top row correspond to the five personality disorders, the next row corresponds to the symptoms, traits, or scales of the framework, and the last row (the last two rows in the case of the last figure), corresponds to the node (s) representing psychological distress. Their color indicates the degree of sensitivity: the more redness, the higher the sensitivity.

Furthermore, green arrows indicate a direct influence, while red arrows would imply an inverse one. The thickness of the arrows shows the strength of the influence.

Discussion

The purpose of this research is, through the incorporation of artificial intelligence techniques, to contribute to the improvement in the evaluation and treatment of personality disorders. These disorders, given their high prevalence and negative impact on all involved, require significant attention, especially considering the limitations that traditional methods have in assessing them.

To the best of our knowledge, no study has been conducted that includes the integration of a broad set of psychological variables useful for the evaluation of personality disorders of the dramatic and emotional type in a single model. Nor are there, to date, studies that combine for this purpose expert knowledge, bibliographical research, and statistical methods to integrate the different frameworks related to personality disorders.

To get these results we built a probabilistic graphical model using an open-source software, OpenMarkov (Arias et al., 2011). We obtained from the scientific literature and a group of experts

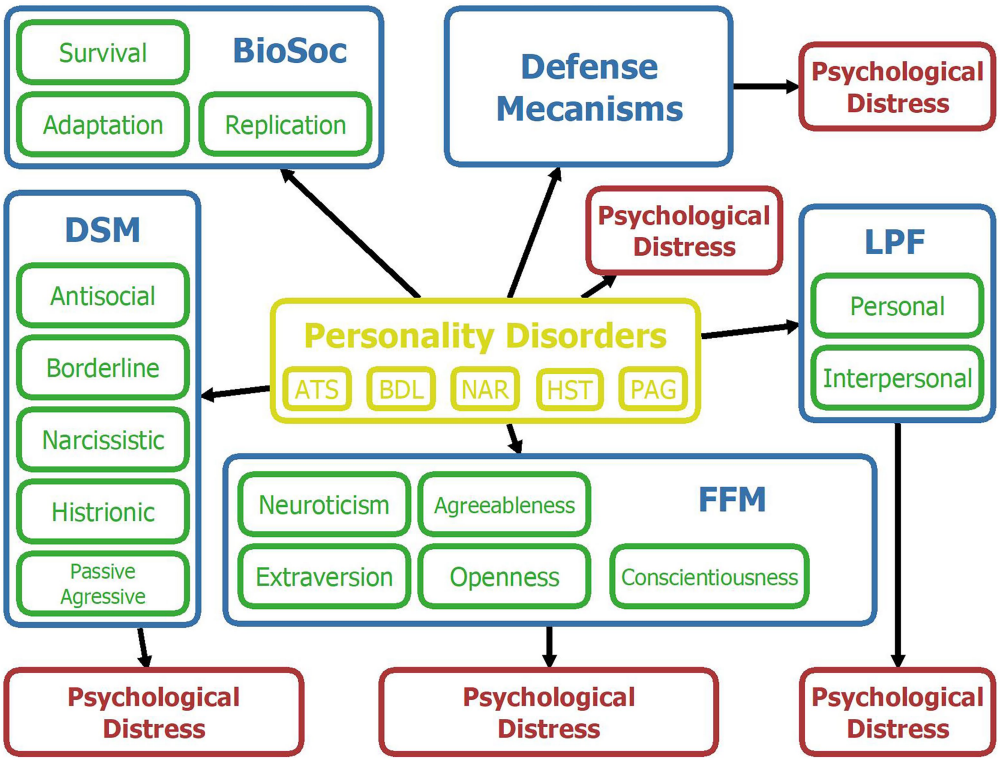


FIGURE 2
Map of variables for the Bayesian network. Yellow=Personality disorders; Blue=Psychological framework; Green=upper-level psychological constructs of a given framework; Red=Psychological distress. ATS=antisocial; BDL=borderline; NAR=narcissistic; HST=histrionic; PAG=passive-aggressive; BioSoc=Biosocial; DSM=Diagnostic and Statistical Manual of mental disorders; FFM=Five-Factor Model; LPF=Level of Personality Functioning.

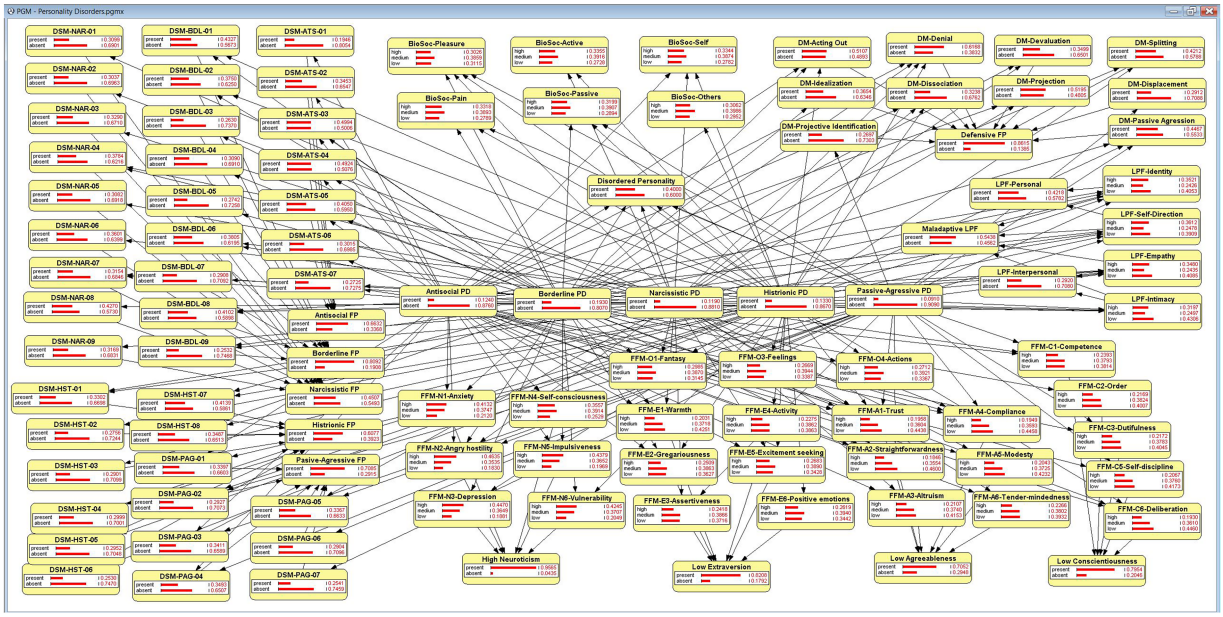


FIGURE 3
The Bayesian network, in OpenMarkov's inference mode.

10.3389/fpsyg.2022.996609

TABLE 7 Symptoms having a positive likelihood ratio (given in parenthesis) higher or equal than 5 for some personality disorder.

ATS	BDL	NAR	HST	PAG
DSM - ATS 07 (7.06)	DSM - BDL 03 (7.75)	DSM - ATS 07 (6.44)	LPF - Intimacy (5.89)	LPF - Intimacy (5.39)
DSM - ATS 01 (6.69)	DSM - BDL 09 (6.36)		DSM - HST 06 (5.59)	LPF - Empathy (5.24)
LPF Intimacy (6.22)	LPF - Intimacy (6.17)		DSM - HST 01 (5.35)	
LPF Empathy (5.67)	LPF - Identity (5.86)		LPF - Identity (5.19)	
	LPF - Empathy (5.05)			
	DSM - BDL 04 (5.00)			
	DSM - BDL 07 (5.00)			

ATS = antisocial; BDL = borderline; NAR = narcissistic; HST = histrionic; PAG = passive-aggressive.

TABLE 8 Symptoms having a positive likelihood ratio (given in parenthesis) lower or equal than 0.2 for some personality disorder.

ATS	BDL	NAR	HST	PAG
LPF Empathy (0.18)	DSM - BDL 03 (0.13)	DSM - NAR 03 (0.11)	DSM - HST 06 (0.14)	MD - Passive-aggressive (0.15)
FFM Compliance (0.19)	LPF - Identity (0.14)	DSM - NAR 04 (0.13)	DSM - HST 01 (0.15)	DSM - PAG 07 (0.17)
MD - Acting out (0.20)	MD - Splitting (0.16)	DSM - NAR 09 (0.18)		
	MD - Devaluation (0.18)	DSM - NAR 02 (0.19)		
	DSM - BDL 02 (0.17)	DSM - NAR 01 (0.19)		
	DSM - BDL 04 (0.17)			

ATS = antisocial; BDL = borderline; NAR = narcissistic; HST = histrionic; PAG = passive-aggressive.

following a Delphi method approach (Hasson et al., 2000; Waggoner et al., 2016). This model represents the relations between a broad set of psychological symptoms and the personality disorders of the dramatic and emotional cluster.

This model facilitates the assessment of personality disorders under a wide range of symptoms from different psychological frameworks. Additionally, with the probabilities obtained through the Delphi method, it has been possible to identify those psychological constructs with the highest diagnostic power for the confirmation or screening of personality disorders.

With respect to the model and its structure, the changes proposed by the experts regarding the relations found in the literature were minimal and, in any case, the changes were to introduce previously absent relations.

The fact that the relations initially included in the model, obtained from the literature, were hardly questioned gives confidence in the correctness of the model. Nevertheless, a bias or carry-over effect should not be ruled out, since the questionnaire specified those relations obtained from the scientific literature. Furthermore, the experts did not propose other psychological variables for inclusion in the model which is a positive indicator that the probabilistic graphical model is exhaustive in terms of the constructs or psychological variables.

Once the structure of the model was defined, the conditional probability tables were obtained from experts by the Delphi method showing that the average degree of agreement between the first and second rounds only increased by around 8%. This modest increase, which would hardly justify an additional Delphi round, occurs mainly because the standard deviation decreases as the scores get closer to the mean, so that, if we keep the same procedure as in the first round, reaching a higher consensus becomes more difficult even though, paradoxically, the results are closer to the mean. This finding is in line with the studies of Hasson et al. (2000) and Woudenberg (1991).

Furthermore, the percentage of items that were modified between the first and second rounds was considerable $(\approx 80\%)$, which seems to indicate a tendency to conform to the mean, probably due to peer pressure.

Given the conditional probabilities obtained for the model, we have been able to determine those symptoms that best allow us to confirm a suspected personality disorder in the clinical population and to rule out its presence in the general population. By identifying the symptoms with a higher positive likelihood ratio, we can develop a reduced measurement instrument to confirm the presence of personality disorders of the dramatic and emotional type in clinical settings. Conversely, by identifying symptoms with a lower negative likelihood ratio we can design a screening instrument to rule out the presence of personality disorders of the dramatic and emotional type in the general population. This would reduce the time needed between an initial consultation, where the patient's clinical history is explored, and the moment of providing the treatment. Furthermore, the creation of a screening tool would allow us to reach more population and provide better access to mental health care without incurring the excessive cost of an indiscriminate complete psychological study.

The advantage of this approach with respect to the traditional method, in which the questionnaires used only include constructs from a single framework, is that, by using a questionnaire that explores the psychological constructs with the greatest likelihood ratio from different frameworks, we obtain a measurement instrument that, with the same extension, has greater diagnostic power (Grimes and Schulz, 2005).

The list of symptoms obtained in this study is quite short, so the presence or absence of these symptoms can be determined either by a questionnaire or by a directed interview in a short time.

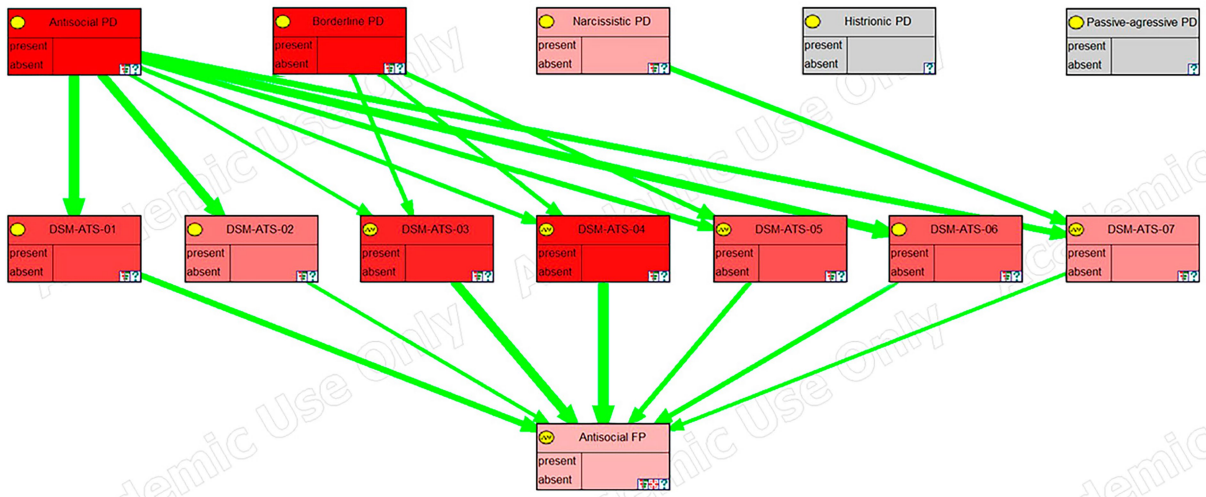


FIGURE 4
Sensitivity analysis, in GeNIE, for antisocial DSM symptoms.

A common cut-off point in the literature has been used, namely $LR+ \geq 5$ and $LR- \leq 0.2$. However, by modifying these cut-off points we can increase or reduce the number of symptoms, which will always be the most relevant, to tailor the desired length of the measurement instrument or the interview.

The most obvious aspect of this list of symptoms is the predominance of those from the DSM model. This was to be expected, since personality disorders are constructs defined on the basis of their symptoms; however, not all symptoms have the same diagnostic power, so this list is useful to rule out those that are either more common in the general population or less common in the clinical population, and can therefore be relegated to a second tier, with minimal loss of diagnostic power.

Other overrepresented symptoms in these lists are the level of personal functioning scales, which are present in the list for all personality disorders except for narcissistic personality disorder, evidence that it is, arguably, the least maladaptive personality disorder of the dramatic and emotional type.

Regarding the defense mechanisms, they appeared only among the symptoms with the lowest negative likelihood ratios. This could be because, although they are highly characteristic of personality disordered individuals, it is not uncommon to find them in the general population, so they are more useful to rule out the disorder than to confirm it. However, given the egosyntonic nature that personality disorders in this cluster tend to have, it is to be expected that coping mechanisms were in play to reduce the psychological distress caused by the effects of the disorder on the person's life.

The five-factor model is hardly represented in the list of the most relevant symptoms for the same reason that defense mechanisms; the prevalence of high or low traits in the normal population is considerable. This supports the study of Rottman et al. (2010) that study that the five-factor model may not be sufficient to diagnose personality disorders. However, one possible solution would be to raise the cut-off points so that, by only considering the variables with the highest (or lowest) and most maladaptive scores as traits present, the prevalence in the normal population would be lowered and the specificity of these traits would be increased. Something similar occurs with Millon's biosocial model whose polarities do not even appear in the list.

Although the model has not yet been validated with a representative sample of patients with personality disorders, the model shows good content validity, as it replicates the findings obtained in other studies using a different methodology. To illustrate this, we performed a sensitivity analysis on some variables of the model using the GeNIE software.

The sensitivity analysis for Antisocial DSM symptoms (Figure 4) showed how the 7 symptoms of this disorder relate primarily to antisocial personality disorder but also, in almost equal measure, to borderline personality disorder despite relating only through 3 of the 7 symptoms. Holthausen and Habel (2018) argued that borderline and antisocial personality disorders are two sides of the same coin and that they have a common underlying factor. They also claimed that the differences between the two disorders come from the way the symptoms manifest and not because of qualitative differences between the disorders. That is the reason why in the graph we see that the symptoms are related to both disorders in almost the same magnitude (depicted by the same intensity of red color).

Likewise, a sensitivity analysis for Borderline DSM symptoms shows its relation with the borderline personality disorder, but also, as mentioned in the previous paragraph, to antisocial personality disorder. However, we can also see that there is an even stronger relation with the histrionic personality disorder. Westen and Shedler (1999b), in one of their studies, make another classification of the disorders using a different methodology from the DSM. They suggest that some of the cases of borderline personality disorder

would be better classified as histrionic personality disorder and in a new category called "emotional dysregulation." Therefore, they propose a new category with symptoms taken from both. These findings are congruent with the graph shown in Figure 5.

A sensitivity analysis corresponding to the psychological variables of the level of personal functioning was also depicted (Figure 6). Sharp et al. (2015) proposed that there is a general factor "g" common to all personality disorders and a specific factor "s" that establishes the differences between the different personality disorders. Our sensitivity analysis showed how the level of personal functioning, measured by its four variables (identity, empathy, intimacy, and self-direction), was affected almost equally by all personality disorders, confirming that we were indeed measuring the "g" factor. However, it also showed how, for the clinically significant psychological distress that this "g" factor produces, the empathy construct had a significantly lower weight. This could be because although empathy is considered a positive attribute, in certain environments, such as finance and politics, is not very adaptive. That is, a lack of empathy is useful to thrive; at the very least, it may not be seen as dysfunctional as the lack of any of the other constructs. This is congruent with some previous work on empathy (Olson, 2012).

The Bayesian network developed in this research has different applications, we will focus on just three.

First, the principal application of a Bayesian network is to compute the posterior probabilities of the states of the variables given a set of findings. In our context, this allows us to determine the probability of each personality disorders given the patient's symptoms. The probability score should not, necessarily, be interpreted in absolute terms, but in relation to the score obtained in the other personality disorders, taking into account that if the x-axis represented the weighted number of symptoms present and the y-axis the probabilities, the function would have a sigmoid shape.

While a therapist is necessary for both the determination of the symptoms and the interpretation of the results, the system can interactively guide the psychological assessment, saving time and facilitating a comprehensive exploration of all the related psychological variables. An advantage with respect to the traditional diagnostic method is the possibility of making a more complete examination, while reducing the evaluator's biases. Although the use of a new tool may initially require an additional effort, this is rewarded with a reduction in the time for the personal interview by being able to directly address the most relevant aspects of the patient's narrative.

The assessment offered by the system is based on the probabilities of both the presence of personality disorders and the likelihood that the evaluated symptoms produce clinically significant psychological distress. The therapist can decide whether to assess all the psychological variables in the model for greater accuracy or to assess a reduced set, in which case the system takes a probabilistic value for the variables whose status is unknown based on the conditional probability tables and the findings entered in the adjacent nodes.

The second application of the system is the possibility of performing a sensitivity analysis—, once the findings have been introduced and an assessment has been obtained,—to determine which symptoms contribute most to the diagnosis. These symptoms constitute the therapeutic targets that may optimize the treatment to reduce the psychological distress as efficiently as possible. However, the fact that a psychological variable has the greatest contribution to the diagnosis does not mean that it is the easiest to be treated, so sensitivity analysis should be regarded as an additional aid to the therapist rather than a straightforward guide.

The third application is the use of the model as an educational tool for psychologists in training. Since there is the possibility of updating, in real-time, a diagnosis based on the symptoms of a patient's psychological profile, a student can see how the diagnosis changes when including or excluding certain symptoms. This, combined with a comprehensive listing of related variables, text boxes with detailed information about symptoms and their characteristics,

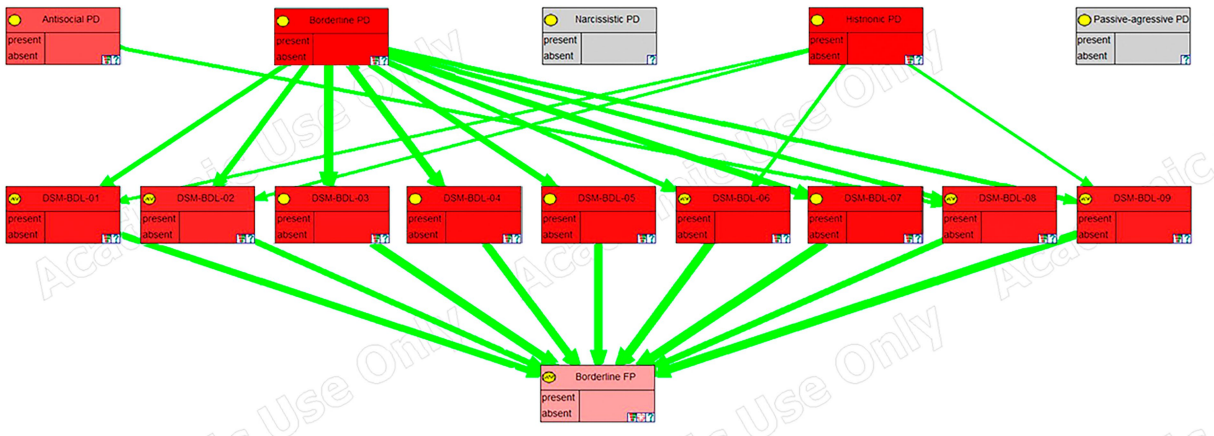


FIGURE 5
Sensitivity analysis, in GeNIE, for borderline DSM symptoms.

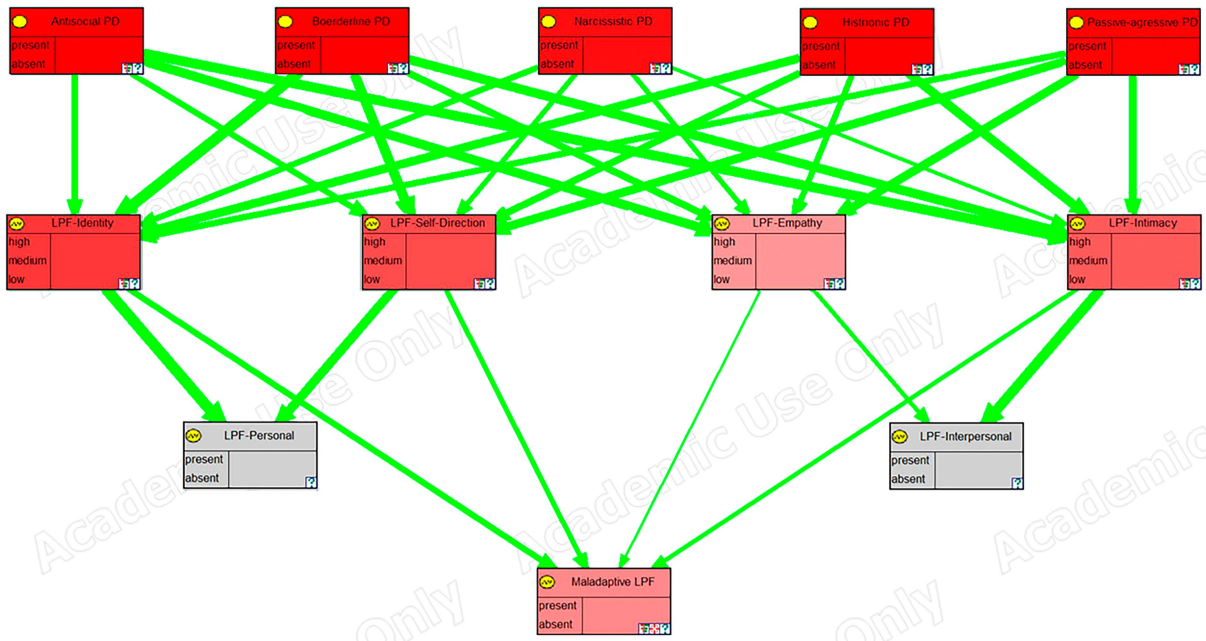


FIGURE 6
Sensitivity analysis, in GeNIE, for level of personality functioning scales.

and color coding of the scores to determine whether the change is positive or negative, we have a simulation tool with great potential to complement other more traditional training methods.

It can be argued that some of the decisions made for the modeling could be somewhat arbitrary, such as the discretization of nodes, the choice of canonical models, or their parameters. However, even the simplest Bayesian networks (i.e., the naive Bayes) are very robust to both imprecise data and approximate assumptions. One of the reasons for such good performance is that, when faced with classification tasks, absolute probabilities between nodes in the model are not as important as the relative probabilities and ranking; that is, if the state of one node is more probable than another, this is be reflected in the model through the probabilities, even if these are not exact (Rish, 2001; Zhang, 2005). This property is maintained with the parameters and the methodology used.

However, one of the next steps to address some of the limitations of this study is to refine the model with statistical data obtained empirically as soon as it is available. Although this statistical data would not be without bias either, it would allow us to fit the model to different populations for a more accurate diagnosis.

Furthermore, in the near future, we will validate the model in a clinical setting to determine its suitability for the assessment and treatment of personality disorders of the dramatic and emotional type. Similarly, it will be of interest to explore the applicability of the model in the training of new psychologists.

Other lines of work aimed at improving the diagnosis and treatment of personality disorders would be taking into account other factors such as ease of treatment and the expectations of success. In this sense, part of the work has already been done by using the Delphi method to measure the psychological distress that each symptom can produce.

The use of artificial intelligence techniques in the field of psychology is an innovative approach that complements traditional techniques used for the investigation and assessment of psychological disorders. Although in this research we have focused on a subset of personality disorders, the methodology is applicable not only to the rest of personality disorders, but also to other psychological conditions whose causality is multifactorial and where empirical data is scarce.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

JG-Fs contribution was in the analysis of the data and the modeling of the Bayesian network. FDs contribution was in the area of artificial intelligence, Bayesian networks, and canonical models. MCs contribution was in the area of psychology, diagnosis and treatment of personality disorders, and data gathering from the group of experts. All authors contributed to the article and approved the submitted version.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.996609/full#supplementary-material

References

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., et al. (2006). The meta-analysis of clinical judgment project: fifty-six years of accumulated research on clinical versus statistical prediction. *Couns. Psychol.* 34, 341–382. doi: 10.1177/0011000005285875

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (DSM-IV-TR)*. United States: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. United States: American Psychiatric Publishing.

Arias, M., Diez, F. J., and Palacios, M. A. (2011). *ProbModelXML. A format for encoding probabilistic graphical models. Tech. rep*. Spain, Universidad Nacional de Educación a Distancia.

Bagby, R. M., Costa, P. T., Widiger, T. A., Ryder, A. G., and Marshall, M. (2005). DSM-IV personality disorders and the five-factor model of personality: a multi-method examination of domain-and facet-level predictions. *Eur. J. Personal.* 19, 307–324. doi: 10.1002/per.563

Berman, S. M., and McCann, J. T. (1995). Defense mechanisms and personality disorders: an empirical test of Millon\textquotesingles theory. *J. Pers. Assess.* 64, 132–144. doi: 10.1207/s15327752jpa6401_9

Bowins, B. (2010). Personality disorders: a dimensional defense mechanism approach. *Am. J. Psychother.* 64, 153–169. doi: 10.1176/appi.psychotherapy.2010.64.2.153

Briganti, G., Scutari, M., and Linkowski, P. (2020). Network structures of symptoms from the Zung depression scale. *Psychol. Rep.* 124, 1897–1911. doi: 10.1177/0033294120942116

Casado-Lumbreras, C., Rodríguez-González, A., Álvarez-Rodríguez, J. M., and Colomo-Palacios, R. (2012). PsyDis: towards a diagnosis support system for psychological disorders. *Expert Syst. Appl.* 39, 11391–11403. doi: 10.1016/j.eswa.2012.04.033

Černis, E., Evans, R., Ehlers, A., and Freeman, D. (2021). Dissociation in relation to other mental health conditions: an exploration using network analysis. *J. Psychiatr. Res.* 136, 460–467. doi: 10.1016/j.jpsychires.2020.08.023

Chen, W., and Huang, S. (2018). Evaluating flight crew performance by a Bayesian network model. *Entropy* 20:178. doi: 10.3390/e20030178

Chiesa, M., Fonagy, P., Holmes, J., Drahorad, C., and Harrison-Hall, A. (2002). Health service use costs by personality disorder following specialist and nonspecialist treatment: a comparative study. *J. Personal. Disord.* 16, 160–173. doi: 10.1521/pedi.16.2.160.22552

Clark, L. A., Livesley, W. J., and Morey, L. (1997). Special feature: personality disorder assessment: the challenge of construct validity. *J. Personal. Disord.* 11, 205–231. doi: 10.1521/pedi.1997.11.3.205

Coid, J. (2003). Epidemiology, public health and the problem of personality disorder. *Br. J. Psychiatry* 182, s3–s10. doi: 10.1192/bjp.182.44.s3

Constantinou, A. C., Fenton, N., and Neil, M. (2016). Integrating expert knowledge with data in Bayesian networks: preserving data-driven expectations when the expert variables remain unobserved. *Expert Syst. Appl.* 56, 197–208. doi: 10.1016/j.eswa.2016.02.050

Costa, P. T., and Widiger, T. A. (Eds.). (2002). *Personality Disorders and The Five-Factor Model of Personality, 2nd Edn*. United States: American Psychological Association.

Cramer, P. (1999). Personality, personality disorders, and defense mechanisms. *J. Pers.* 67, 535–554. doi: 10.1111/1467-6494.00064

de Barros, D. M., and de Pádua Serafim, A. (2008). Association between personality disorder and violent behavior pattern. *Forensic Sci. Int.* 179, 19–22. doi: 10.1016/j.forsciint.2008.04.013

Diez, F. J., and Druzdzel, M. J. (2006). Canonical probabilistic models for knowledge engineering. Techreport, UNED, UNED. Retrieved from http://www.cisiad.uned.es

Druzdzel, (1999). SMILE: structural modeling, inference, and learning engine and GeNIe: a development environment for graphical decision-theoretic models. *Aaai/Iaai*, 902–903.

Druzdzel, M. J., and van der Gaag, L. C. (2000). Building probabilistic networks: where do the numbers come from? Guest editors\textquotesingle introduction. *IEEE Trans. Knowl. Data Eng.* 12, 481–486. doi: 10.1109/tkde.2000.868901

Ellouze, M., Mechti, S., and Belguith, L. H. (2021). *Approach based on ontology and machine learning for identifying causes affecting personality disorder disease on twitter. Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), 12817 LNAI*, Springer, Cham. 659–669.

Fernando, I., Henskens, F., and Cohen, M. (2011). A domain specific expert system model for diagnostic consultation in psychiatry. 2011 12th ACIS international conference on software engineering, artificial intelligence, Networking and Parallel/Distributed Computing, IEEE. doi:10.1109/snpd.2011.38

Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H.-C., et al. (2019). Artificial intelligence for mental health and mental illnesses: an overview. *Curr. Psychiatry Rep.* 21:116. doi: 10.1007/s11920-019-1094-0

Grimes, D. A., and Schulz, K. F. (2005). Refining clinical diagnosis with likelihood ratios. *Lancet* 365, 1500–1505. doi: 10.1016/s0140-6736(05)66422-7

Hasson, F., Keeney, S., and McKenna, H. (2000). Research guidelines for the Delphi survey technique. *J. Adv. Nurs.* 32, 1008–1015. doi: 10.1046/j.1365-2648.2000.t01-1-01567.x

Hayden, S. R., and Brown, M. D. (1999). Likelihood ratio: a powerful tool for incorporating the results of a diagnostic test into clinical Decisionmaking. *Ann. Emerg. Med.* 33, 575–580. doi: 10.1016/s0196-0644(99)70346-x

Heckerman, D. (1990). "A tractable inference algorithm for diagnosing multiple diseases" in *Uncertainty in artificial intelligence*, vol. *10* (Netherlands: Elsevier), 163–171.

Heckerman, D., and Breese, J. S. (1994). A new look at causal Independence. In Uncertainty proceedings 1994 , eds. M. Henrion, D. Shachter, Ross, N. Kanal, Laveen, Lemmer and F. John. (pp. 286–292). Elsevier. doi:10.1016/b978-1-55860-332-5.50041-9

Holthausen, B. S., and Habel, U. (2018). Sex differences in personality disorders. *Curr. Psychiatry Rep.* 20:975. doi: 10.1007/s11920-018-0975-y

Hopwood, C. J., Good, E. W., and Morey, L. C. (2018). Validity of the DSM–5 levels of personality functioning scale–self report. *J. Pers. Assess.*, 100, 650–659. doi: 10.1080/00223891.2017.1420660

Hsu, C.-C., and Sandford, B. A. (2007). The Delphi technique: making sense of consensus. *Pract. Assess. Res. Eval.* 12, 1–8. doi: 10.7275/pdz9-th90

Huang, Y., Kotov, R., de Girolamo, G., Preti, A., Angermeyer, M., Benjet, C., et al. (2009). DSM–IV personality disorders in the WHO world mental health surveys. *Br. J. Psychiatry* 195, 46–53. doi: 10.1192/bjp.bp.108.058552

Hutsebaut, J., Feenstra, D. J., and Kamphuis, J. H. (2016). Development and preliminary psychometric evaluation of a brief self-report questionnaire for the assessment of the DSM–5 level of personality functioning scale: the LPFS brief form (LPFS-BF). *Personal. Disord. Theory Res. Treat.* 7, 192–197. doi: 10.1037/per0000159

Khazbak, M., Wael, Z., Ehab, Z., Gerorge, M., and Eliwa, E. (2021). MindTime: deep learning approach for borderline personality disorder detection. 2021 international Mobile, intelligent, and ubiquitous computing conference (MIUCC), (pp. 337–344). doi:10.1109/MIUCC52538.2021.9447620

Kjærulff, U. (1994). "Reduction of computational complexity in Bayesian networks through removal of weak dependences" in *Uncertainty proceedings*, Craig Boutilier and Moisés Goldszmidt. (Netherlands: Elsevier), 374–382.

Krysinska, K., Heller, T. S., and Leo, D. D. (2006). Suicide and deliberate self-harm in personality disorders. *Curr. Opin. Psychiatry* 19, 95–101. doi: 10.1097/01.yco.0000191498.69281.5e

Laijawala, V., Aachaliya, A., Jatta, H., and Pinjarkar, V. (2020). Classification algorithms based mental health prediction using data mining. 2020 5th international conference on communication and electronics systems (ICCES), (pp. 1174–1178). doi:10.1109/ICCES48766.2020.9137856

Latif, R. A., Mohamed, R., Dahlan, A., and Nor, M. Z. (2016). Using Delphi technique: making sense of consensus in concept mapping structure and multiple choice questions (MCQ). *Education in Medicine Journal* 8:421. doi: 10.5959/eimj.v8i3.421

Lenzenweger, M. F., Lane, M. C., Loranger, A. W., and Kessler, R. C. (2007). DSM-IV personality disorders in the National Comorbidity Survey Replication. *Biol. Psychiatry* 62, 553–564. doi: 10.1016/j.biopsych.2006.09.019

Livesley, J. (2012). Tradition versus empiricism in the current DSM-5 proposal for revising the classification of personality disorders. *Crim. Behav. Ment. Health* 22, 81–90. doi: 10.1002/cbm.1826

Luxton, D. D. (2014). Artificial intelligence in psychological practice: current and future applications and implications. *Prof. Psychol. Res. Pract.* 45, 332–339. doi: 10.1037/a0034559

Lynam, D. R., and Widiger, T. A. (2001). Using the five-factor model to represent the DSM-IV personality disorders: an expert consensus approach. *J. Abnorm. Psychol.* 110, 401–412. doi: 10.1037/0021-843x.110.3.401

Meehl, P. E. (1954). *Clinical versus Statistical prediction: A Theoretical Analysis and A Review of The Evidence*. United States: University of Minnesota Press.

Millon, T. (2011). Classifying personality disorders: an evolution-based alternative to an evidence-based approach. *J. Personal. Disord.* 25, 279–304. doi: 10.1521/pedi.2011.25.3.279

Millon, T., and Grossman, S. D. (2007). *Moderating Severe Personality Disorders: A Personalized Psychotherapy Approach*. United States: Wiley.

Millon, T., and Grossman, S. D. (2007a). *Overcoming resistant personality disorders: A Personalized Psychotherapy Approach*. United States: Wiley.

Millon, T., and Grossman, S. D. (2007b). *Resolving Difficult Clinical Syndromes: A Personalized Psychotherapy Approach*. United States: Wiley.

Mullins-Sweatt, S. N., and Widiger, T. A. (2007). Millon\textquotesingles dimensional model of personality disorders: a comparative study. *J. Personal. Disord.* 21, 42–57. doi: 10.1521/pedi.2007.21.1.42

Nunes, L. C., Pinheiro, P. R., and Pequeno, T. C. (2009). An expert system applied to the diagnosis of psychological disorders. 2009 IEEE international conference on intelligent computing and intelligent systems, 3, pp. 363–367. doi:10.1109/ICICISYS.2009.5358164

Olson, G. (2012). "Corporations: empathy–devoid psychopaths" in *Empathy imperiled* (New York: Springer), 53–60.

Oniśko, A. (2001). Evaluation of the HEPAR II system for diagnosis of liver. *Prace Instytutu Podstaw Informatyki Polskiej Akademii Nauk*, 62–65. doi: 10.1016/S0888-613X(01)00039-1

Oniśko, A., Druzdzel, M. J., and Wasyluk, H. (2001). Learning Bayesian network parameters from small data sets: application of Noisy-OR gates. *Int. J. Approx. Reason.* 27, 165–182. doi: 10.1016/s0888-613x(01)00039-1

Piersma, H. L., Ohnishi, H., Lee, D. J., and Metcalfe, W. E. (2002). An empirical evaluation of Millon's dimensional polarities. *J. Psychopathol. Behav. Assess.* 24, 151–158. doi: 10.1023/a:1016006616346

Pompili, M., Ruberto, A., Girardi, P., and Tatarelli, R. (2004). Suicidality in DSM IV cluster B personality disorders. An overview. *Ann. Ist. Super. Sanita* 40, 475–483. PMID: 15815115

Randa, C. P., and Permanasari, A. E. (2014). Development of diagnosis expert system for personality disorders. 2014 Makassar international conference on electrical engineering and informatics (MICEEI), (pp. 180–183). doi:10.1109/MICEEI.2014.7067335

Rish, I. (2001). An empirical study of the naive Bayes classifier. IJCAI 2001 workshop on empirical methods in artificial intelligence, 3, pp. 41–46.

Rottman, B. M., Kim, N. S., Ahn, W. K., and Sanislow, C. A. (2010). Can personality disorder experts recognize DSM-IV personality disorders from five-factor model descriptions of patient cases? *J. Clin. Psychiatry* 72, 630–639. doi: 10.4088/jcp.09m05534gre

Saibene, A., Assale, M., and Giltri, M. (2021). Expert systems: definitions, advantages and issues in medical field applications. *Expert Syst. Appl.* 177:114900. doi: 10.1016/j.eswa.2021.114900

Samuel, D. B., and Widiger, T. A. (2004). Clinicians' personality descriptions of prototypic personality disorders. *J. Personal. Disord.* 18, 286–308. doi: 10.1521/pedi.18.3.286.35446

Samuels, J. (2011). Personality disorders: epidemiology and public health issues. *Int. Rev. Psychiatry* 23, 223–233. doi: 10.3109/09540261.2011.588200

Samuels, J., Eaton, W. W., Bienvenu, O. J., Brown, C. H., Costa, P. T., and Nestadt, G. (2002). Prevalence and correlates of personality disorders in a community sample. *Br. J. Psychiatry* 180, 536–542. doi: 10.1192/bjp.180.6.536

Sharp, C., Wright, A. G., Fowler, J. C., Frueh, B. C., Allen, J. G., Oldham, J., et al. (2015). The structure of personality pathology: both general (`g') and specific (`s') factors? *J. Abnorm. Psychol.* 124, 387–398. doi: 10.1037/abn0000033

Singh, R., Subramani, S., Du, J., Zhang, Y., Wang, H., Ahmed, K., et al. (2020). Deep learning for multi-class antisocial behavior identification from twitter. *IEEE Access* 8, 194027–194044. doi: 10.1109/ACCESS.2020.3030621

Soeteman, D. I., Roijen, L. H.-V., Verheul, R., and Busschbach, J. J. (2008). The economic burden of personality disorders in mental health care. *J. Clin. Psychiatry* 69, 259–265. doi: 10.4088/jcp.v69n0212

Suhasini, A., Palanivel, S., and Ramalingam, V. (2011). Multimodel decision support system for psychiatry problem. *Expert Syst. Appl.* 38, 4990–4997. doi: 10.1016/j.eswa.2010.09.152

Sulistiani, H., Muludi, K., and Syarif, A. (2021). Implementation of various artificial intelligence approach for prediction and recommendation of personality disorder patient. *J. Physics: Conference Series* 1751:2040. doi: 10.1088/1742-6596/1751/1/012040

Szalai, J. (2021). The potential use of artificial intelligence in the therapy of borderline personality disorder. *J. Eval. Clin. Pract.* 27, 491–496. doi: 10.1111/jep.13530

Torgersen, S. (2014). "Prevalence, sociodemographics, and functional impairment" in *The American Psychiatric Publishing textbook of personality disorders*. eds. J. M. Oldham, A. E. Skodol and D. S. Bender. *2nd Edn* (United States: American Psychiatric Publishing), 109–129.

Tuena, C., Chiappini, M., Repetto, C., and Riva, G. (2020). "Artificial intelligence in clinical psychology" in *Reference Module in Neuroscience and Biobehavioral Psychology* ed. Giuseppe Riva (Netherlands: Elsevier)

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecol. Model.* 203, 312–318. doi: 10.1016/j.ecolmodel.2006.11.033

Verheul, R., and Widiger, T. A. (2004). A meta-analysis of the prevalence and usage of the personality disorder not otherwise specified (PDNOS) diagnosis. *J. Personal. Disord.* 18, 309–319. doi: 10.1521/pedi.2004.18.4.309

Waggoner, J., Carline, J. D., and Durning, S. J. (2016). Is there a consensus on consensus methodology? Descriptions and recommendations for future consensus research. *Acad. Med.* 91, 663–668. doi: 10.1097/acm.0000000000001092

Westen, D., and Arkowitz-Westen, L. (1998). Limitations of Axis II in diagnosing personality pathology in clinical practice. *Am. J. Psychiatr.* 155, 1767–1771. doi: 10.1176/ajp.155.12.1767

Westen, D., and Shedler, J. (1999a). Revising and assessing Axis II, part I: developing a clinically and empirically valid assessment method. *Am. J. Psychiatry* 156, 258–272. doi: 10.1176/ajp.156.2.258

Westen, D., and Shedler, J. (1999b). Revising and assessing Axis II, part II: toward an empirically based and clinically useful classification of personality disorders. *Am. J. Psychiatry* 156, 273–285. doi: 10.1176/ajp.156.2.273

Westen, D., and Shedler, J. (2000). A prototype matching approach to diagnosing personality disorders: toward DSM-V. *J. Personal. Disord.* 14, 109–126. doi: 10.1521/pedi.2000.14.2.109

Whisman, M. A., and Schonbrun, Y. C. (2009). Social consequences of borderline personality disorder symptoms in a population-based survey: marital distress, marital violence, and marital disruption. *J. Personal. Disord.* 23, 410–415. doi: 10.1521/pedi.2009.23.4.410

Widiger, T. A. (2007). "Alternatives to DSM-IV: Axis II" in *Personality disorders: Toward the DSM-V* eds. W. O'Donohue , K. A. Fowler and S. O. Lilienfeld (New York: SAGE Publications, Inc.), 21–40.

Widiger, T. A., and Costa, P. T. (Eds.). (2013). *Personality disorders and the five-factor model of personality*, *3rd*. United States: American Psychological Association.

Widiger, T. A., and Lowe, J. R. (2011). "Personality disorders" in *Handbook of assessment and treatment planning for psychological disorders*. eds. M. M. Antony and D. H. Barlow. *2nd Edn* (New York: Guilford press), 571–605.

Widiger, T. A., Trull, T. J., Clarkin, J. F., Sanderson, C., and Costa, P. T. (2002). "A description of the DSM-IV personality disorders with the five-factor model of personality" in *Personality disorders and the five-factor model of personality*. *2nd* ed eds. Paul T. Costa Jr. and Thomas A. Widiger (United States: American Psychological Association), 89–99.

Woudenberg, F. (1991). An evaluation of Delphi. *Technol. Forecast. Soc. Chang.* 40, 131–150. doi: 10.1016/0040-1625(91)90002-w

Wu, J., Fang, W., Hu, Z., and Hong, B. (2018). Application of Bayesian approach to dynamic assessment of flood in urban underground spaces. *Water* 10:1112. doi: 10.3390/w10091112

Zaheer, J., Links, P. S., and Liu, E. (2008). Assessment and emergency Management of Suicidality in personality disorders. *Psychiatr. Clin. N. Am.* 31, 527–543. doi: 10.1016/j.psc.2008.03.007

Zhang, H. (2005). Exploring conditions for the optimality of naive Bayes. *Int. J. Pattern Recognit. Artif. Intell.* 19, 183–198. doi: 10.1142/s0218001405003983

Zimmerman, M., Rothschild, L., and Chelminski, I. (2005). The prevalence of DSM-IV personality disorders in psychiatric outpatients. *Am. J. Psychiatr.* 162, 1911–1918. doi: 10.1176/appi.ajp.162.10.1911

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 22 December 2022
DOI 10.3389/fpsyg.2022.965623

OPEN ACCESS

EDITED BY
Salvador Chacón-Moscoso,
Sevilla University,
Spain

REVIEWED BY
Joachim Funke,
Heidelberg University,
Germany
Emanuela Furfaro,
University of California,
Davis, United States

*CORRESPONDENCE
Christoph Dominik Güss
dguess@unf.edu

SPECIALTY SECTION
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

RECEIVED 09 June 2022
ACCEPTED 15 November 2022
PUBLISHED 22 December 2022

CITATION
Smith W, Hermida J and Güss CD (2022) "Oh no, the forest is burning!" cultural differences in the complex problem-solving process only under high uncertainty.
Front. Psychol. 13:965623.
doi: 10.3389/fpsyg.2022.965623

"Oh no, the forest is burning!" cultural differences in the complex problem-solving process only under high uncertainty

Willow Smith, Joanna Hermida and Christoph Dominik Güss*

Department of Psychology, University of North Florida, Jacksonville, FL, United States

What do people in different cultures do when they encounter complex problems? Whereas some cross-cultural research exists about complex problem-solving predictors and performance, the process has rarely been studied. We presented participants from Brazil, Germany, the Philippines, and the United States with two computer-simulated dynamic problems, one where quick action was required – the WinFire simulation – and one where cautious action was required – the Coldstore simulation. Participants were asked to think aloud in their native language while working on these two tasks. These think-aloud protocols were digitally recorded, transcribed, and coded by coders in each country in terms of the steps involved in complex problem solving and dynamic decision making. For the current study, we developed a program to calculate transition frequencies from one problem solving step to another and analyzed only those protocols with more than 15 transitions. For WinFire, these were 256 think-aloud protocols from the four countries with a total of 12,542 statement, for Coldstore, these were 247 participants with a total of 15,237 statements. Based on previous, limited cross-cultural research, we predicted that after identifying a problem, Brazilians would make emotional and self-related statements, Germans would engage primarily in planning, Filipinos would gather additional information, and Americans would primarily state solutions. Results of latent transition analysis partially support these hypotheses, but only in the highly uncertain Coldstore situation and not in the more transparent WinFire situation. Transition frequencies were then also analyzed regarding community clusters using the spinglass algorithm in R, igraph. Results highlight the importance of process analyses in different tasks and show how cultural background guides people's decisions under uncertainty.

KEYWORDS

verbal protocol, think-aloud, cross-cultural, complex problem solving, dynamic decision making, lag sequential analysis, log-linear analysis, transition frequencies

Introduction

According to Popper (1999), "All life is problem solving." Problem solving is a part of everyday life. Complex problem solving can be defined as overcoming barriers and reaching goals in complex, dynamic, and non-transparent problem situations (e.g., Brehmer and Dörner, 1993; Funke, 2010; Dörner and Funke, 2017): Complex refers to the multitude of variables involved and their interrelatedness; dynamics refers to a state of affairs in which circumstances are changing over time; and non-transparent means that key aspects of the problem are unknown to the problem solver.

There is extensive research on complex problem solving (e.g., Frensch and Funke, 1995; Dörner, 1996; Greiff and Scherer, 2018; Schoppek et al., 2018) and dynamic decision making (e.g., Fox et al., 2013; Gonzalez et al., 2017). However, only in recent years have complex problem-solving and dynamic decision making been studied in different cultures and among different nationalities (e.g., Güss et al., 2010; Gonzalez and Martin, 2011; Lovett and Forbus, 2011; Wüstenberg et al., 2014). The present research aims to fill a gap in the existing research by testing hypotheses regarding cultural differences in complex problem-solving among Brazilian, German, Filipino, and U.S. American participants with a special focus on the problem-solving process.

Problem solving and decision making involve a series of steps: gathering information, setting goals, making plans, structuring a decision, and making a final decision (e.g., Galotti, 2002). Bransford and Stein's (1993) classic IDEAL model proposes a similar set of steps: Identify, Define, Explore, Act, and Look back. Complex problem solving and dynamic decision making involve, however, not only cognitive, but also motivational, emotional, and self-related processes (Güss et al., 2010, 2017; Dörner and Funke, 2017), but the exact sequence of these processes and the cultural differences related to this sequence have not yet been investigated.

Cross-cultural hypotheses regarding reactions toward problems

We expect to find cultural differences in regard to dealing with complex problem situations that are a result of knowledge that is transmitted, processed and stored by individuals (e.g., Hirschfeld and Gelman, 1994; Sperber and Hirschfeld, 1999) in a specific social, cultural, and historical context (Lave and Wenger, 1991; Cole, 1996). As Hutchins (1995, p. 354) expressed, "culture is an adaptive process that accumulates partial solutions to frequently encountered problems." In the following sections, we attempt to derive culturally preferred patterns for different cultural groups about what to do first when confronted with a problem. Since there is a lack of research many of these hypotheses are tentative. Of course, there is also variation within cultures, but we focus here only on differences between cultures.

Prioritization of individuality and emotional expressiveness are important parts of Brazilians' culture (Véras and Véras, 2011). Véras and Véras studied problem solving within business relationships in Brazil and China. They found that in order to successfully conduct business with Brazilians, one must build and cherish an emotional relationship before negotiating any business deals, and that Brazilians normally do not rely on strict or formal rules of communication. For example, if a Brazilian feels the need to make a statement about their own opinion, or how they feel individually, they will freely do so (Véras and Véras, 2011). Other studies (e.g., Strohschneider and Güss, 1998) have also found intense emotion expressiveness among Brazilians. We hypothesize that Brazilians will make emotional and self-related statements when encountering a problem situation more often than the other cultural groups in our study.

Germans primarily focus on orderliness and are committed to their decisions (Müller et al., 2008). In other research, German companies have been characterized as well-oiled machines (Hofstede, 2001) with a long-term time orientation (Hölter, 2013). In cross-cultural comparisons on problem solving and planning, German students have shown more detailed planning (e.g., Strohschneider and Güss, 1998). Based on these findings, we hypothesize that German participants will focus more on planning when encountering a problem situation than Brazilian, Filipino or United States participants.

Filipinos have been observed to be more hesitant to accept the reliability of information and its sources. In a study comparing decision making in the Philippines, the United States, and Hong Kong (Builtjens and Noorderhaven, 1996), Filipinos collected more information than Americans or Hong Kong residents, and delayed decision making when solving dynamic and complex problems. Thus, we expect that Filipinos will exhibit a more cautious approach to decision making (also partly due to high collectivism, Hofstede, 2001) than the other cultural groups in our study, with a focus on gathering additional information when confronted with a problem situation.

In evaluating statements, Americans focus more on what is said than on how a statement is made, which implies that Americans make explicit and direct statements when communicating and making decisions (Lisø, 2011). Other cross-cultural studies have shown that North Americans are more decisive in decision making than East Asians, i.e., they can commit to an action with more ease than East Asians (e.g., Ng and Hynie, 2014), potentially due to high United States individualism (Hofstede, 2001). These findings are in line with the contention of American philosophers John Dewey and Charles Sanders Pierce, in which pragmatism is a unique American cultural feature. Based on these findings we hypothesize that Americans will be likely to search right away for applicable solutions when confronted with a complex problem situation.

To summarize, we predict that after perceiving a problem situation, Brazilians will make primarily emotional and self-related statements; Germans will engage primarily in planning; Filipinos will gather additional information; and Americans will primarily state solutions.

Materials and methods

Selection of comparison countries

We attempted to select (a) countries from different continents, (b) countries that differ widely on macro-cultural variables (e.g., gross-national product, climate) and other cultural dimensions such as cultural values (Hofstede et al., 2005; Schwartz et al., 2010), (c) countries where participants have experience working with computers, and (d) countries whose languages we, the researchers, speak and where we have existing research contacts, because we intended to travel to these countries and help with the data collection and train the researchers and coders.

Participants

In every country, participants were students from two universities who worked on both simulations (see Güss et al., 2010). In the current study we analyzed 256 WinFire think-aloud protocols and 247 Coldstore think-aloud protocols. These were 78 WinFire and 73 Coldstore protocols from Brazil, 69 WinFire and 68 Coldstore protocols from Germany, 53 WinFire and 56 Coldstore protocols from the Philippines, and 56 WinFire and 50 Coldstore protocols from the United States.

The overall age of participants ranged from 18 to 50 years old ($M = 22.66$ years, SD = 4.65). The mean age was 23.78 for the Brazilian sample (SD = 4.98), 23.01 for the German sample (SD = 3.93), 20.90 for the Filipino sample (SD = 3.11), and 22.35 for the U.S. American sample (SD = 5.76). The average age differed significantly among the four cultural groups, $F(3, 241) = 4.28$, $p = 0.006$. The Filipino sample was according to Tukey post-hoc tests significantly younger than the other samples.

Overall, 70% of the participants were female and 30% were male. In the Brazilian sample, there were 74.3% female and 25.7% male participants. In the German sample, there were 73.1% female and 26.9% male participants. In the Filipino sample, there were 65.4% female and 34.6% male participants. In the U.S. American sample, there were 67.9% female and 32.1% male participants. The gender distributions did not differ among countries, χ^2 (df = 3, $N = 246$) = 1.57, $p = 0.67$.

For the current study, we only included those more elaborate think-aloud protocols that had more than 15 transitions, since we wanted to focus on transitions of the thought process and because WinFire lasted 11 min and Coldstore lasted 13 min. Answers to questions or comments of the experimenters were not coded and did not count as transitions. Other statements were not included in the number of transitions. We also had a sample from India, but transitions for most of the Indian sample were so low that we did not include it in the current study. Thus, for this study, we analyzed 256 think-aloud protocols for WinFire with a total of 12,542 statements (78 from Brazil with a total of 4,046 statements, 69 from Germany with 4,101 statements, 53 from the Philippines with 1,755 statements, and 56 from the United States with 2,640 statements). For Coldstore, there were 247 participants with a total of 15,237 statements from the analyzed think-aloud protocols (73 from Brazil with 4,742 statements, 68 from Germany with 5,696 statements, 56 from the Philippines with 1,877 statements, and 50 from the United States with 2,922 statements).

Participants were either paid for their participation in the study or received course credit. Students in all countries were from schools of social sciences, arts and sciences, and business; and 70% of the participants were psychology majors. None of the participants had taken part in other complex problem-solving (CPS) experiments prior to this study. Responding to the demographic question about work, none of the participants indicated that they had ever been employed as a firefighter – one of the simulations, WinFire, was about fighting fires.

Instruments: Microworlds and thinking aloud

Participants worked on the two microworlds, WinFire and Coldstore. We intentionally chose these two microworlds with different task demands to not "favor" one culture over another. Winfire is a simulation that requires quick action. Coldstore is a microworld that requires cautious actions. Another factor that influenced our choice of these two simulations was the duration. Other simulations take 90–120 min and we wanted to use short simulations, as we planned recording and analyzing think-aloud protocols (Güss, 2018).

A microworld is a simulated problem in which a participant is instructed to access information and make decisions. These decisions are then implemented, the problem changes according to these decisions and other system changes, and the participant continues making decisions and solving the problem. In WinFire (Gerdes et al., 1993), participants take the role of commanding officers of a fire brigade and were instructed to save three cities and the forest from approaching fires. On the screen, participants are presented with a forest, three cities, fire-fighting trucks, helicopters, water dikes, and a cemented area. According to the criteria of microworlds, WinFire is high in complexity and dynamics, and moderate in transparency. The system consists of many variables (complexity) that are interdependent (leading to non-transparency) and that develop in a nonlinear way (dynamics). WinFire is highly complex in that it offers four main (and a few other) command options to 12 units at any given time. A person has the choice of a minimum of $4 \times 12 + 4 \times 11 + \ldots + 4 \times 1 = 312$ alternatives. WinFire is highly dynamic because the situation changes even without any intervention from the participant, such that in 11 min, 15 fires break out at programmed times. WinFire is moderate in transparency because some information between input and output are not accessible to the participant and it is hard to predict when and where fires will start and how quickly they will spread. The success criterion in WinFire is the percentage of forests that remain protected at the end of the game; the more forest was

protected the better. The simulation had 111 time intervals/cycles with 6 s each lasting in total 11 min.

In Coldstore (Reichert and Dörner, 1988, following McKinnon and Wearing, 1985), participants are required to make decisions under conditions with delayed effects. Participants take the role of a supermarket manager who must manually control the temperature of a Coldstore containing perishable goods because the automatic control device has broken down.

The Coldstore microworld can be described as low in complexity, moderate in dynamics, and low in transparency. These characterizations were validated by surveys assessing participants' subjective perceptions of the task (Güss et al., 2005). Coldstore is low in complexity because it consists of few simulated variables and the participant has only one option to intervene which involves manipulating a control wheel. Coldstore is moderate in dynamics because it is a non-linear time-delayed response to the actions of the participant. It is low in transparency because the temperature does not immediately react to changes on the control wheel. Participants must plan and then make decisions about how to reach and maintain the optimal temperature.

Aside from the thermometer, there is a manual control (described as a control wheel), with a scale from 0 to 200, that can be used to regulate the cooling system. The closer the participant manages to bring the actual temperature to the target temperature (4 degrees Celsius or 39.2 degrees Fahrenheit), the better. The success criterion in Coldstore was the sum of the deviations between the actual temperature and target temperature. The less the temperature deviates from the target temperature, the better the performance. The Coldstore simulation had 100 cycles with 8 s each lasting in total 13 min.

While participants worked on the microworlds, they were instructed to think aloud, i.e., to say out loud everything that went through their mind without interpreting or analyzing it (see Ericsson and Simon, 1993; Güss, 2018). To analyze the verbal protocols, we decided to use a microanalytic coding system. The focus of a microanalytic coding system is on all behaviors recorded in a certain time period, that is, in our case, every single sub-sentence, such as "I send the helicopter to the fire." In our system, raters from all countries coded single statements that enabled us to stay close to the data and leave little room for interpretation in our cross-national comparisons. The microanalytic coding system's rigor worked for our purposes in terms of detailing the categories and differentiating these from one country to another (see Supplementary Appendix A online for the coding system with the 10 main categories and think-aloud examples;). Whereas the previously published results of these think-aloud protocols focused on country differences regarding the frequencies of the categories, for example how often did participants gather information in Brazil, India, Germany, the United States, and the Philippines (Güss et al., 2010), the current study was focused on the process. In our analysis of the protocols, we aimed to answer the question: During the problem-solving process, which step follows which step; and particularly which step follows the perception of a problem?

Procedure

The University of North Florida Institutional Review Board together with the ethics committees in every country – where required – approved the research study in every country. Informed consent was obtained from all individual participants prior to beginning the study. The duration of each session was ~1.5 h, and all participants were tested individually with an experimenter present. The instructions and administration conditions were identical for each country. WinFire extended for 11 min; Coldstore for 13 min. After signing informed consent forms, participants received a three-page printed introduction to WinFire and a two-page introduction to Coldstore on the computer screen. These introductions were translated into German and Brazilian Portuguese using a translation-back-translation procedure as discussed by Brislin (1970). Participants played test games as practice for 3 min before each microworld to gain familiarity with the screens and commands and to practice thinking aloud. Experimenters, who were seated about three feet obliquely beside the participant, reminded the participant to continue to think aloud (i.e., "keep talking," "continue to say what's going through your mind"). Experimenters were instructed to give no more than seven reminders.

Experimenters were professors and an additional three to four research assistants from each country. The last author went to each of the countries and was always part of the data collection process. Participants spoke in their native tongues. Participants in the Philippines were offered the option of speaking in their native languages if they preferred, that is, Tagalog or Ilocano (for a critical discussion of the think-aloud method in cross-cultural research, see Güss, 2018). All protocols were analyzed in the language in which they were spoken and only translated to report results.

Data analysis

Every statement in the think-aloud protocols was tape-recorded, transcribed into Microsoft Excel, and coded per idea unit by the professors, who were fluent in Brazilian Portuguese, German, English, Tagalog, and Ilocano, and by several student research assistants from each country. For example, the statement "I send helicopter 3 to the big forest fire and truck 2 to the city" has two idea units; the first referring to the helicopter, and the second referring to the truck. Raters who were not authors participated in about 10 h of training on the coding system. The following example statement has two different idea units: "I send truck 5 to the city, and then the helicopter to the water dike." Statements triggered by the experimenter's comments were not included in further analyses.

A coding system was developed following a theory-driven top-down and a data-driven bottom-up approach, resulting in the 10 main categories (Güss et al., 2010): (1) situation description (SD), (2) problem identification (PI), (3) formulation of goals

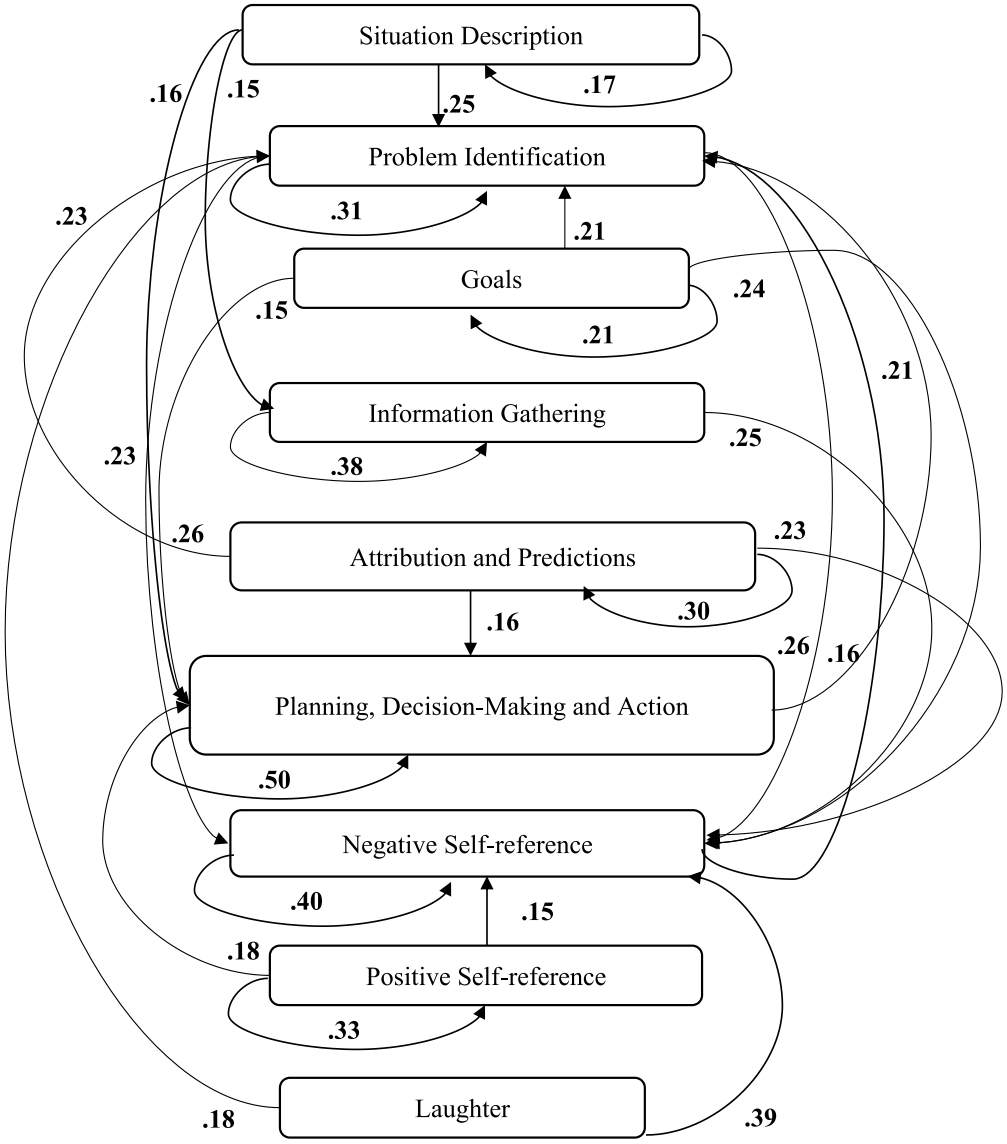


FIGURE 1
Brazil – WINFIRE conditional probabilities greater 15 percent with significant z-score values.

(GO), (4) information gathering (INFO), (5) attributions and predictions (ATPR), (6) planning, decision making, and action (PLDM), (7) positive self-evaluations and emotions (SR+), (8) negative self-evaluations and emotions (SR−), (9) laughter (L), and (10) other (O).

The same coding system was used in each country for the analysis of both the WinFire and the Coldstore data (see Supplementary Appendix A). The coding system consisted of the main categories with 2–4 subcategories under each, resulting in 22 subcategories. One additional category--hypotheses about the system's functioning – was added to the Coldstore coding. These 22 subcategories were then summarized in the 10 main categories: eight main steps (Situation description – SD, Problem identification – PI, Formulation of goals – GO, Gathering of information – Info, Attributions and predictions – ATPR, Planning, decision making, and action – PLDM, Positive self-evaluations and emotions – SR+, Negative self-evaluations and emotions – SR−), plus the Laughter and Other category (see Supplementary Appendix A). A sample of participants' verbal protocols with their assigned codes is presented in the Supplementary Appendix online.

Raters in every country participated in about 10 h of training on the coding system where also specific examples were discussed. For example, the following statement has two different idea units: "I send the helicopter to the city / /and then the truck to the lake," coded each with the category "Planning, Decision Making, and Action" (PLDM). Initial calculations of inter-rater reliability using Cohen's Kappa (Cohen, 1960) were conducted on a sample of

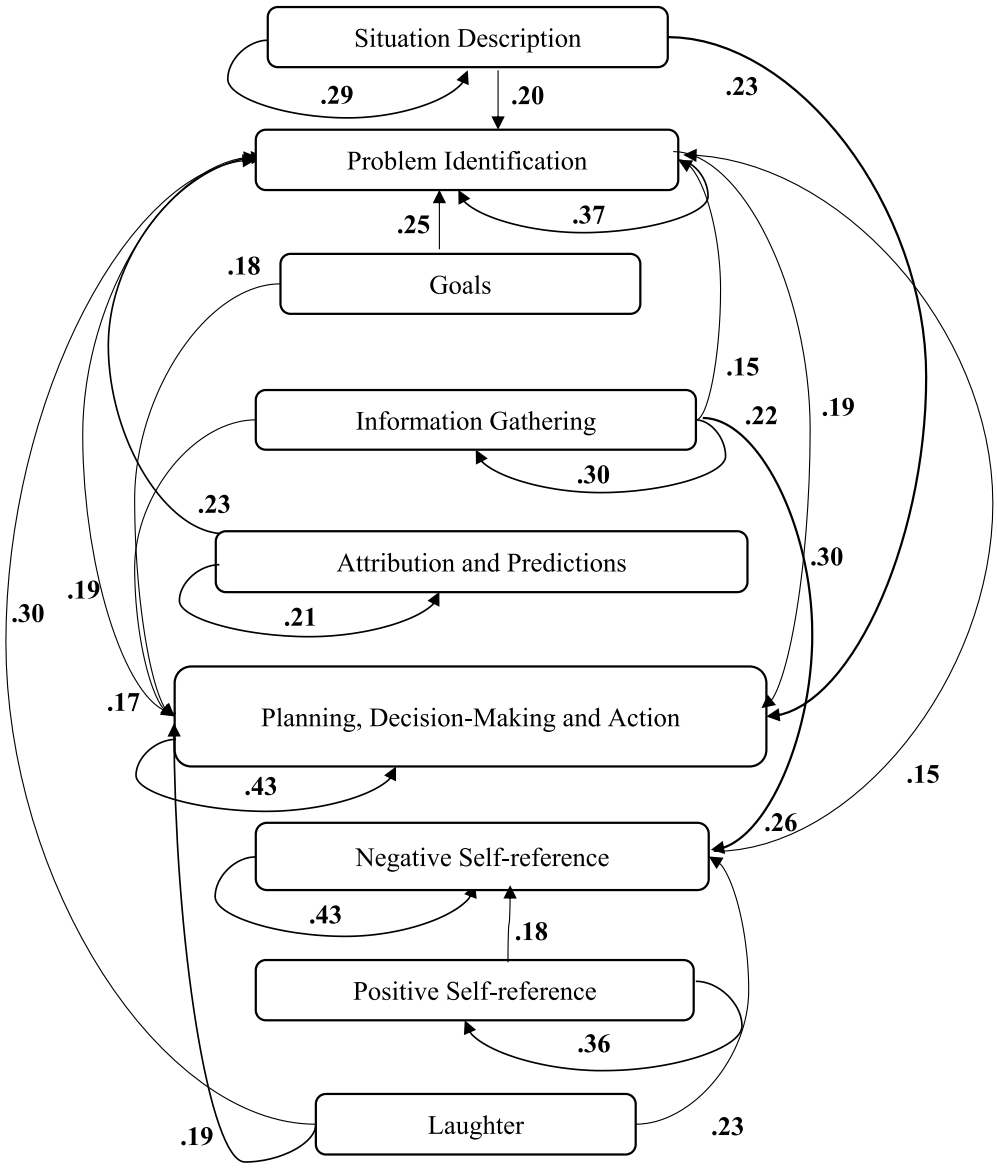


FIGURE 2
Brazil – COLDSTORE conditional probabilities greater 15 percent with significant z-score values.

think-aloud protocols from each country to check the credibility of the coding system for the think-aloud protocols of WinFire and Coldstore and the coders' training (see Güss et al., 2010). According to Fleiss (1981), Kappas over 0.75 are excellent, between 0.60 and 0.75 are good, and between 0.40 and 0.60 are fair. Inter-rater reliability was 0.59 (Brazil), 0.60 (United States), 0.66 (Philippines), and 0.83 (Germany). Disagreements were resolved through discussion among the coders that was aimed at seeking consensus. The resulting Kappas ranged from good to excellent (Fleiss) and were considered satisfactory for such a complex coding system and such heterogeneous samples.

A computer program was created by Bhattacharya (2019) that read the transcribed think-aloud and coded think-aloud files of every participant saved in the Excel program for WinFire and Coldstore and then calculated the transition frequencies for each step. For example, how often were goals (GO) formulated after a problem was identified (PI)? How often did planning (PLDM) happen after a problem was identified (PI)? All statements and transitions that related to responses of experimenter interventions or "other" statements (e.g., repeating exact same statement, the first statement as a response to experimenter statement, "Mmmmh," "Ahhh") were excluded. One output file shows the transitions among the 22 subcategories, the other output file shows the transitions among the nine main categories for each WinFire and Coldstore (see Coding System in Supplementary Appendix A). The transitional probability (TP) – also called

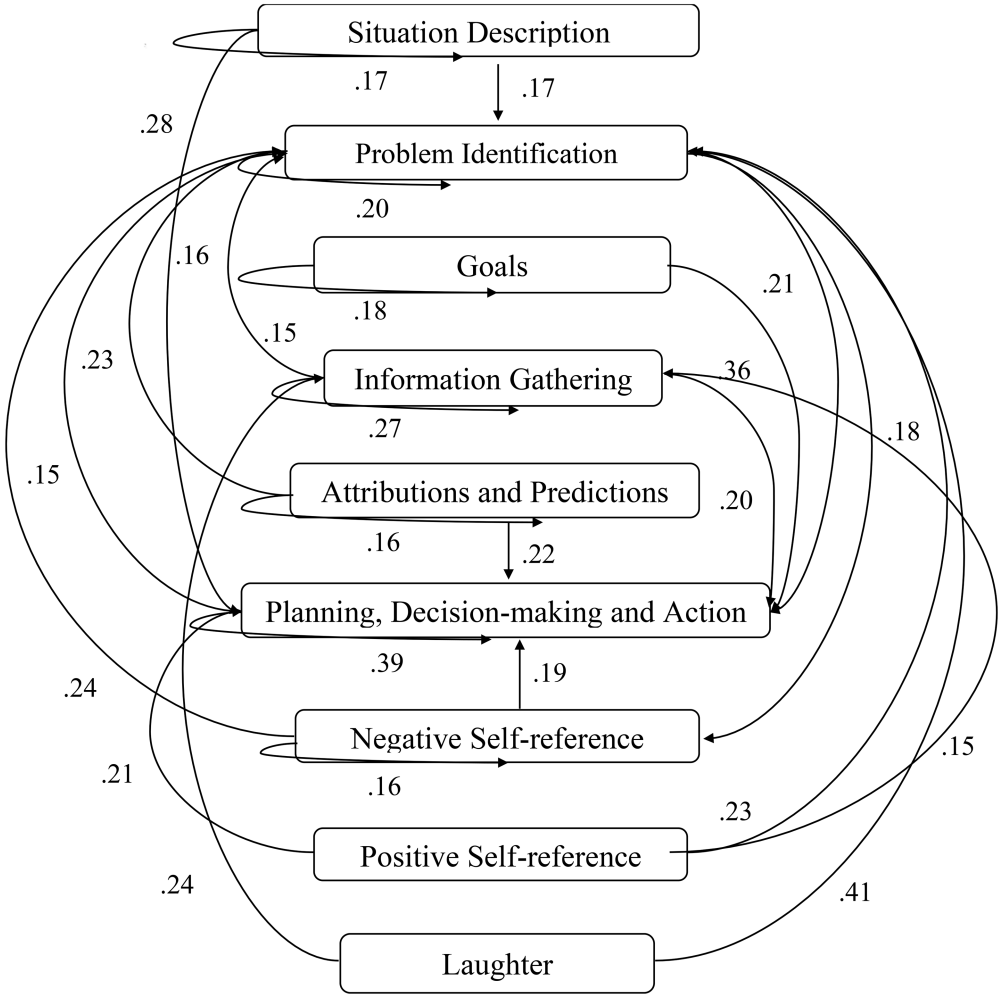


FIGURE 3
Germany – WINFIRE conditional probabilities greater 15 percent with significant z-score values.

conditional probability – from any category x to another category y is given by TP (x → y) = frequency (xy)/frequency (x). These analyses are also called lag analysis or latent transition analysis (Collins and Lanza, 2010). Transition frequencies were summarized and averaged for each country for WinFire and Coldstore (see Figures 1–8). The figures show only transitions >15% for easier comprehension. In order to calculate significant transitions (see Gottman and Roy, 1990), we also calculated observed frequencies, conditional probabilities, expected frequencies, and z-scores for each microworld and each country (Data for this study can be found in the Supplementary material section, Supplementary Appendix B). Z-scores less than or equal to the value of −1.645 and z-scores higher or equal to the value of 1.645 are significant at the 0.05 level.

We then used these data files on conditional/transitional probabilities to represent communities in a graph. "A community is defined as a subset of nodes within the graph such that connections between the nodes are denser than connections with the rest of the network" (Radicchi et al., 2004). We used the spinglass algorithm in R, igraph, to graph these communities (Reichardt and Bornholdt, 2006, see Supplementary Appendix C for the R code).

Results

The four hypotheses were as follows: After presenting participants with problem scenarios, Brazilians will emphasize making emotional and self-related statements (SR−), Germans will engage primarily in planning (PLDM), Filipinos will focus on gathering additional information (INFO), and Americans will primarily state solutions (PLDM). Statement transitions over 15% for each country are shown in Figures 1–10 for WinFire and Coldstore (Figures 11–24).

We compared the percentages in the three most common transitions from Problem identification (PI) to Problem

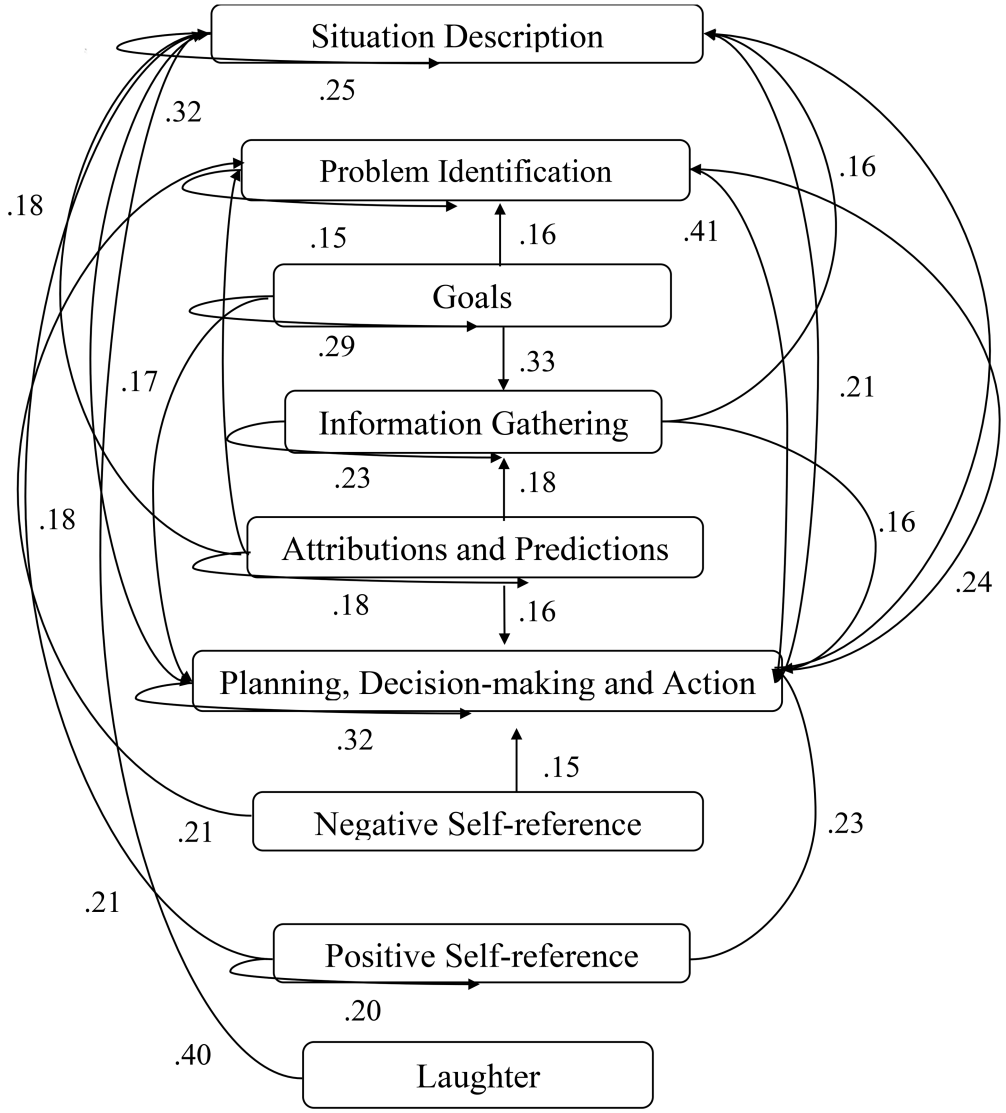


FIGURE 4
Germany – COLDSTORE conditional probabilities greater 15 percent with significant z-score values.

identification; from Problem identification to Negative self-reference (SR−); and from Problem identification to Planning, decision making, and action (PLDM) in WinFire. The percentages did not differ significantly among the four countries in WinFire, χ^2 (6) = 6.27, $p > 0.05$, ns. For WinFire (see Table 1), Brazilians made primarily Problem identification (31%) and Negative self-reference (23%) statements after problem identification (PI). Germans had the most statement transitions regarding "Planning, decision making, and action" (21%) followed by Problem identification (20%) after PI. Filipinos showed primarily Problem identification (29%) and Negative self-reference (18%) statements after PI. US participants made primarily statements regarding Problem identification (26%) and Negative self-reference (25%) after PI.

We also compared the percentages in the three most common transitions from Problem identification (PI) to Problem identification; from Problem identification to Negative self-reference (SR−); and from Problem identification to Planning, decision making, and action (PLDM) in Coldstore. The percentages differed significantly among the four countries in Coldstore, χ^2 (6) = 30.03, $p < 0.001$. For Coldstore (see Table 1), the most frequent problem-solving statements after problem identification (PI) for Brazilians were Problem identification (37%) and Planning, decision making, and action (19%). Germans made mostly Planning, decision making and action (41%) and Problem identification (15%) after PI. Filipinos made mostly Problem identification (29%) and Information gathering (13%) and Negative self-reference (13%) statements after PI. The US participants showed

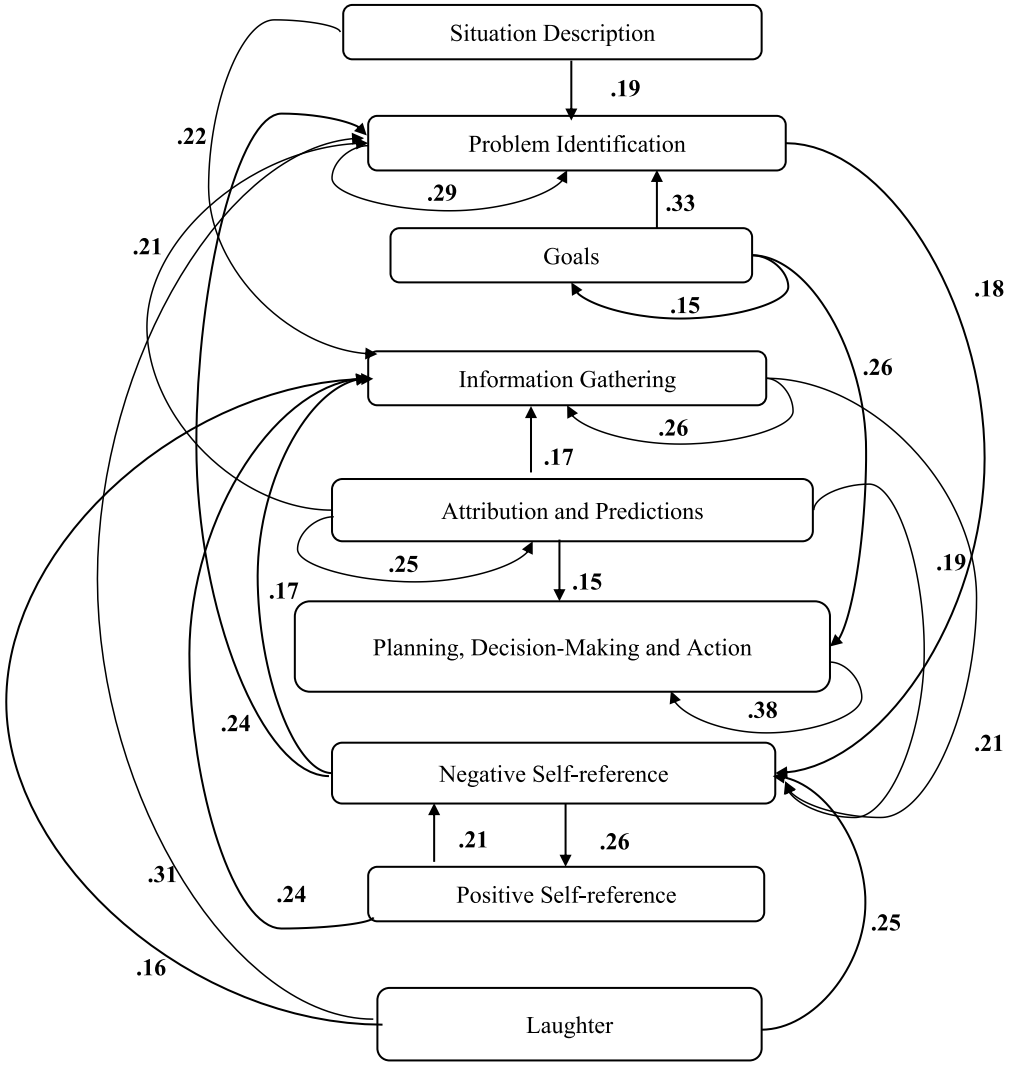


FIGURE 5
Philippines – WINFIRE conditional probabilities greater 15 percent with significant z-score values.

statements regarding Planning, decision making, and action (26%) and problem identification (22%) after PI.

We then compared the same three transition percentages between Winfire and Coldstore. They differed significantly between the two simulations, χ^2 (2) = 13.82, $p < 0.001$. The transition to Problem identification was similar in both simulations, but more Negative self-reference (SR−) transitions were found in Winfire and more Planning, decision making, and action (PLDM) transitions were found in Coldstore.

Additionally, we run further analyses in the R-program on state-transitions of the complex problem-solving steps. For COLDSTORE, Brazilian, German, and Filipino data showed each 3 community clusters, and US data showed 4 community clusters. For all countries, the two steps Problem identification (PI) and Planning, decision making, and action (PLDM) were always in the same cluster. COLDSTORE data from all four countries have in common that negative self-reflections and emotions (SR−) were always in the same cluster as information gathering (INFO).

For WINFIRE, Brazilian, Filipino, and US data showed 3 community clusters, German data showed 2 community clusters. Only for the Brazilian and US samples, the two steps PI and PLDM were in the same cluster. For WINFIRE, SR− was only in the same cluster as INFO in one sample, the US sample.

The clusters containing PI varied among cultures. For the German sample, for example, only one other step in COLDSTORE and two other steps in WINFIRE were in the same cluster as PI. For the US sample, 3 other steps in COLDSTORE and 4 other steps in WINFIRE were in the same cluster as PI.

From a cognitive perspective, it is surprising that negative emotions and self-references (SR−) were always central

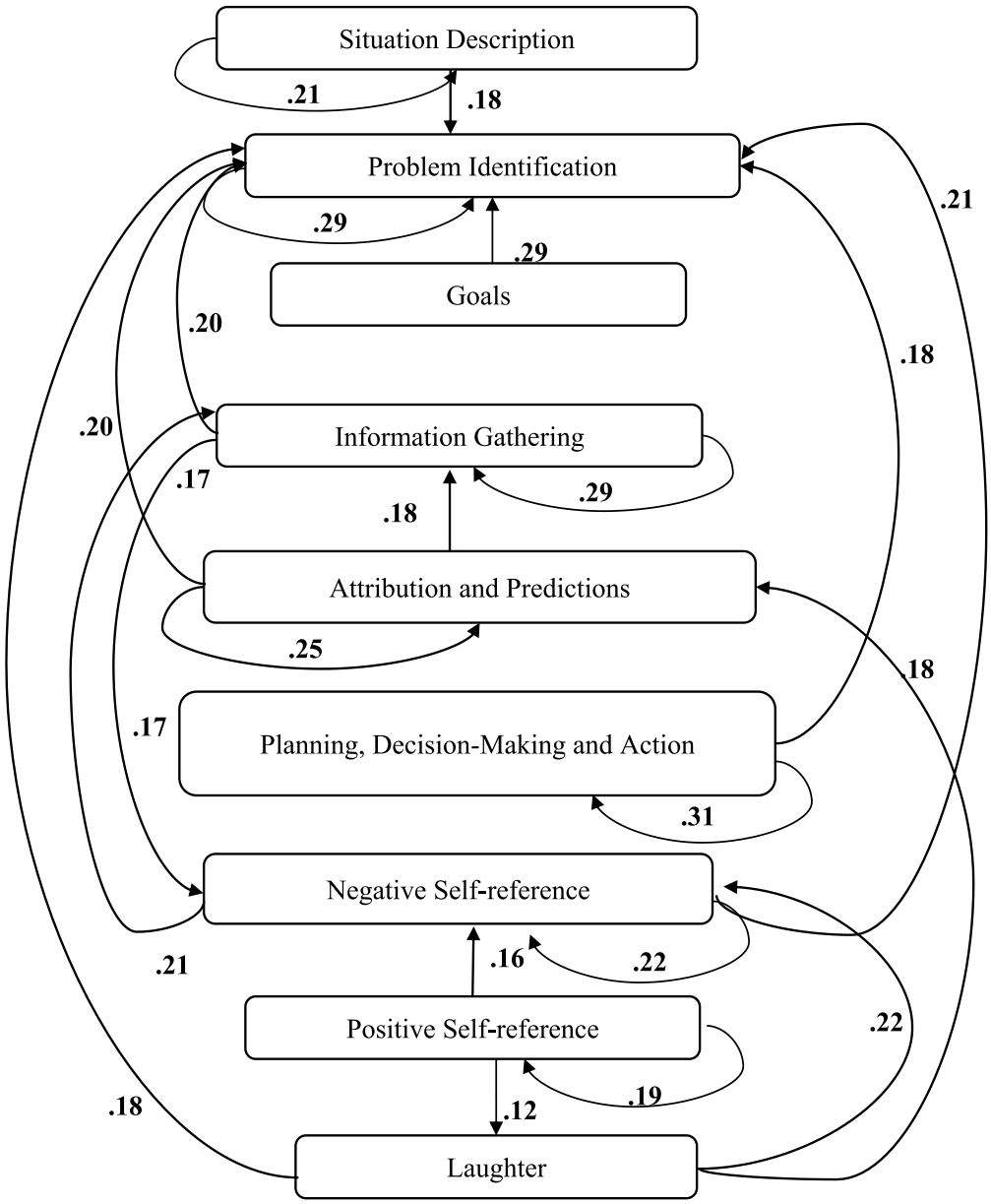


FIGURE 6
Philippines – COLDSTORE conditional probabilities greater 15 percent with significant z-score values.

among the problem-solving clusters. In complex situations, negative emotions seem to provide situational impressions that can then motivate and lead to modifications in the problem-solving approach.

Discussion

The purpose of this study was to examine the process of complex problem-solving and cultural differences among Brazil, Germany, the Philippines, and the United States. Our hypotheses included, Brazilians making primarily emotional and self-related statements, Germans engaging primarily in planning, Filipinos gathering additional information, and Americans stating solutions.

The first hypothesis regarding Brazil was partially supported. Data showed 23% of Brazilians making negative self-references (SR−) following problem identification (PI) in WinFire. These findings are partially consistent with previous research showing the high emotional expressiveness of Brazilians ranking number 10 out of 32 countries (Matsumoto, 2008). However, in both WinFire and Coldstore, Brazilians continue to elaborate on the problem situation with 37% of Brazilians stating problem identification statements (PI) after

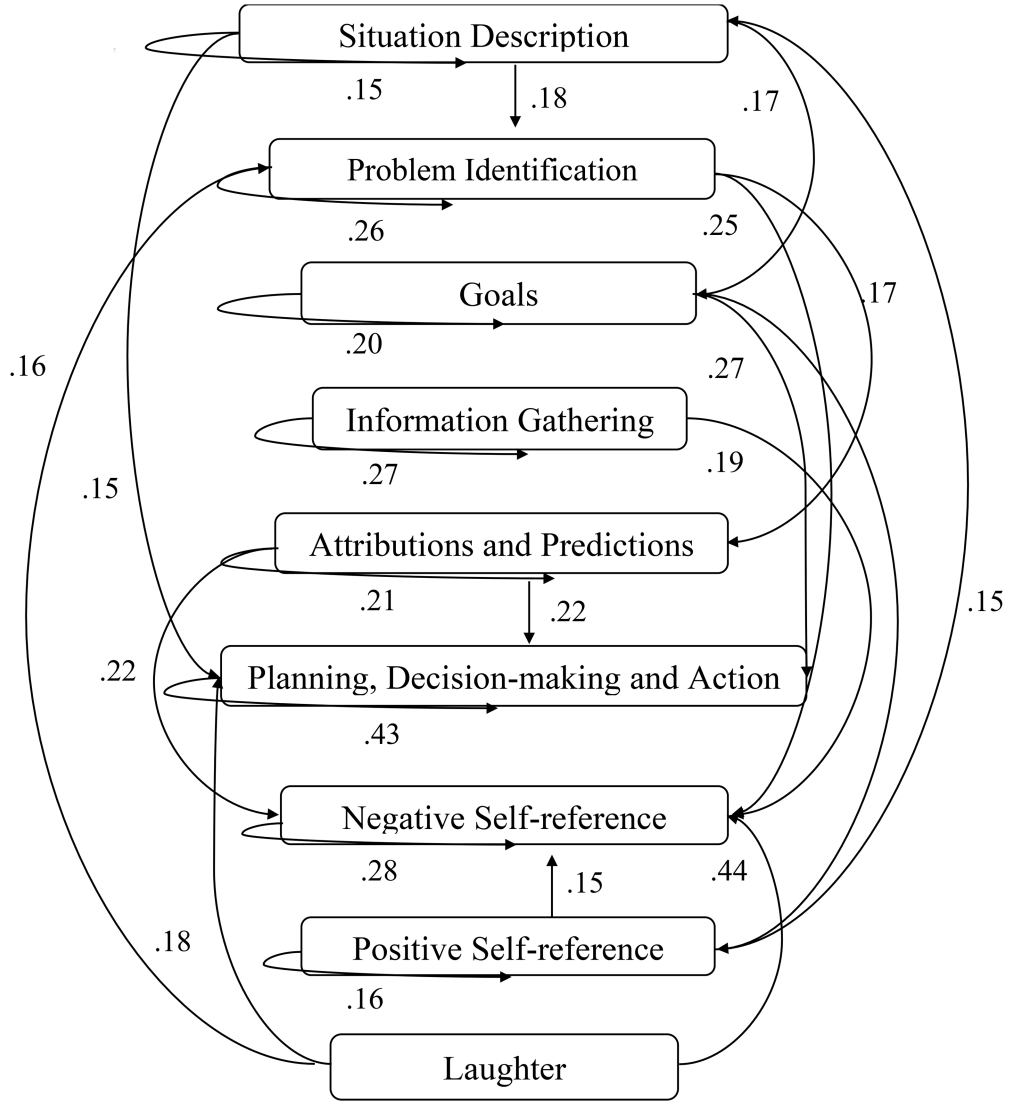


FIGURE 7
United States – WINFIRE conditional probabilities greater 15 percent with significant z-score values.

problem identifications in Coldstore and 31% in WinFire. Thus, elaborating on the problem is the most frequent transition.

The second hypothesis regarding German participants was supported; 21% of German participants responded to problem identification with planning, decision making and action (PLDM) in WinFire, and 41% of German participants responded with planning, decision making, and action (PLDM) to problem identification (PI) in Coldstore. These results are consistent with previous research showing the importance of planning in the German culture (e.g., Strohschneider and Güss, 1998; Hofstede, 2001). Additionally, these results reflect Germans' desire for thinking, planning ahead, and taking action in the opportunities given (Tipandjan et al., 2012).

The third hypothesis regarding Filipino participants was partially supported, which showed 29% of Filipinos in WinFire and 29% of Filipinos in Coldstore were involved in problem identification (PI) after problem identifications and 13% made statements related to information gathering (INFO) in Coldstore. Problem identification is a part of gathering further information; one elaborates on the problem, further analyzes it, and tries to understand it. It shows a more cautious approach to problem solving. These results are consistent with the analysis of dynamic decisions (and not thinking aloud) in Coldstore where Filipinos, in comparison to the participants from the other countries, showed the most cautious approach (Güss and Dörner, 2011). These results are also supported by Builtjens and Noorderhaven (1996) who showed that the cautious decision-making process is partially due to a culture which is highly context dependent.

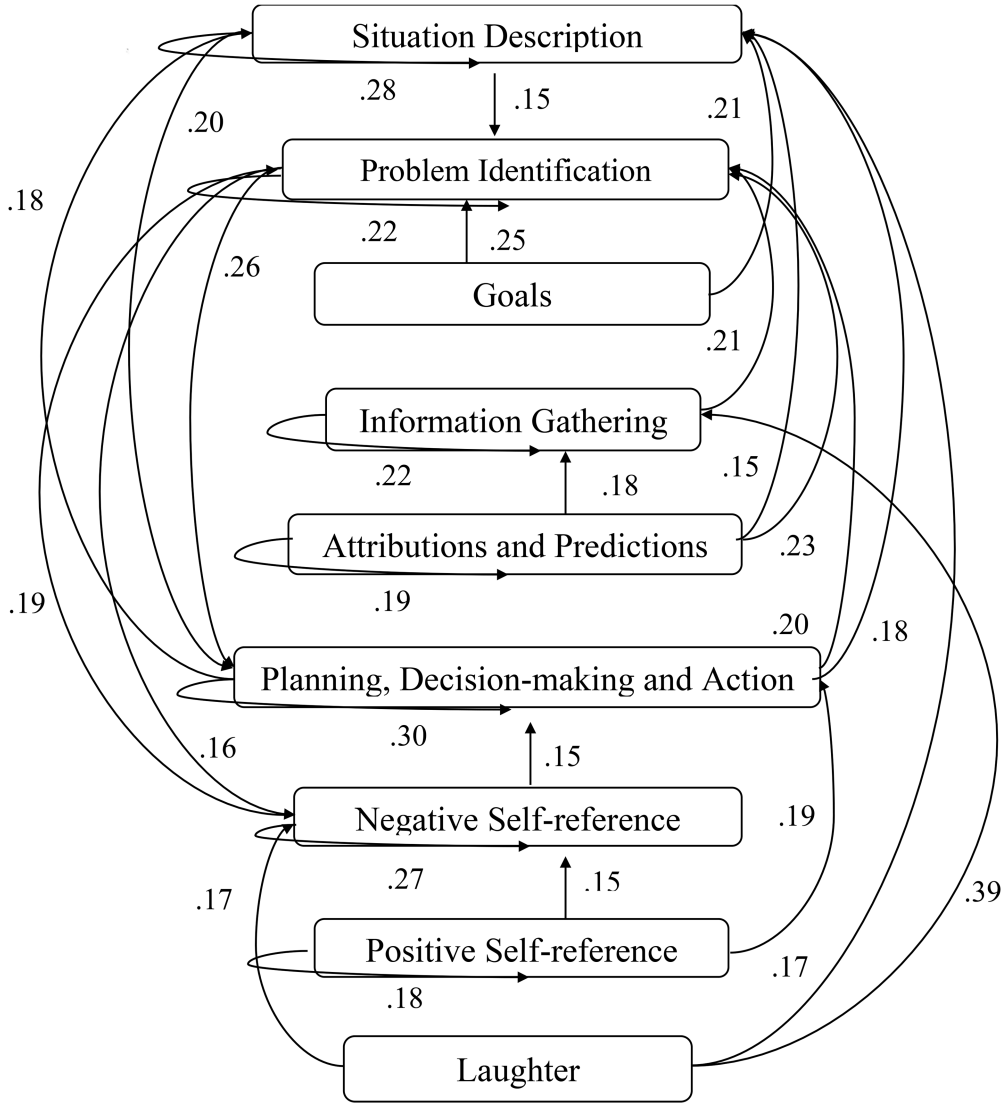


FIGURE 8
United States – COLDSTORE conditional probabilities greater 15 percent with significant z-score values.

The fourth hypothesis regarding US Americans was only supported for Coldstore which showed 26% of American participants involved in planning, decision making, and action (PLDM) after identifying a problem. Further problem identification played an important role in WinFire with 26% and in Coldstore with 22% of the statements following problem identification (PI). The results reveal a need to understand the problem situation better, before making decisions. Other research reveals Americans being very conscious about the decision-making process as well as making choices depending on the information that is given instead of the loyalty of a company or brand (Leng and Botelho, 2010).

Summarizing the results focusing on the simulation, we find for WinFire, that in three of the four countries (Brazil, Philippines, United States), the primary reactions to recognizing a problem were problem-related statements and negative self-references, i.e., negative emotions and negative statements regarding one's self-wroth (SR−). Only German participants showed primarily planning and decision making (PLDM) after problem identification (PI). The countries did not differ significantly in their transition percentages.

For Coldstore, the countries differed significantly in their transition percentages. The most frequent problem-solving statements after problem identification (PI) in three of the four countries (Brazil, Germany, United States) were planning, decision making, and action (PLDM); and problem identification – in different order though. Filipino participants also made statements regarding problem identification, but also regarding information gathering (INFO) and negative self-references (SR−).

Two general findings are worth mentioning contradicting our predictions. First, in both situations, the most frequent

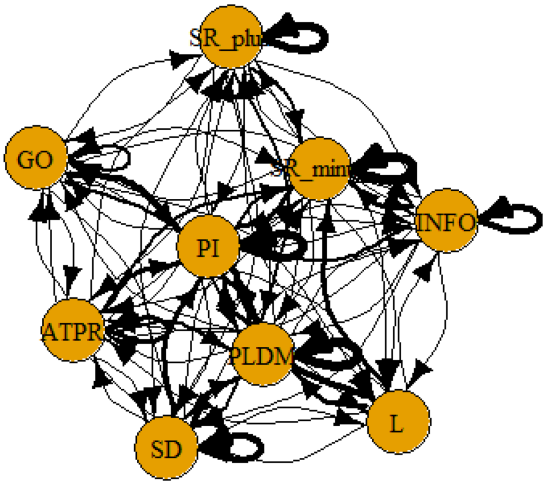


FIGURE 9
Brazil – COLDSTORE, all conditional probabilities.

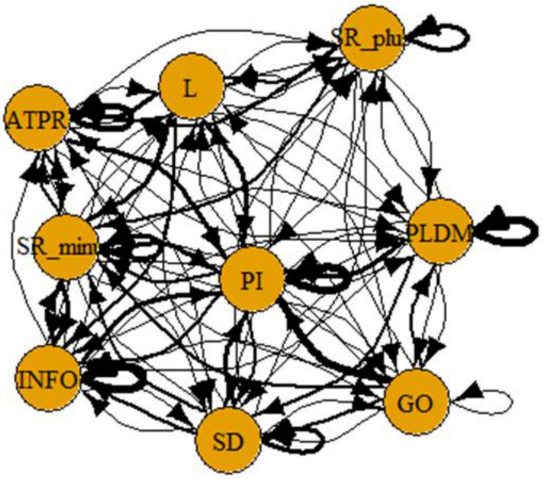


FIGURE 11
Philippines – COLDSTORE, all conditional probabilities.

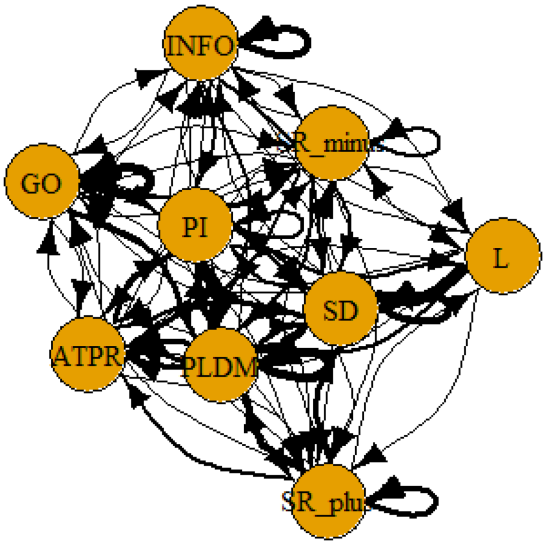


FIGURE 10
Germany – COLDSTORE, all conditional probabilities.

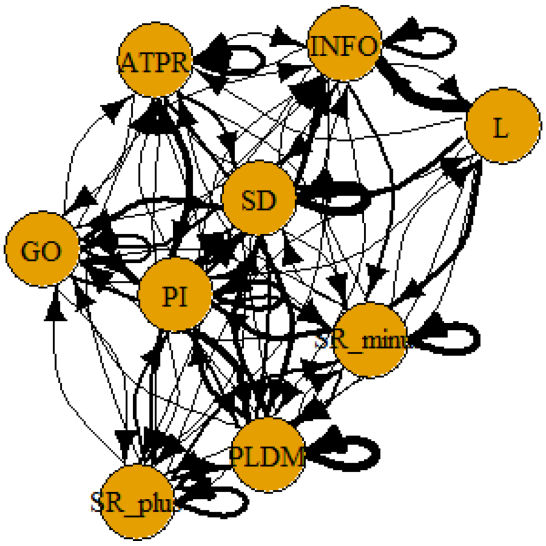


FIGURE 12
United States – COLDSTORE, all conditional probabilities.

statements following problem identification (PI) are further problem identification statements (PI), negative self-references (SR−), and planning, decision making, and action (PLDM). Most participants, regardless of their cultural background, further elaborate on the problem situation, but then show either negative emotions such as stress or frustration or show self-derogatory statements like "I will never be a good fire fighter" or "I am not good at this." before they go on with further planning and decision making. Contradicting the rational "homo economicus" decision-making model, people are affected by complex and dynamic problems: they often react overwhelmed, stressed, and their feeling of competence is attacked. Emotions, motivations, and cognition interact when people deal with complex problems (see Dörner and Güss, 2013; Dörner and Güss, 2022).

Second, there were no cultural differences in WinFire regarding transitions after problem identification (PI). In Coldstore, however, we found significant differences and see more cultural differences in the transitions after problem identification (PI). One possible explanation is related to the transparency and uncertainty related to the two simulations. In WinFire, it is obvious when a fire starts, where it starts, and if it is an imminent

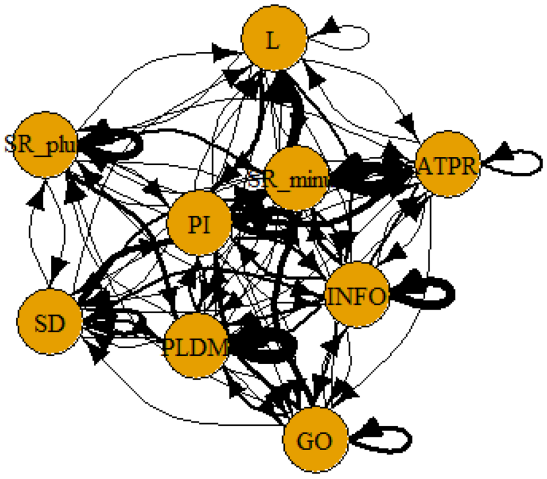


FIGURE 13
Brazil – WINFIRE, all conditional probabilities.

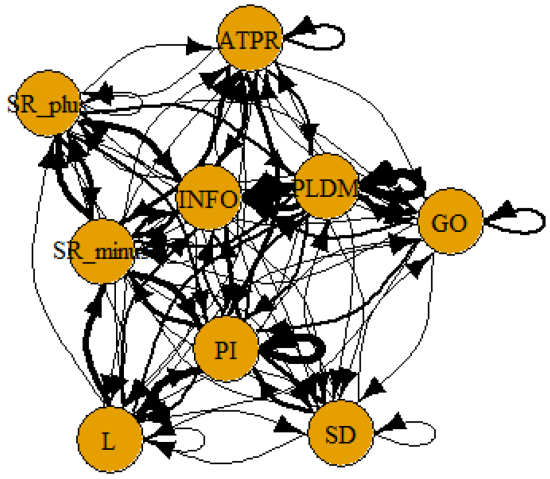


FIGURE 15
Philippines – WINFIRE, all conditional probabilities.

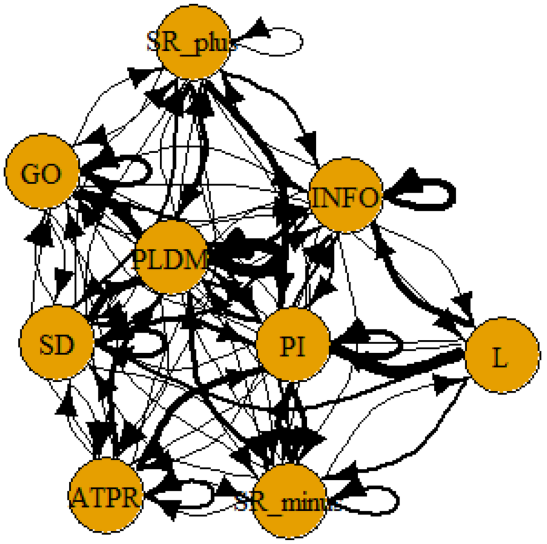


FIGURE 14
Germany – WINFIRE, all conditional probabilities.

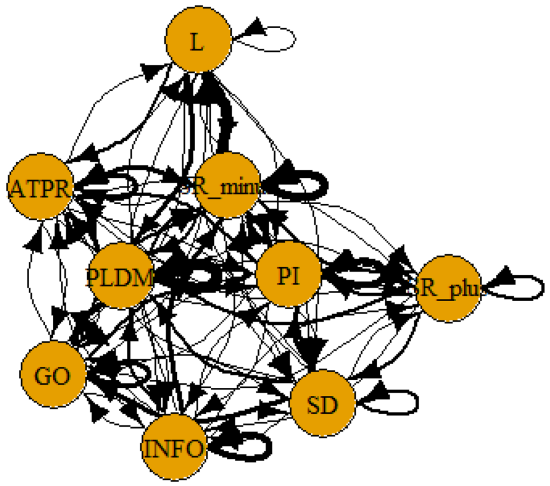


FIGURE 16
United States – WINFIRE, all conditional probabilities.

threat to a city or not. It is obvious how distant trucks and helicopters are from the fires. Wind strength and direction are also shown. Thus, even though the problem is highly dynamic, it is relatively transparent. Coldstore, however, is more uncertain. It is unclear why the temperature oscillates, and why the temperature does not seem to react to the changes on the control wheel. This uncertainty leaves room for interpretation and for discovery and thus allows the cultural decision-making strategies to shine. This discussion about the different task demands of WinFire and Coldstore is also reflected in the significant differences of the transition probabilities between the two simulations in the three most frequent steps (from Problem identification PI to Problem identification PI, to Negative self-reference SR−, and to Planning, decision making, and action PLDM.) The transition to Problem identification was similar in both simulations, but more Negative self-reference transitions (SR−) were found in Winfire and more Planning, decision making, and action (PLDM) transitions were found in Coldstore.

Our study also has limitations. The first limitation is within-country diversity. It is apparent that within every culture there also exist various subcultures and we only include two student samples from different cities in each country. Another limitation is related to the think-aloud method. For some participants, it was unusual and

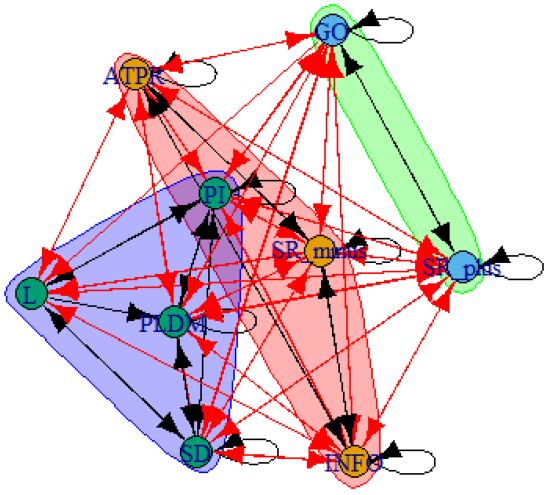


FIGURE 17
Brazil – COLDSTORE: Community clusters for transitional probabilities.

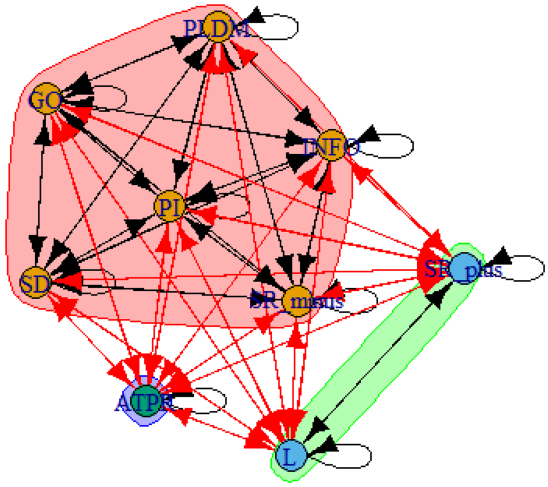


FIGURE 19
Philippines – COLDSTORE: Community clusters for transitional probabilities.

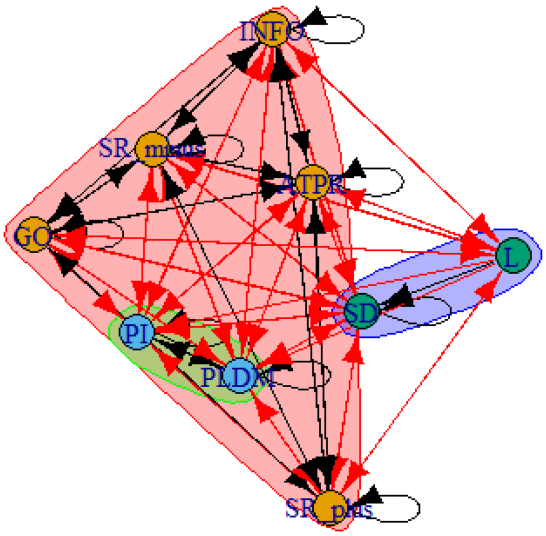


FIGURE 18
Germany – COLDSTORE: Community clusters for transitional probabilities.

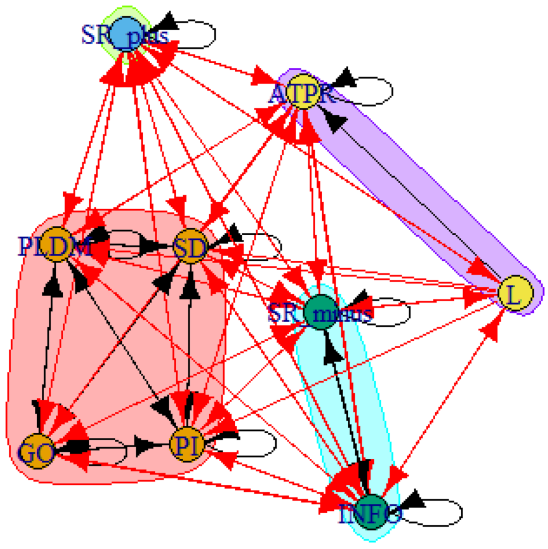


FIGURE 20
United States – COLDSTORE: Community clusters for transitional probabilities.

uncomfortable to talk aloud, especially for the Filipino participants (see also differences in think-aloud between European and Asian Americans, Kim, 2002; on language and reasoning see also Papafragou et al., 2007). Future research could, for example, consider group-versus individual-related differences in complex problem solving across various cultural regions. A third limitation is related to the spinglass community detection method in igraph, as items (transition frequencies in our case) deterministically belong to one community only.

In sum, this study showed an in-depth log-linear analysis of the complex and dynamic problem-solving strategies in four countries during two different simulated tasks. Findings showed unique cultural preferences in strategies when identifying problems when the task was uncertain, but not when the task was relatively transparent. It is the cultural upbringing in our mind that triggers preferred ways in dealing with uncertain problem situations.

Data availability statement

The original contributions presented in the study are included in the article. Data for this study can be found in the

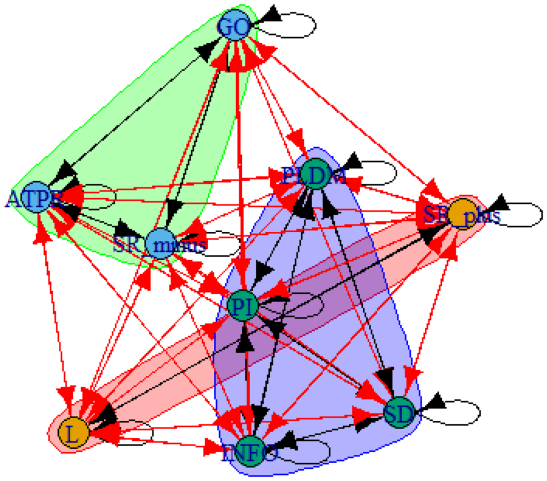


FIGURE 21
Brazil – WINFIRE: Community clusters for transitional probabilities.

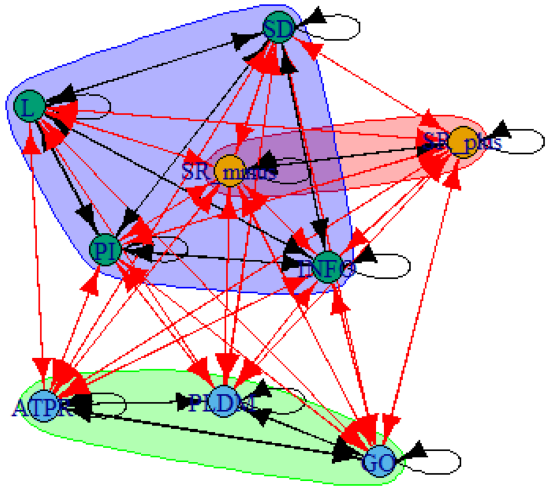


FIGURE 23
Philippines – WINFIRE: Community clusters for transitional probabilities.

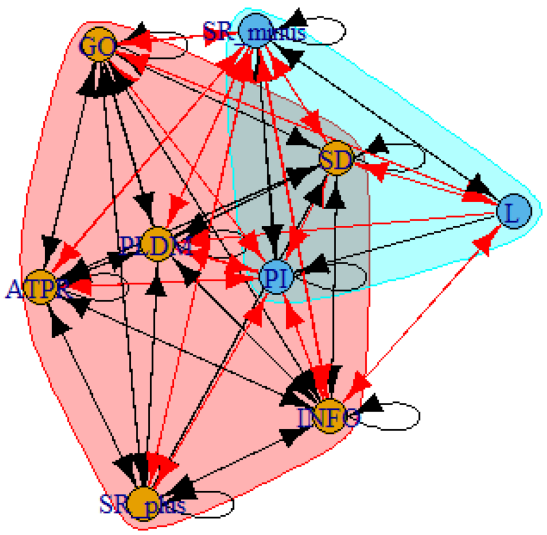


FIGURE 22
Germany – WINFIRE: Community clusters for transitional probabilities.

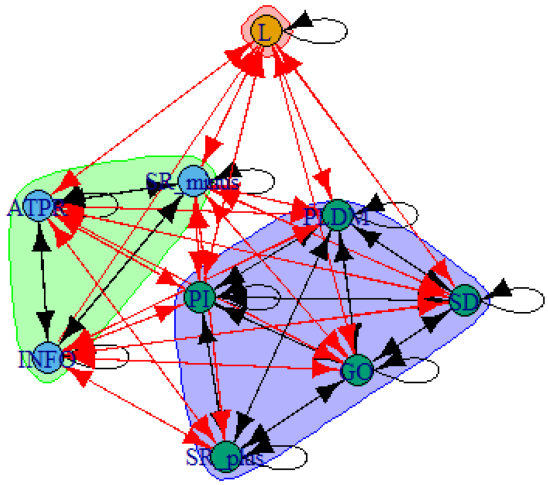


FIGURE 24
United States – WINFIRE: Community clusters for transitional probabilities.

Supplementary material section. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by The University of North Florida Institutional Review Board together with the ethics committees in every country – where required. The participants provided their written informed consent to participate in this study.

Author contributions

WS and JH were primarily analyzing the transition data and creating the figures. CG was conducting the R-analyses, creating the figures in R, and writing most parts of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported in part by the National Science Foundation grant #0218203 to the last author and by the University of North Florida's John A. Delaney Presidential Professorship to the last author.

TABLE 1 Highest average statement transition percentages stemming from problem identification for WINFIRE and COLDSTORE.

Country	WINFIRE		COLDSTORE	
	Highest statement transition following Problem Identification	Statement transition percent	Highest statement transition following Problem Identification	Statement transition percent
Brazil	Problem identification	31%	Problem identification	37%
	Negative self-reference	23%	Planning, decision making, and action	19%
	Planning, decision making, and action	14%	Negative self-reference	15%
Germany	Planning, decision making, and action	21%	Planning, decision making, and action	41%
	Problem identification	20%	Problem identification	15%
	Negative self-reference	18%	Situation description	12%
Philippines	Problem identification	29%	Problem identification	29%
	Negative self-reference	18%	Information gathering	13%
	Laughter	13%	Negative self-reference	13%
United States	Problem identification	26%	Planning, decision making, and action	26%
	Negative self-reference	25%	Problem identification	22%
	Planning, decision making, and action	17%	Negative self-reference	16%

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.965623/full#supplementary-material

References

Bhattacharya, A. (2019). *LagAn: Computer Program for Lag Analyses; to Analyze Transitions in Think-Aloud Protocols*. Jacksonville, FL: University of North Florida.

Bransford, J. D., and Stein, B. S. (1993). *The Ideal Problem Solver* (*2nd ed*). New York, NY: Freeman.

Brehmer, B., and Dörner, D. (1993). Experiments with computer-simulated microworlds: escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Comput. Hum. Behav.* 9, 171–184. doi: 10.1016/0747-5632(93)90005-D

Brislin, R. W. (1970). Back translation for cross-cultural research. *J. Cross-Cult. Psychol.* 1, 185–216. doi: 10.1177/135910457000100301

Builtjens, R. P. M., and Noorderhaven, N. (1996). *The Influence of National Culture on Strategic Decision Making: A Case Study of the Philippines.* Tilburg, Netherlands: Institute for Research on Intercultural Cooperation.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Cole, M. (1996). *Cultural Psychology. A Once and Future Discipline*. Cambridge, MA: Belknap and Harvard University Press.

Collins, L. M., and Lanza, S. T. (2010). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. Hoboken, NJ: Wiley.

Dörner, D. (1996). *The Logic of Failure*. New York, NY: Holt.

Dörner, D., and Funke, J. (2017). Complex problem solving: what it is and what it is not. *Front. Psychol.* 8:1153. doi: 10.3389/fpsyg.2017.01153

Dörner, D., and Güss, C. D. (2013). PSI: a computational architecture of cognition, motivation, and emotion. *Rev. Gen. Psychol.* 17, 297–317. doi: 10.1037/a0032947

Dörner, D., and Güss, C. D. (2022). Human error in complex problem solving and dynamic decision making: a taxonomy of 24 errors and a theory. *Comput. Hum. Behav. Rep.* 7:100222. doi: 10.1016/j.chbr.2022.100222

Ericsson, K. A., and Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data* (*Rev. ed.*). Cambridge, MA: The MIT Press.

Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions*. New York, NY: Wiley.

Fox, J., Cooper, R. P., and Glasspool, D. W. (2013). A canonical theory of dynamic decision-making. *Front. Psychol.* 4:150. doi: 10.3389/fpsyg.2013.00150

Frensch, P. A., and Funke, J. (Eds.) (1995). *Complex Problem Solving: The European Perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Funke, J. (2010). Complex problem solving: a case for complex cognition? *Cogn. Process.* 11, 133–142. doi: 10.1007/s10339-009-0345-0

Galotti, K. M. (2002). *Making Decisions that Matter. How People Face Important Life Choices*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Gerdes, J., Dörner, D., and Pfeiffer, E. (1993). *Interaktive Computersimulation "WINFIRE" [the interactive computer simulation "WINFIRE"]*. Germany: Lehrstuhl Psychologie II: Otto-Friedrich-Universität Bamberg.

Gonzalez, C., Fakhari, P., and Busemeyer, J. (2017). Dynamic decision making: learning processes and new research directions. *Hum. Factors* 59, 713–721. doi: 10.1177/0018720817710347

Gonzalez, C., and Martin, J. M. (2011). "Dynamic decision making and cultural affiliation," in *Cultural Factors in Systems Design*. eds. R. W. Proctor, S. Noff and Y. Yih (Boca Raton, FL: CRC Press), 43–62.

Gottman, J. M., and Roy, A. K. (1990). *Sequential Analysis: A Guide for Behavioral Researchers*. Cambridge: Cambridge University Press.

Greiff, S., and Scherer, R. (2018). Complex problem solving and its position in the wider realm of the human intellect. *J. Intelligence* 6, 1–4. doi: 10.1027/1015-5759/a000487

Güss, C. D. (2018). What is going through your mind? Thinking aloud as a method in cross-cultural psychology. *Front. Psychol.* 9:1292. doi: 10.3389/fpsyg.2018.01292

Güss, C. D., Burger, M. L., and Dörner, D. (2017). The role of motivation in complex problem solving. *Front. Psychol.* 8:851. doi: 10.3389/fpsyg.2017.00851

Güss, C. D., and Dörner, D. (2011). Cultural differences in dynamic decision-making strategies in a non-linear, time-delayed task. *Cogn. Syst. Res.* Special Issue on Complex Cognition 12, 365–376. doi: 10.1016/j.cogsys.2010.12.003

Güss, C. D., Glencross, E., Tuason, M. T., Summerlin, L., and Richard, F. D. (2005). "Task complexity and difficulty in two computer-simulated problems: cross-cultural similarities and differences," in *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*. eds. K. Forbus, D. Gentner and T. Regier (Mahwah, NJ: Cognitive Science Society and Lawrence Erlbaum Associates), 511–516.

Güss, C. D., Tuason, M. T., and Gerhard, C. (2010). Cross-national comparisons of complex problem-solving strategies in two microworlds. *Cogn. Sci. Soc.* 34, 489–520. doi: 10.1111/j.1551-6709.2009.01087.x

Hirschfeld, L. A., and Gelman, S. A. (Eds.) (1994). *Mapping the Mind. Domain Specificity in Cognition and Culture*. New York, NY: Cambridge University Press.

Hofstede, G. (2001). *Culture's Consequences* (*2nd ed*). Thousand Oaks, CA: Sage.

Hofstede, G., Hofstede, G. J., and Minkov, M. (Eds.) (2005). Cultures and organizations: Software of the mind (Vol. 2). New York: Mcgraw-hill.

Hölter, E. (2013). "Time horizon in German management: goal-orientated helix," in *Time and Management from a Cross-Cultural Perspective*. eds. H. Helfrich, E. Hölter and I. V. Arzhenovskiy (Cambridge, MA: Hogrefe Publishing), 161–173.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Kim, H. S. (2002). We talk, therefore we think? A cultural analysis of the effect of talking on thinking. *J. Pers. Soc. Psychol.* 83, 828–842. doi: 10.1037/0022-3514.83.4.828

Lave, J., and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge, England: Cambridge University Press.

Leng, C. Y., and Botelho, D. (2010). How does national culture impact on consumers decision-making styles? A cross cultural study in Brazil, the United States and Japan. *Braz. Administration Rev.* 7, 260–275. doi: 10.1590/S1807-76922010000300004

Lisø, S. O. (2011). *The Influence of Culture on Decision-Making in Strategic Alliance*. University of Agder, University of Agder, Kristiansand, Norway: Department of Economics and Business Administration.

Lovett, A., and Forbus, K. (2011). Cultural commonalities and differences in spatial problem-solving: a computational analysis. *Cognition* 121, 281–287. doi: 10.1016/j.cognition.2011.06.012

Matsumoto, (2008). Mapping expressive differences around the world. The relationship between emotional display rules and individualism versus collectivism. *J. Cross-Cult. Psychol.* 39, 55–74. doi: 10.1177/0022022107311854

McKinnon, A. J., Wearing, A. J. (1985). Systems analysis and dynamic decision making. *Acta Psychologica*, 58, 159–172. doi: 10.1016/0001-6918(85)90005-8

Müller, R., Spang, K., and Özcan, S. (2008). "Cultural differences in decision-making among project teams: examples from Swedish and German project teams," in *Paper Presented at PMI® Research Conference: Defining the Future of Project Management, Warsaw, Poland* (Newtown Square, PA: Project Management Institute).

Ng, A. H., and Hynie, M. (2014). Cultural differences in indecisiveness: the role of naïve dialecticism. *Personal. Individ. Differ.* 70, 45–50. doi: 10.1016/j.paid.2014.06.022

Papafragou, A., Li, P., Choi, Y., and Han, C. (2007). Evidentiality in language and cognition. *Cognition* 103, 253–299. doi: 10.1016/j.cognition.2006.04.001

Popper, K. (1999). *All life is Problem Solving*. New York, NY: Routledge.

Radicchi, F., Castellano, C., Cecconi, F., and Parisi, D. (2004). Defining and identifying communities in networks. *Proc Natl Acad Sci U S A*. 101, 2658–2663. doi: 10.1073/pnas.0400054101

Reichardt, J., and Bornholdt, S. (2006). Statistical mechanics of community detection. *Phys. Rev. E* 74:016110. doi: 10.1103/PhysRevE.74.016110

Reichert, U., and Dörner, D. (1988). Heurismen beim Umgang mit einem, einfachen dynamischen System [Heurisms in handling a "simple" dynamic system]. *Sprache Kognition* 7, 12–24.

Schwartz, S. (2010). Basic values: How the motivate and inhibit prosocial behavior. *In Prosocial* motives, emotions, and behavior: The better angels of our nature, eds. M. Mikulincer and P. R. Shaver, (pp. 221–241). American Psychological Association. doi: 10.1037/12061-012

Schoppek, W., Kluge, A., Osman, M., and Funke, J. (2018). Editorial: complex problem solving beyond the psychometric approach. *Front. Psychol.* 9:1224. doi: 10.3389/fpsyg.2018.01224

Sperber, D., and Hirschfeld, L. A. (1999). The cognitive foundations of cultural stability and diversity. *Trends Cogn. Sci.* 8, 40–46. doi: 10.1016/j.tics.2003.11.002

Strohschneider, S., and Güss, D. (1998). Planning and problem solving: differences between Brazilian and German students. *J. Cross-Cult. Psychol.* 29, 695–716. doi: 10.1177/0022022198296002

Tipandjan, A., Schäfer, T., Sundaram, S., and Sedlmeier, P. (2012). What are the important decisions in the lives of German and Indian university students? The structure of real-life decision-making processes. *Integr. Psychol. Behav. Sci.* 46, 205–234. doi: 10.1007/s12124-011-9189-0

Véras, E. Z., and Véras, D. B. (2011). Cultural differences between countries: the Brazilian and the Chinese ways of doing business. *J. Innovation Sustainability* 2, 77–83. doi: 10.24212/2179-3565.2011v2i2p77-83

Wüstenberg, S., Greiff, S., Molnár, G., and Funke, J. (2014). Cross-national gender differences in complex problem solving and their determinants. *Learn. Individ. Differ.* 29, 18–29. doi: 10.1016/j.lindif.2013.10.006

frontiers | Frontiers in Psychology

TYPE Hypothesis and Theory
PUBLISHED 25 January 2023
DOI 10.3389/fpsyg.2022.993660

OPEN ACCESS

EDITED BY
Miguel Ángel Carrasco,
National University of Distance Education (UNED), Spain

REVIEWED BY
Barbara Hanfstingl,
University of Klagenfurt, Austria
Silvia Lopes,
Universidade de Lisboa, Portugal

*CORRESPONDENCE
Ali Ünlü
✉ aligalibuenlue@gmail.com

SPECIALTY SECTION
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

RECEIVED 22 July 2022
ACCEPTED 12 December 2022
PUBLISHED 25 January 2023

CITATION
Ünlü A (2023) Qualitative motivation with sets and relations. *Front. Psychol.* 13:993660. doi: 10.3389/fpsyg.2022.993660

Qualitative motivation with sets and relations

Ali Ünlü*

School of Social Sciences and Technology, Technical University of Munich, Munich, Germany

In self-determination theory (SDT), multiple conceptual regulations of motivation are posited. These forms of motivation are especially qualitatively viewed by SDT researchers, and there are situations in which combinations of these regulations occur. In this article, instead of the commonly used numerical approach, this is modeled more versatilely by sets and relations. We discuss discrete mathematical models from the theory of knowledge spaces for the combinatorial conceptualization of motivation. Thereby, we constructively add insight into a dispute of opinions on the unidimensionality vs. multidimensionality of motivation in SDT literature. The motivation order derived in our example, albeit doubly branched, was approximately a chain, and we could quantify the combinatorial details of that approximation. Essentially, two combinatorial dimensions reducible to one were observed, which could be studied in other more popular scales as well. This approach allows us to define the distinct, including even equally informative, gradations of any regulation type. Thus, we may identify specific forms of motivation that may otherwise be difficult to measure or not be separable empirically. This could help to dissolve possible inconsistencies that may arise in applications of the theory in distinguishing the different regulation types. How to obtain the motivation structures in practice is demonstrated by relational data mining. The technique applied is an inductive item tree analysis, an established method of Boolean analysis of questionnaires. For a data set on learning motivation, the motivation spaces and co-occurrence relations for the gradations of the basic regulation types are extracted, thus, enumerating their potential subforms. In that empirical application, the underlying models were computed within each of the intrinsic, identified, introjected, and external regulations, in autonomous and controlled motivations, and the entire motivation domain. In future studies, the approach of this article could be employed to develop adaptive assessment and training procedures in SDT contexts and for dynamical extensions of the theory, if motivational behavior can go in time.

KEYWORDS

self-determination theory, motivation, knowledge space theory, set, relation, inductive item tree analysis, motivational implication, dimensionality

1. Introduction

Knowledge space theory (KST) was introduced by Doignon and Falmagne (1985, 1999), refer also Falmagne and Doignon (2011), which is a relatively recent psychometric theory. Initially, that theory was developed for the assessment and training of knowledge, but has evolved into a broader range of applications (e.g., Albert and Lukas, 1999; Falmagne et al., 2013). KST, as compared with the statistical item response theory (e.g., Van der Linden and Hambleton, 1997), for example, is inherently combinatorial; it did not develop from classical numerical scales in the first place (for a conceptual comparison of these theories, see Ünlü, 2007). The behavioral orientation of KST is good for qualitative modeling. Since numerical values (e.g., person ability or item difficulty) are more strongly aggregated numbers, in particular, restricted by their own natural ordering, the use of more fine-grained combinatorial structures (e.g., persons represented by their sets of skills they possess) allows for greater flexibility in more diagnostic modeling, as we will describe in this article. In a numerical approach, two persons are typically reduced to their aggregate motivation degrees, say $1.27 < 2.35$, which are always comparable (real numbers). This is in contrast to representing persons by the sets of motivations they possess, which are not necessarily comparable, for example, if represented by $\{a, c, e\}$ and $\{b, c, d\}$, where neither is a subset of the other. To avoid misconception at this point, obviously, there may be situations where the former perspective is desirable or sufficient for a use case, but if the aim is to have a qualitative assessment or conceptualization, the latter approach may be better suited. It depends on what the research aims are. For example, if ordering motivational behavior along the relative autonomy continuum is preferred, if continuously unidimensional, a quantitative single-valued scoring rule can suffice. However, if the goal is to see which motivations more strongly interrelate to what substantive outcomes, a qualitative model can be more useful. This article assumes that the latter perspective is desired for the problem of interest.

We describe an application of KST to self-determination theory (SDT). SDT was proposed by Deci and Ryan (1985, 2000), refer also Ryan and Deci (2000, 2002). The need for qualitative treatment of motivation was raised by researchers in SDT, in particular by Chemolli and Gagné (2014), with further pertinent SDT references therein, describing empirical problems that required a qualitative conceptualization of motivation (e.g., Koestner et al., 1996; Vansteenkiste et al., 2009). KST can contribute to this endeavor, as this article aims at demonstrating sets and relations among motivations (for advantages and a limitation, refer to section "Usefulness of this approach for motivation research and limitation").

Why do we treat the dichotomous data case first with this article? We are aware that SDT instruments use Likert scales and that dichotomization of polytomous or continuous data or variables can be controversial, but we have the following reasons. The theory of knowledge spaces is far more advanced in the dichotomous formulation, with a plethora of easier-to-grasp and better-accessible results that can readily be applied in motivation research. For example, if you take Birkhoff's theorem, an important mathematical as well as methodological result, this theorem is way easier to formulate and understand than its polytomous counterparts. In KST, the polytomous case is an ongoing research, still, many powerful dichotomous concepts of KST have to be generalized and developed for polytomous items, if possible. In addition, dichotomous indicators can provide useful information about binary classification problems, for example, whether a person is intrinsically motivated or not, or more generally, which motivations may or may not be present in the total motivational profile of a person. That is, dichotomous indicators can still be informative enough for such use cases involving binary decisions. Dichotomous results may also be viewed as necessary conditions for a polytomous model, for example, when violated, they may give evidence against the model. In general, results obtained by dichotomous analyses are, albeit rougher, easier to interpret. In particular, what we will see in this article is that the approach based on sets and relations in dichotomous formulation allows us to quantify, combinatorially and qualitatively, how far multidimensionality may be away from unidimensionality, thus contributing to the debated issue of dimensionality in SDT (Chemolli and Gagné, 2014; Sheldon et al., 2017). Anyhow, we have to see the merits of this approach after SDT researchers have given this method a try by testing it across their motivation scales or empirical studies.

How to derive KST relations among motivations from empirical data? We discuss one possibility based on the data-analytic approach of the inductive item tree analysis (IITA). There are other methods as well (e.g., Albert and Lukas, 1999), for example, by querying experts; theory-driven, based on skills or competencies; or by data mining (which is nearest to what we present). Publications to learn more about IITA are Schrepp (1999a,b, 2003, 2006), Sargin and Ünlü (2009), Ünlü and Sargin (2010), and Ünlü and Schrepp (2021). As a well-established method of Boolean analysis of questionnaires, IITA takes as input a data set (e.g., of motivation scores), and in this article, it is treated for the dichotomous case and derives the set of implications deemed to be plausible for the data set according to some faithful criteria. That is, detailed later, the IITA algorithm can be used to extract motivation co-occurrence relations, and thus, by application of Birkhoff's theorem, quasi-ordinal motivation spaces, from data motivation variables.

In addition to data analysis, rather theoretically, we believe that the application of KST to SDT can in particular be useful for the representation of the logical structure of motivations. Based on mathematical considerations, we may obtain principled definitions of such pertinent concepts as self-determination and derive from these more abstract definitions, results about their universal (mathematical or axiomatic) properties in population instead of sample quantities (Ünlü, 2022).

This article has the following structure. In section "Self-determination theory," the theory of self-determination is reviewed, and in section "Knowledge space theory, Birkhoff's theorem, and inductive item tree analysis," the knowledge space theory. In particular, section "Knowledge space theory, Birkhoff's theorem, and inductive item tree analysis" also includes a short introduction to IITA and Birkhoff's theorem. In section "Sets and relations among motivations," the application of KST sets and relations to motivation is described. The SDT analogs of the basic concepts of KST are the motivation domain, motivation structure, motivation state, motivation co-occurrence relation, and quasi-ordinal motivation space. In section "An empirical application," an empirical example is provided, which concerns learning motivation. In section "Usefulness of this approach for motivation research and limitation," we outline why this approach to the modeling and analysis of motivation is useful and a limitation of this study. This article ends with a conclusion in section "Conclusion," and with **Appendices A, B** containing the scale and binary data sets used for the empirical application, respectively.

2. Self-determination theory

Self-determination theory maintains a comprehensive website at https://selfdeterminationtheory.org. As a theory of motivation, SDT investigates what drives people to act (Ryan(ed.), 2019; Conesa et al., 2022; Gagné et al., 2022). The basic concepts of SDT are best represented by **Table 1**, a table very often reported in SDT publications, and here, we present a slightly modified interpretation of it.

Briefly, intrinsic motivation represents behavior enjoyed doing it for its own sake, extrinsic motivation is instrumental behavior, and amotivation is the lack of motivation. SDT does not further differentiate intrinsic motivation and amotivation, thus their single regulations are interpreted accordingly. But extrinsic motivation, which is assumed, can have varying internalization. Extrinsic motivation is differentiated into three (or four) gradually internalized regulation types, in increasing order of internalities, namely, external regulation, introjected regulation, identified regulation, and integrated regulation. Again briefly, external regulation is behavior dictated by purely external factors such as reward or punishment. Introjected regulation is more internalized than external regulation and includes proving oneself worthy or avoiding guilt. Identified regulation is even more internalized than introjected regulation and is described as acting to express rational values without being accompanied by unforced interval motives such as fun or inherent satisfaction. Integrated regulation, although extrinsic motivation, is assumed to be only internal motivation (different from intrinsic regulation), located on the internalization continuum of SDT between identified regulation and intrinsic regulation. With integrated regulation, a person's identified values are even further internalized and integrated with each other. Integrated regulation seems to be difficult to measure (e.g., Roth et al., 2009; Gagné et al., 2014). In literature, additional forms of motivation have also been proposed, for example, negative introjection (left) and positive introjection (right) in Sheldon et al. (2017) or avoidance introjection (left) and approach introjection (right) in Assor et al. (2009). In each of these cases, both forms of introjection are between external regulation and identified regulation. Probably other forms of motivation may be possible here and there. Why do we list them? In the approach based on sets and relations, which provides a general framework, all of these motivations can be easily included in the formulation of the models. These motivations define what will be called the motivation domain, the set of all motivations of interest. The set and relation

TABLE 1 Self-determination continuum also called SDT's taxonomy of motivation (cf., Deci and Ryan, 2000).

Behavior continuous	Nonself-determined	Self-determined			
Type of motivation	Amotivation	Extrinsic motivation			Intrinsic motivation
Type of regulation	Non-regulation	External regulation	Introjected regulation	Identified regulation	Intrinsic regulation
	Lack of external and internal controls and motives Complement of motivation	Constraints of external controls Lack of internal motives	Constraints of internal controls Forced external motives	Internally identify with external value Quasi-unforced internal motives	Lack of external and internal controls Unforced internal motives
Locus of causality continuous	Impersonal	External	Somewhat external	Somewhat internal	Internal
Trichotomy of form	Uncontrolled Non-autonomous	Controlled Non-autonomous	Controlled Non-autonomous	Quasi-uncontrolled Quasi-autonomous	Uncontrolled Autonomous

We see three main types of motivation (intrinsic motivation, extrinsic motivation, amotivation) with their corresponding regulatory styles (intrinsic regulation, identified regulation, introjected regulation, external regulation, non-regulation) and loci of causality (internal, somewhat internal, somewhat external, external, impersonal). These regulations describe increasingly less self-determined (with far left nonself-determined) behaviors, of different qualitative forms (uncontrolled/autonomous, quasi-uncontrolled/quasi-autonomous, controlled/non-autonomous, controlled/non-autonomous, uncontrolled/non-autonomous).

representations defined are built upon the specified domain of motivations.

In this article, we show how KST can be applied to qualitatively model these motivations or regulations of SDT's (extended) taxonomy. Instead of, typically by confirmatory factor analysis, introducing one or more latent continuous dimensions and factors, to represent each of the regulation types, we use sets and relations among those motivations or regulations. A strength of the latter approach is that it allows for very general combinatorial structures and, thus, offers more flexibility in modeling motivation qualitatively (section "Usefulness of this approach for motivation research and limitation").

3. Knowledge space theory, Birkhoff's theorem, and inductive item tree analysis

We give a short introduction to KST (Doignon and Falmagne, 1985), including the theorem by Birkhoff (1937), and IITA (Schrepp, 1999b; Sargin and Ünlü, 2009). These three components are the building blocks of the methodology used in this article. The core contributions of our study are to apply KST models to represent motivation in SDT, with equivalent mathematical representations at the levels of persons and motivations by Birkhoff's theorem as a byproduct, and the concrete implementation of those models in data through the use of IITA.

3.1. Knowledge space theory and Birkhoff's theorem

An application with examples of the concepts in motivation presented here can be found in section "Sets and relations among motivations."

3.1.1. Surmise relation

The starting point for a theory of knowledge assessment is to assume that in a knowledge domain of interest, the pieces of knowledge may imply each other. For example, in the knowledge domain of natural numbers, the mastery of the arithmetic problem b "$3 \cdot 2 \ =$ " may imply the mastery of the arithmetic problem a "$2+2 \ =$ ". That is, the mastery of problem a is assumed to be a prerequisite for the mastery of problem b. Mathematically, this is represented by the ordered pair $a \sqsubseteq b$ of a binary relation $\sqsubseteq$ on the knowledge domain. In accordance with the interpretation of mastery, that relation is assumed to be reflexive and transitive, a quasi-order, or as it is also called in KST, a surmise relation.

Definition 1. Let Q be a non-empty and finite set of dichotomous items, the (knowledge) domain. Let $\sqsubseteq$ be a binary relation on Q, that is, a subset of $Q \times Q$. We call $\sqsubseteq$ a quasi-order or surmise relation (on Q) if and only if it is reflexive and transitive, that is, if and only if, respectively, $x \sqsubseteq x$ for all $x \in Q$, and $x \sqsubseteq y$, $y \sqsubseteq z$ implies $x \sqsubseteq z$ for all $x, y, z \in Q$.

Typically, Q can be a psychological or educational test consisting of dichotomous questions or problems that can either be solved (coded 1) or failed (coded 0) by examinees, and a surmise relation on that test then captures the mastery dependencies among the test items.

3.1.2. Knowledge structure and space

The implications of the surmise relation entail that only certain mastery patterns, represented by subsets of the domain, are admissible, which are called the knowledge states. For example, the subset of items is mastered by a student's, her or his, knowledge state. If it contains the multiplication item b, then it must also contain the addition item a, since $a \sqsubseteq b$. The collection of all so compatible knowledge states is called the knowledge structure. In ideal conditions, if no response errors occur, the only response patterns possible would be the knowledge states.

Definition 2. Let Q be a domain. A knowledge structure $\mathcal{H}$ on Q is a set of subsets of Q, which contains at least the empty set $\emptyset$ and Q itself. The elements of $\mathcal{H}$ are called knowledge states.

Knowledge states are subsets of Q. Thus, we can take their union $\cup$ and intersection $\cap$, which yield the important special case of a knowledge structure, a quasi-ordinal knowledge space.

Definition 3. Let $\mathcal{H}$ be a knowledge structure on the domain Q. We call $\mathcal{H}$ a knowledge space (on Q) if and only if $G \cup H \in \mathcal{H}$ for all $G, H \in \mathcal{H}$. The knowledge structure $\mathcal{H}$ is a closure space (on Q) if and only if $G \cap H \in \mathcal{H}$ for all $G, H \in \mathcal{H}$. A quasi-ordinal knowledge space is a knowledge structure that is both a knowledge space and a closure space.

3.1.3. Birkhoff's theorem

The quasi-ordinal knowledge space model is the set representation at the level of persons and the surmise relation model is the order representation at the level of items. They correspond to a person's ability and item difficulty of numerical item response theory. In contrast to the numerical approach, in KST, the two representations are connected by a central mathematical theorem, which is Birkhoff's theorem. The details of this theorem can be found in Falmagne and Doignon (2011, section 3.8, pp. 56–58).

Theorem 1 (Birkhoff, 1937). There exists a one-to-one correspondence between the collection of all quasi-ordinal knowledge spaces $\mathcal{H}$ on a domain Q and the collection of all surmise relations $\sqsubseteq$ on Q, defined by the two equivalences, for $p, q \in Q, H \subseteq Q$,

$$\mathrm{p} \sqsubseteq q :\Longleftrightarrow (\forall G \in \mathcal{H}, \mathrm{q} \in G \Longrightarrow p \in G),$$

$$\mathrm{H} \in \mathcal{H} :\Longleftrightarrow (\forall r \sqsubseteq \mathrm{s}, s \in H \Longrightarrow r \in H)\,.$$

This theorem mathematically links two different levels of empirical interpretations. It will be applied to SDT and exemplified in section "Sets and relations among motivations."

3.1.4. Validation

Validation procedures were proposed in the literature based on probabilistic extensions of knowledge structures. The most prominent one is the basic local independence model, a latent class scaling model (Doignon and Falmagne, 1999, chapter 7). For logistic and generalized normal ogive statistical validation procedures in KST, refer to Stefanutti (2006) and Ünlü (2006), respectively. Latent class analysis exploratory, estimation, and testing procedures were also studied for knowledge structures (Schrepp, 2005; Ünlü, 2011). For recent developments in performance- and competence-based knowledge space theory, including further, also more qualitative, validation procedures, refer to Falmagne et al. (2013).

3.2. Inductive item tree analysis

In IITA, competing relations are generated and a fit measure is computed for each of these relations in order to find that relation which most adequately describes the data. Since traditional inference-based methods, such as (asymptotic) chi-squared goodness-of-fit tests (available as well), do almost always reject the wished model (placed in the null hypothesis), this class of IITA relational mining techniques is generally effective and more useful. The R (The R Core Team, 2022) package DAKS (Ünlü and Sargin, 2010) freely available at https://CRAN.R-project.org/package=DAKS implements the IITA procedures, in addition to software by Schrepp (2006).

3.2.1. General problem

What is the general problem addressed by IITA? Assume that you have noisy indicators for latent variables of interest and that among those latent variables there exist latent logical, that is, deterministic, implications. The goal of IITA is to detect these implications from the information on the indicators. A typical KST example of an implication, refer section "Knowledge space theory and Birkhoff's theorem," is that the mastery of a math problem may imply the mastery of another math problem, where the questions of a math test are the indicators. In section "Sets and relations among motivations," we apply this idea to motivation in SDT. There, the indicators are the test items of a motivation questionnaire (https://selfdeterminationtheory.org/questionnaires), which measure their underlying, for example, intrinsic and external regulations. Then, a logical implication between motivations assumes that the latent possession or occurrence of a motivation implies the possession or occurrence of another motivation. If in a study such motivational implications are deemed to be realistic or of interest, IITA is a technique that can be used to uncover those implications from the motivation questionnaire data.

3.2.2. Computational components

The IITA algorithm consists of three computational components. First, it constructs a selection set of competing quasi-orders on the domain of latent (e.g., motivation) variables. That construction is inductive. For varying numbers of observed counterexamples of an (e.g., motivational) implication (premises true but conclusions false), anchored with the simplest quasi-order consisting of all implications that are not violated in the data, in successive steps, quasi-orders (e.g., motivation co-occurrence relations) are constructed. This is realized by adding specific implications that have no more than the predefined numbers of counterexamples in the data set and that do not violate transitivity. In this way, a maximum of sample size plus one, typically a much smaller number than this, of increasingly more complex quasi-orders are derived. Second, the fit of each constructed quasi-order to the data set is quantified by a measure of the average squared differences between the observed and under the model expected numbers of counterexamples for the implications. Third, a best-fitting (e.g., motivation co-occurrence) relation of the selection set with the computed minimum discrepancy is chosen as the final solution. Subsequently, we outline these components of the technique, and more details can be found in Ünlü and Schrepp (2021, section 2).

3.2.3. Inductive construction

We start with notation. Let $Q = \{i_1, ..., i_m\}$ be the domain of $m \geq 2$ dichotomous items. Let $D = \{d_1, ..., d_N\}$ be the data set (with repetitions) of observed response patterns (mappings) $d_n : Q \to \{0, 1\}$, where $d_n(i) = 0$ or 1 stands for the response of the subject $n = 1, ..., N$ to the item $i \in Q$. For $i, j \in Q$, let $b_{ij} = \left|\{d \in D : d(i) = 1 \wedge d(j) = 0\}\right|$ be the number of subjects of the sample who solved item i but failed item j. If we postulate the implication $i \longrightarrow j$ (**Definition 5**), b_{ij} is the number of observed counterexamples for that implication. We can define the binary relation $\sqsubseteq_0$ of all implications that are not violated in data D by $j \sqsubseteq_0 i :\Longleftrightarrow b_{ij} = 0$ for all $i, j \in Q$. This relation $\sqsubseteq_0$ is a quasi-order on Q (Van Leeuwe, 1974). However, this is not the final quasi-order that IITA returns. To accept $\sqsubseteq_0$ is generally not satisfactory since this is data fitting, which does not account for response errors. Thus, IITA allows for varying numbers of observed counterexamples $L = 0, 1, ..., N$.

The construction of the quasi-orders is as follows. The IITA algorithm starts with $\sqsubseteq_0$, but inductively constructs bigger quasi-orders. The procedure to construct the $L+1$ step quasi-order $\sqsubseteq_{L+1}$ from the L step quasi-order $\sqsubseteq_L$, anchoring with $\sqsubseteq_0$, for $L = 0, 1, ..., N-1$, consists of the following steps 1, 2, and 3:

1. To determine the set A_{L+1} of all item pairs that are not already contained in $\sqsubseteq_L$ and have no more than $L+1$ observed counterexamples in D.
2. To iteratively repeat the following two operations a and b:

a. For each element of A_{L+1}, check if it violates transitivity in $\sqsubseteq_L \cup A_{L+1}$. If so, mark that element.
b. If no element is marked in operation a, stop step 2. Otherwise, delete all marked elements from A_{L+1} and restart the process in operation a with this reduced new set A_{L+1}.

3. When the process in step 2 stops, the remaining implications in A_{L+1} do not violate transitivity in $\sqsubseteq_L \cup A_{L+1}$. By construction, $\sqsubseteq_{L+1} := \sqsubseteq_L \cup A_{L+1}$ is the $L+1$ step quasi-order.

Thus, by IITA, the increasingly bigger quasi-orders $\sqsubseteq_0 \subseteq \sqsubseteq_1 \subseteq \ldots \subseteq \sqsubseteq_N$ are inductively constructed.

3.2.4. Fit measure

Among these relations, IITA proposes the following method to determine the best-fitting quasi-order. We quantify the fit of any of the IITA quasi-orders $\sqsubseteq_L$, $L = 0, 1, \ldots, N$ to the data set D by the measure:

$$\text{diff}\left(\sqsubseteq_L, D\right) = \frac{\sum_{i \neq j}\left(b_{ij} - t_{ij}\right)^2}{m\left(m-1\right)},$$

where the sum is taken over all item pairs $(j, i) \in Q \times Q$, $i \neq j$, and m is the number of items. In addition, b_{ij} is the observed number of subjects who solved item i but failed item j, and t_{ij} is the corresponding theoretical value expected and derived under the assumption that $\sqsubseteq_L$ is the correct quasi-order underlying the data set D.

The derivation of the t_{ij} estimators is intricate and necessitate a few considerations.

1. Assume that $\sqsubseteq_L$, for a given L, is the quasi-order of true logical implications between the items. How many violations for a true implication $i \longrightarrow_L j$, $i, j \in Q$ can we expect? If we assume a single response error probability by which a true implication may be violated, that rate can be estimated by:

$$\gamma_L = \frac{\sum_{\mathcal{G}} \frac{b_{ij}}{p_i N}}{|\mathcal{G}|} \text{ if } \mathcal{G} \neq \emptyset, \text{ or } 0 \text{ if } \mathcal{G} = \emptyset,$$

where $\mathcal{G} = \{j \sqsubseteq_L i : i \neq j \wedge p_i \neq 0\}$, and $p_i = \frac{|\{d \in D : d(i) = 1\}|}{N}$, $i \in Q$ is the relative frequency of subjects of the sample who solved item i. Thus, γ_L is an estimated average amount of random response errors in the data, under the assumption that $\sqsubseteq_L$ is the underlying true quasi-order. For further motivation for this choice of estimator, refer to Ünlü and Schrepp (2021).

2. Under the assumption that $\sqsubseteq_L$ is the correct quasi-order, thus based on the corresponding estimated error probability γ_L, for any item pair $(j, i) \in Q \times Q$, $i \neq j$, the, under $\sqsubseteq_L$ derived, theoretical values t_{ij} used in the definition of the diff measure can be estimated as follows: three cases are distinguished (for more details, refer to Ünlü and Schrepp, 2021). First, if $j \sqsubseteq_L i$, $i \neq j$, use the estimation equation $t_{ij} = \gamma_L p_i N$. Second, if $j \not\sqsubseteq_L i$ and $i \not\sqsubseteq_L j$, assume that the items are stochastically independent, and set $t_{ij} = \left(1 - p_j\right) p_i N$. Third, if $j \not\sqsubseteq_L i$ but $i \sqsubseteq_L j$, the estimator is $t_{ij} = max\left(0, \left(p_i - p_j + p_j \gamma_L\right) N\right)$.

3.2.5. Selection

A validation procedure is obtained that gives information about which model to pick. With the above ingredients, in data D, the fit measure diff is computed for each quasi-order obtained by inductive construction. Thus, a non-negative real value is associated with any of the competing quasi-orders. Since diff quantifies an average squared difference between observed and expected variables, smaller values of the measure are interpreted to indicate a better fit. In particular, the decision rule is to select that quasi-order among $\sqsubseteq_0, \sqsubseteq_1, \ldots, \sqsubseteq_N$, which has the minimum diff value. This is the final solution returned by the IITA algorithm.

4. Sets and relations among motivations

We apply the basic concepts of KST to SDT and describe them in motivation.

Before presenting the definitions, a general remark is in order. The program of this article has far-reaching consequences (section "Usefulness of this approach for motivation research and limitation"). The KST models applied to motivation can be used to develop routines in motivation research for the adaptive assessment and training of motivation using computers. Adaptive testing is the major strength and point of origin of KST (Falmagne et al., 2013, with references therein). Noteworthy, there is an essential difference between the educational vs. motivational applications of the models. The number of conceivable states of motivation in SDT may not be that large as compared with the states of knowledge studied in KST, with several million feasible knowledge states in large-scale empirical studies. This is clearly an advantage of the SDT application, in particular, for combinatorial as well as statistical reasons.

We want to motivate the basic concepts by an example, and then give the definitions for motivation. Consider the regulations, for ease of presentation with no gradations of the regulations (generalized below), external regulation a, introjected regulation b, identified regulation c, and intrinsic regulation d. These motivations make up the motivation domain of interest, that is, the set of all considered motivations $M = \{a, b, c, d\}$ (other choices for the domain are discussed later). The central assumption is that these motivations can only occur in certain combinations in the population of reference, called the motivation states, which are subsets of the motivation domain. For example, a student could be externally and introjectedly motivated at the same time, represented by the state or subset $\{a, b\}$ of the motivation domain. Or, this

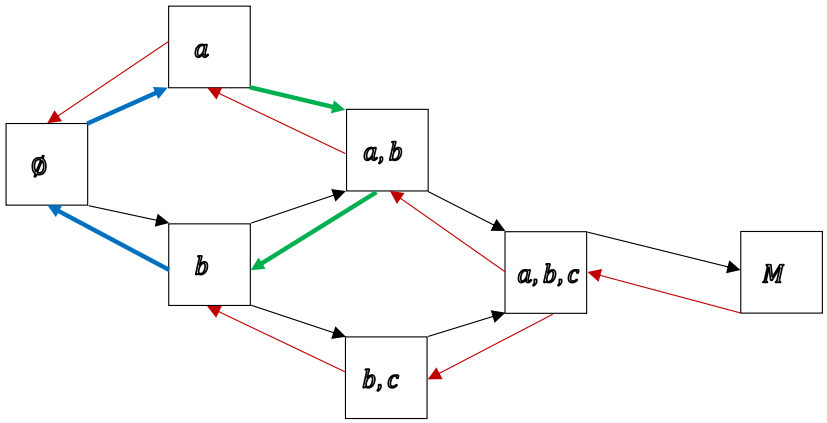


FIGURE 1
The directed graphs, in black and red, respectively, of the motivation structure $\mathcal{M} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{b, c\}, \{a, b, c\}, M\}$, with $M = \{a, b, c, d\}$. A black arc (directed edge) $A \longrightarrow B$ linking the motivation state A (left) to the motivation state B (right) means $A \subset B$ (A subset of B) and there is no other motivation state between the two. A red arc $B \longrightarrow A$ linking the motivation state B (right) to the motivation state A (left) stands for $B \supset A$ (B superset of A) and there is no other state in between. A trajectory describing the transition from the state of pure external motivation $\{a\}$ to the state of pure introjected motivation $\{b\}$, along directed edges (of both colors) in only states of the motivation structure, is shown in green, $\{a\} \longrightarrow \{a, b\}$ (black arc) and $\{a, b\} \rightarrow \{b\}$ (red arc). Another trajectory, now from state $\{b\}$ to state $\{a\}$, posited under this model, probably critical empirically, is to become amotivated first, purging the existent introjected motivation, before the other external motivation can be gained, shown in blue. In particular, there is no direct arc linking the two motivation states, thus, according to this model, external motivation cannot be directly converted into introjected motivation and vice versa, but instead, for example, under the green trajectory, must be attained jointly first before the initial one is forfeited to end up with the other. In these graphs, motivations are gained or lost one by one in progressions from state to state. Restrictions of this sort imply serviceable mathematical properties.

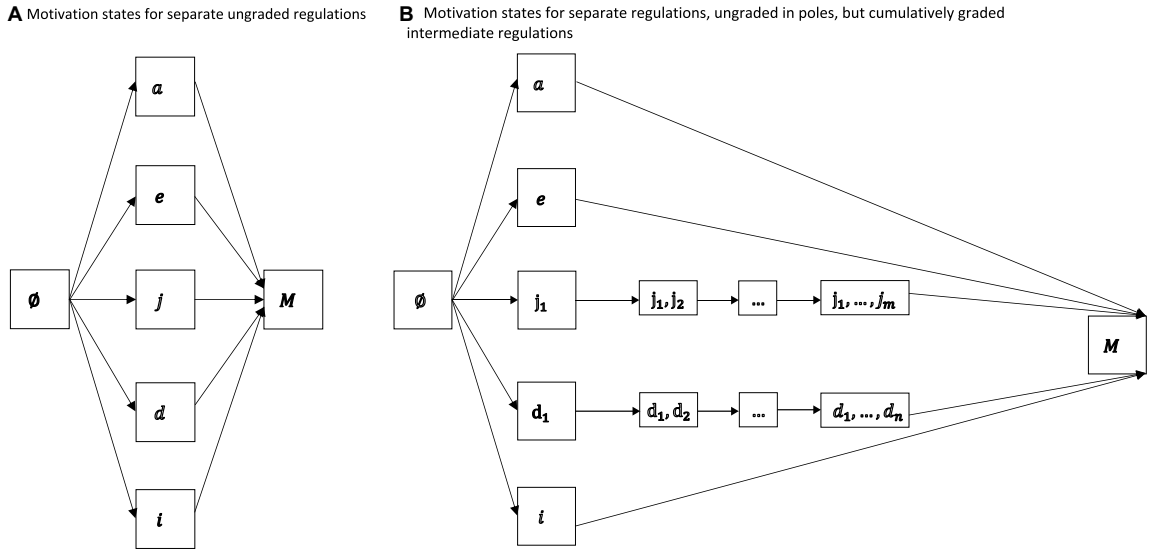


FIGURE 2
Motivation states for separate ungraded regulations **(A)** and separate but cumulatively graded intermediate regulations **(B)**. In both representations, non-regulation, external regulation, and intrinsic regulation are singletons. The linear orders imposed on the gradations of introjected regulation and identified regulation could be associated with the varying cumulative internalities that the gradations may have along an internalization continuum.

is an assumption that can change depending on the model used; for a student to be intrinsically motivated, the student necessarily needs to have the other three motivations. In this case, $M = \{a, b, c, d\}$ is the only possible motivation state containing intrinsic regulation, and other combinations containing intrinsic regulation such as $\{a, d\}$ (externally and intrinsically motivated) or $\{d\}$ (only intrinsically motivated) are excluded according to the posited model. Since amotivation is understood to be the state of no motivation, this could be modeled by a specific subset, the empty set $\emptyset$. We can

collect together all feasible combinations of motivations, the motivation states, into a set on its own, called the motivation structure. In this example, the combinations $\{a, b\}$ and M belong to the motivation structure, but not $\{a, d\}$ and $\{d\}$. As mentioned, if this model is deemed to be empirically inadequate, it can be replaced by another model; the mathematical definition allows us to flexibly define the states, that is, the combinations of motivations considered to be feasible or occurring in an empirical study. The practical derivation of the motivation structure can be (inclusive "or") theory-driven, derived from querying experts, or obtained by statistics and data analysis, where the latter is the approach pursued with this article. Let us assume that a researcher has identified the following motivation structure in the motivation domain of her study,

$$\mathcal{M} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{b, c\}, \{a, b, c\}, M\}. \quad (1)$$

This motivation structure captures the logical organization of the motivations of interest in the population of reference. Only those combinations of the motivations are feasible, in the sense that they have shares of the population. Mathematically though, the motivation structure is defined to always include the empty set and the motivation domain, for technical reasons. Leaving out these extreme states from the definition of a motivation structure, in principle you could, will complicate, or probably invalidate, mathematical results, at least in their common formulations. We will stay with the definition used in KST and include as stated always the empty set and the motivation domain itself. So, amotivation, if modeled as the empty set, in contrast to the element(s) of the domain, and the possibility of possessing all regulations jointly are assumed to be motivation states under any deterministic model. Thus, if these extreme motivation states do not occur, any specified deterministic model will be wrong in those states. Otherwise, the deterministic model can always be chosen to be correct in all other states. However, this "methodological artifact" is not restrictive from an empirical viewpoint. In practice, if any of these two states is deemed to be empirically implausible, its probability of occurrence in the population of reference will (virtually) be zero, thus it will be discarded in the probabilistic formulation and extension of the deterministic model. However, as seen below, at the level of motivations viewed as "items" composing the states, this subtle issue is resolved and does not appear.

The motivation structure $\mathcal{M}$ in **Eq. 1** induces the two directed graphs shown in **Figure 1**, which suggests a process of how motivation may progress from state to state, for example, over time. **Figure 1** helps to indicate the general idea, and in particular, the flexibility that comes with such a discrete mathematical structure. Depending on the empirical situation, a proper model could be specified (e.g., quantitatively by exploratory data analysis, below), allowing for the qualitative, or even dynamic over time, analysis of motivation.

With motivation progression, things can get mathematically more tractable, if only "learning" is possible. That is, if only additional, new motivations are attained, one by one, and no motivations are lost during progression, moving from left to right in the graph of **Figure 1**, along the black arcs only. In such a more restrictive model, in this example, a student cannot reach the states of pure external motivation or pure introjected motivation if she or he is initially only introjectedly or externally motivated, respectively. This would be prohibited by model assumption. This case of only "learning," that is, merely moving from left to right along black arcs, cumulative in the "items" or motivations "learned," one at a time, is the case that has been extensively studied in dichotomous KST. It is especially interesting, if justifiable in a study on motivation, since it entails a rich mathematical theory, implying strong mathematical measurement properties (Falmagne and Doignon, 2011).

Here is the definition in terms of motivation.

Definition 4. Let M be a non-empty set of motivations or regulations (examples below). Let $\mathcal{M}$ be a collection of subsets of M, which contains at least $\emptyset$ and M. Then, $\mathcal{M}$ is called a motivation structure on (or in) the motivation domain M. The elements of $\mathcal{M}$ are the motivation states.

Example 1. We consider the general case. The motivation domain M could consist of k forms (gradations) of non-regulation $a_1, ..., a_k$; l forms of external regulation $e_1, ..., e_l$; m forms of introjected regulation $j_1, ..., j_m$; n forms of identified regulation $d_1, ..., d_n$; u forms of integrated regulation $g_1, ..., g_u$; and o forms of intrinsic regulation $i_1, ..., i_o$. In this case, the extreme motivation state M describes hypothetical behavior that is regulated by all types jointly. The other extreme state $\emptyset$ may represent a sort of totally unregulated behavior, unexplainable by any of the types (regarding the extreme states, refer to the aforementioned text). The motivation structure in this domain can contain arbitrary combinations of these regulations, such as the state consisting of the last form of external regulation and the first two (if $o \geq 2$) of intrinsic regulation $\{e_l, i_1, i_2\}$, which represents a student extrinsically and intrinsically motivated in respective gradations of the regulations.

Example 2. Let the notation be as in **Example 1**. Assume that the motivations are not further graded, $k = l = m = n = o = 1$, where integrated regulation is not of interest (in the sequel). If non-regulation a, external regulation e, introjected regulation j, identified regulation d, and intrinsic regulation i can only occur separately,

$$\mathcal{M}_1 = \{\emptyset, \{a\}, \{e\}, \{j\}, \{d\}, \{i\}, M\}.$$

If, in this representation, the intermediate introjected and identified regulations have cumulative gradations $j_1 \preccurlyeq_j ... \preccurlyeq_j j_m$ and $d_1 \preccurlyeq_d ... \preccurlyeq_d d_n$, indexed in the increasing rankings of their

TABLE 2 Frequency distributions of the dichotomous scores across motivation variables of, from top to bottom, the individual regulation types, of autonomous motivation and controlled motivation, of the entire motivation domain, and the subsets *A*, *B*, and *C* of the data set, respectively.

$N = 948$	i_1	i_2	i_3	i_4	i_5														
0	419	381	442	429	427														
1	529	567	506	519	521														
$N = 1028$	e_1	e_2	e_3	e_4	e_5														
0	795	789	382	465	609														
1	233	239	646	563	419														
$N = 1073$	d_1	d_2	d_3	d_4	d_5														
0	302	277	300	332	662														
1	771	796	773	741	411														
$N = 1168$	j_1	j_2	j_3	j_4															
0	574	675	806	779															
1	594	493	362	389															
$N = 482$	i_1	i_2	i_3	i_4	i_5	d_1	d_2	d_3	d_4	d_5									
0	195	171	207	196	195	136	138	150	168	285									
1	287	311	275	286	287	346	344	332	314	197									
$N = 550$	e_1	e_2	e_3	e_4	e_5	j_1	j_2	j_3	j_4										
0	425	418	225	276	342	289	326	384	378										
1	125	132	325	274	208	261	224	166	172										
$N = 180$	i_1	i_2	i_3	i_4	i_5	d_1	d_2	d_3	d_4	d_5	j_1	j_2	j_3	j_4	e_1	e_2	e_3	e_4	e_5
0	77	69	85	79	76	55	54	56	64	110	89	101	114	113	122	123	61	83	92
1	103	111	95	101	104	125	126	124	116	70	91	79	66	67	58	57	119	97	88
$N = 342$	i_1	i_3	i_4	i_5	d_5	j_1	j_2	j_3	j_4	e_3	e_4	e_5							
0	166	182	171	172	228	181	201	231	234	120	163	206							
1	176	160	171	170	114	161	141	111	108	222	179	136							
$N = 577$	i_1	i_2	i_3	i_4	i_5	d_1	d_2	d_3	d_4										
0	224	197	239	228	227	156	159	172	191										
1	353	380	338	349	350	421	418	405	386										
$N = 706$	d_5	j_1	j_2	j_3	j_4	e_1	e_2	e_5											
0	516	403	447	514	503	567	567	465											
1	190	303	259	192	203	139	139	241											

The sample sizes *N* used in their corresponding IITA analyses of these motivation variables are also shown.

10.3389/fpsyg.2022.993660

TABLE 3 The scale (Müller et al., 2007) used for the empirical application (Appendix A).

Ich arbeite und lerne in diesem Fach, ...	Stimmt völlig	Stimmt eher	Stimmt teils/teils	Stimmt eher nicht	Stimmt überhaupt nicht
1 ... weil es mir Spaß macht.	O	O	O	O	O
2 ... weil ich möchte, dass mein Lehrer denkt, ich bin ein/e gute/r Schüler/in.	O	O	O	O	O
3 ... um später eine bestimmte Ausbildung machen zu können (z.B. Schule, Lehre oder Studium).	O	O	O	O	O
4 ... weil ich sonst von zu Hause Druck bekomme.	O	O	O	O	O
5 ... weil ich neue Dinge lernen möchte.	O	O	O	O	O
6 ... weil ich ein schlechtes Gewissen hätte, wenn ich wenig tun würde.	O	O	O	O	O
7 ... weil ich damit mehr Möglichkeiten bei der späteren Berufswahl habe.	O	O	O	O	O
8 ... weil ich sonst Ärger mit meinem/r Lehrer/in bekomme.	O	O	O	O	O
9 ... weil ich es genieße, mich mit diesem Fach auseinander zu setzen.	O	O	O	O	O
10 ... weil ich möchte, dass die anderen Schüler/innen von mir denken, dass ich ziemlich gut bin.	O	O	O	O	O
11 ... weil ich mit dem Wissen im Fach später einen besseren Job bekommen kann.	O	O	O	O	O
12 ... weil ich sonst schlechte Noten bekomme.	O	O	O	O	O
13 ... weil ich gerne Aufgaben aus diesem Fach löse.	O	O	O	O	O
14 ... weil ich mich vor mir selbst schämen würde, wenn ich es nicht tun würde.	O	O	O	O	O
15 ... weil ich die Sachen, die ich hier lerne, später gut gebrauchen kann.	O	O	O	O	O
16 ... weil ich es einfach lernen muss.	O	O	O	O	O
17 ... weil ich gerne über Dinge dieses Faches nachdenke.	O	O	O	O	O

This is a copy of page 2 in https://ius.aau.at/wp-content/uploads/2016/01/mui_fragebogen.pdf (as of 30 October 2022). Since this scale was only validated in German, thus to avoid introducing any artifact, we have abstained from translating it into English.

linear orderings $\preccurlyeq_j$ and $\preccurlyeq_d$ (orders are treated later),

$$\mathcal{M}_2 = \{\emptyset, \{a\}, \{e\}, \{j_1\}, \{j_1, j_2\}, ..., \{j_1, j_2, ..., j_m\},$$
$$\{d_1\}, \{d_1, d_2\}, ..., \{d_1, d_2, ..., d_n\}, \{i\}, M\}.$$

We present the corresponding directed graphs of Example 2 in Figure 2.

Besides motivation structures, the set representation, there is the other representation based on relations. We can adumbrate how the two representations are connected by reconsidering the example with $\mathcal{M} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{b, c\}, \{a, b, c\}, M\}$ in Eq. 1. These states are the combinations of motivations that people can have. Under this model, if a person possesses motivation c, that is, if she or he is in one of the states $\{b, c\}$, $\{a, b, c\}$, or $\{a, b, c, d\}$, then this person must also possess motivation b, since b is in all motivation states that contain c. That is, (for any person) if motivation c occurs/is possessed, then motivation b occurs/is possessed. We also express this by saying that motivation c implies motivation b (always in the interpretation of occurrence or possession), denoted by $c \longrightarrow b$ or $b \sqsubseteq c$. Similarly, we see that if motivation d is possessed, then necessarily all other motivations must also be possessed,

TABLE 4A Intrinsic regulation, with $N = 948$, $k = 5$ items, and $u = 25$ unique response patterns (Appendix B).

Pattern	Frequency
00000	360
00001	3
00010	4
00100	2
00101	1
00111	2
01000	34
01001	6
01010	1
01011	1
01100	1
01110	1
01111	3
10000	3
10011	1
10101	1
10111	4
11000	7
11001	3
11010	3
11011	16
11100	6
11101	2
11110	5
11111	478

that is, d implies a, b, and c; for, the only state containing d is M, which contains all other motivations. Thus, $a, b, c \sqsubseteq d$. In addition, the motivations a and b do not imply each other, since the motivation states $\{a\}$ and $\{b\}$ contain the motivations a and b, but not b and a, respectively. Note that this does not exclude that both motivations can occur jointly, for example, with state $\{a, b\}$. In this way, we inspect all pairs of motivations to determine whether these motivations are in relation or not, thereby yielding the following motivational implications or relation pairs $a \sqsubseteq d$ and $b \sqsubseteq c \sqsubseteq d$ (including $b \sqsubseteq d$). Thus, this construction, which is one part of Birkhoff's theorem, the direction from set to relation, is concrete. The other direction from relation to set is also comprehensible and accordingly obtained. We define those subsets of the motivation domain to be motivation states that are consistent with all implications or pairs of the relation. For example, given the above relation $\sqsubseteq$, since d implies a, b, and c, we cannot take $\{a, d\}$ to be a motivation state since it is not consistent with the relation $\sqsubseteq$, which requires that if d is possessed, then all of the other motivations must also be possessed. This subset $\{a, d\}$ of the motivation domain contains d, but not the required, implied, motivations b and c. Since motivation a has no prerequisite motivation in the relation $\sqsubseteq$, that is, since it does not imply any other motivation, it can occur as the only motivation. That is, $\{a\}$ is consistent with $\sqsubseteq$, and thus, a derived motivation state. In this way, we can check for any subset of the motivation domain whether this subset contains with any of its motivations also all of the motivations implied by this motivation in the underlying relation. This constitutes the other part of Birkhoff's theorem. In accordance with the interpretation of motivation in possession or occurrence, we call this relation $\sqsubseteq$ corresponding to $\mathcal{M}$,

TABLE 4B External regulation, with $N = 1028$, $k = 5$ items, and $u = 30$ unique response patterns (Appendix B).

Pattern	Frequency
00000	233
00001	46
00010	44
00011	28
00100	100
00101	32
00110	118
00111	110
01000	2
01001	1
01010	4
01011	4
01100	7
01101	5
01110	30
01111	31
10000	5
10010	4
10011	6
10100	12
10101	7
10110	18
10111	26
11000	1
11010	1
11011	3
11100	7
11101	7
11110	23
11111	113

TABLE 4C Identified regulation, with $N = 1073$, $k = 5$ items, and $u = 29$ unique response patterns (Appendix B).

Pattern	Frequency
00000	197
00001	26
00010	9
00011	8
00100	2
00101	1
00110	4
00111	2
01000	1
01001	1
01100	5
01101	1
01110	21
01111	24
10000	12
10001	4
10010	4
10011	1
10100	2
10101	2
10111	3
11000	21
11001	6
11010	8
11011	2
11100	37
11101	14
11110	339
11111	316

which is the collection of all these derived pairs of motivations, a motivation co-occurrence relation (formally defined below).

Basically, the two representations are mathematically equivalent, but empirically they are interpreted at two different levels. From a practical viewpoint, the representation based on motivation structures is at the level of persons, whereas the co-occurrence relation is at the level of motivations "viewed as items." What do we mean by this? A motivation structure describes the feasible combinations of the motivations people can have. An element of the structure is the motivation state of a person, a collection of regulations that jointly characterize a person. In contrast, the representation based on relations asks for valid hierarchical dependencies, a relation $\sqsubseteq$, among the regulations, similar to ordering items, for example, by item difficulty. For a pair of motivations x and y, we set $x \sqsubseteq y$, if possessing motivation y entails possessing motivation x. This could result from or include, for example, when temporally one motivation (x) is attained before or at the same time as the other motivation (y). This implicational interpretation of motivation is general. In fact, if the concept of motivation structure is deemed to be empirically plausible, this entails the plausibility of implications between the motivations as the two formulations are connected mathematically as well as by interpretation. In addition, a special case of co-occurrence relation is the trivial relation, the diagonal, according to which no implications between the regulations are assumed, except for the reflexive implications, which are tautologies. Thus, mathematically also this case of completely unrelated regulations is contained in the definition of motivation co-occurrence relation.

Definition 5. Let M be the motivation domain. Assume that we can form pairs (x, y) of the regulations of the domain (e.g., if possession of y implies the possession of x). Let the set of all these pairs of regulations be denoted as $\sqsubseteq$. For a pair (x, y) in $\sqsubseteq$, we can also write $x \sqsubseteq y$. We call this set $\sqsubseteq$ a motivation co-occurrence relation if and only if it has the following additional properties, for any choice of regulations x, y, z of the domain:

1. $x \sqsubseteq x$ (reflexivity);
2. if $x \sqsubseteq y$ and $y \sqsubseteq z$, then $x \sqsubseteq z$ (transitivity).

That is, a motivation co-occurrence relation is a quasi-order (e.g., Davey and Priestley, 2002) on the motivation domain. If $x \sqsubseteq y$, we say that y implies x, and write $y \longrightarrow x$.

TABLE 4D Introjected regulation, with $N = 1168$, $k = 4$ items, and $u = 16$ unique response patterns (Appendix B).

Pattern	Frequency
0000	458
0001	21
0010	7
0011	3
0100	41
0101	33
0110	3
0111	8
1000	122
1001	14
1010	32
1011	18
1100	63
1101	54
1110	53
1111	238

TABLE 4E Autonomous motivation, with $N = 482$, $k = 10$ items, and $u = 59$ unique response patterns (Appendix B).

Pattern	Frequency	Pattern	Frequency
0000000000	89	0101011100	1
0000000001	5	0101111101	1
0000000010	1	0110011110	1
0000000100	1	0111000001	1
0000000110	1	0111111111	1
0000001100	1	1000000000	1
0000001111	1	1000000001	1
0000010000	5	1001111111	1
0000010001	2	1011111111	1
0000010011	1	1100000010	1
0000011000	7	1100111110	1
0000011001	3	1101011110	2
0000011010	1	1101111110	4
0000011011	1	1101111111	5
0000011100	21	1110011111	3
0000011101	7	1111000011	1
0000011110	12	1111100000	5
0000011111	4	1111100001	6
0000100000	1	1111100010	3
0001010000	1	1111100011	2
0001010010	1	1111101110	4
0010100011	1	1111101111	4
0100000000	5	1111110010	1
0100001110	1	1111110111	1
0100010000	1	1111111000	1
0100011000	1	1111111010	1
0100011110	6	1111111101	1
0100011111	4	1111111110	100
0100111110	3	1111111111	137
0100111111	2		

The axioms of reflexivity and transitivity are empirically necessary. It is obvious that under the motivation possession interpretation, the properties of reflexivity and transitivity necessarily hold. Reflexivity is logically trivial. Possession of x implies the possession of x. We must also have transitivity. If possession of z implies the possession of y, and possession of y, in turn, implies the possession of x, then possession of z must also imply the possession of x. Thus, if we derive implications between motivations consistently in this interpretation, the resulting relationship will be reflexive and transitive, a motivation co-occurrence relation.

The two representations based on motivation structure and motivation co-occurrence relation are basically equivalent. You have to additionally assume that the motivation structure is closed under (set) union and intersection.

Definition 6. Let $\mathcal{M}$ be a motivation structure. A motivation structure is closed under union or intersection if the unions or intersections of motivation states are again motivation states, respectively. We call the motivation structure $\mathcal{M}$ a motivation space if $\mathcal{M}$ is closed under union. The motivation structure $\mathcal{M}$ is called a motivation closure space if $\mathcal{M}$ is closed under intersection. If the motivation structure $\mathcal{M}$ is closed under union and intersection, $\mathcal{M}$ is called a quasi-ordinal motivation space. That is, quasi-ordinal motivation spaces are motivation spaces and motivation closure spaces.

In this article, because of Birkhoff's theorem, we are mainly concerned with quasi-ordinal motivation spaces.

Example 3. The motivation structures $\mathcal{M}_1$ and $\mathcal{M}_2$ in **Example 2** are motivation closure spaces, but not motivation spaces, and thus, not quasi-ordinal motivation spaces. The motivation structure $\mathcal{M}$ in **Eq. 1** is a quasi-ordinal motivation space.

We recap the old, but important theorem by Birkhoff (1937) informally, which allows us to switch between the two representations of motivation, on the one hand as a motivation structure and on the other as a motivation co-occurrence relation. In applications, you could choose between the two representations depending on the focus of the study. For example, if progressions in motivation states of persons during dynamic motivation assessment are tracked (**Figure 1**), the representation by sets can be adequate. Or, if dependencies between motivations regarding their occurrences are of interest (**Figure 3**), the representation by relations is more useful. We loosely present **Theorem 1** in its interpretation in motivation as a corollary, but the corresponding constructions were discussed in the example before.

Corollary 1. Let M be any motivation domain. For each motivation structure on M, which is a quasi-ordinal motivation space, you can construct a unique motivation co-occurrence relation on M, which is, in the above coherent sense, consonant with this space. Vice versa, for each motivation co-occurrence relation on M, a corresponding unique and consonant quasi-ordinal motivation space on M can be constructed. Since this is bijectively possible, meaning the two constructions define inverse transformations to each other, respecting the particular discrete properties, there is no loss of combinatorial information when changing from one representation to the other.

5. An empirical application

In this section, we discuss an illustrative example of the approach based on sets, relations, and IITA. We used data kindly provided by Müller et al. (2007). The data comprise

TABLE 4F Controlled motivation, with $N = 550$, $k = 9$ items, and $u = 134$ unique response patterns (Appendix B).

Pattern	Frequency	Pattern	Frequency	Pattern	Frequency	Pattern	Frequency
000000000	114	010000110	6	100100111	1	110110100	1
000000001	5	010000111	2	100110011	1	110110110	2
000000010	16	010001110	1	101000000	1	110111110	1
000000011	2	010001111	1	101000001	3	110111111	4
000000100	33	010010010	1	101000011	1	111000000	2
000000101	3	010010100	1	101000111	3	111000001	2
000000110	28	010100000	1	101001011	1	111000011	1
000000111	4	010100011	1	101011011	1	111000100	1
000001000	1	010100100	1	101100011	2	111000110	1
000001100	3	010100101	2	101100101	2	111000111	5
000001110	6	010100110	5	101100111	2	111001110	1
000001111	2	010101110	1	101101101	1	111001111	1
000010000	1	010111110	2	101110100	1	111010111	2
000010100	5	011000111	1	101111101	1	111011111	5
000010101	1	011100000	1	110000000	5	111100000	4
000010110	3	011111111	1	110000010	1	111100001	4
000010111	1	100000000	19	110000100	4	111100011	3
000011100	3	100000001	2	110000110	5	111100100	2
000011110	3	100000010	2	110000111	4	111100101	4
000011111	3	100000011	2	110001011	1	111100110	1
000100000	1	100000100	7	110001110	1	111100111	19
000100110	1	100000101	4	110001111	2	111101000	1
000101101	1	100000110	3	110010100	1	111101011	1
000111110	2	100000111	4	110010110	1	111101100	1
000111111	1	100001001	1	110011110	1	111101111	11
001000000	2	100001101	1	110011111	1	111110010	1
001000001	1	100001110	2	110100000	1	111110011	2
001000111	1	100010000	1	110100011	1	111110110	1
001010000	1	100010010	1	110100100	1	111110111	8
001100001	1	100010100	1	110100101	1	111111101	3
001111111	1	100010101	2	110100110	2	111111110	1
010000000	3	100011101	1	110100111	2	111111111	49
010000010	4	100011110	1	110101110	4		
010000100	4	100100000	1	110101111	1		

the learning motivation subscale scores of Austrian pupils in different school class subject areas. Investigated were five items each for the intrinsic regulation and external regulation subscales, and five and four items for the intermediate identified regulation and introjected regulation subscales, respectively. For further information about the scale and for the data sets of the empirical application, see **Appendices A, B**. Free software to run IITA analyses can be found in Schrepp (2006) and Ünlü and Sargin (2010).

In the original data, the scores ranged from (integer) 1−5, including missing values, with a higher score indicating a higher level of endorsement of a regulation. The data were dichotomized relative to the center at the mid-value 3 (balanced score) of the scale. The scores $x \leq 2$ (non-affirmative scores) were set to 0 and scores $x \geq 4$ (affirmative scores) were set to 1.

TABLE 4G All regulations, with $N = 180$, $k = 19$ items, and $u = 109$ unique response patterns (Appendix B).

Pattern	Frequency	Pattern	Frequency	Pattern	Frequency
0000000000000000000	15	0000011111100111111	1	1111110000111100000	1
0000000000000000010	3	0001000001000001110	1	1111110001111000001	1
0000000000000000100	3	0100000000111000111	1	1111110001111000100	2
0000000000000000110	7	0100000101111000111	1	1111110001111100101	2
0000000000000000111	1	0100001000000000000	1	1111110010000100111	1
0000000000000001100	2	0100001011111111110	1	1111110011111000000	1
0000000000000001110	1	0100011000000000010	1	1111110101111101011	1
0000000000000011100	1	0100100001111000100	2	1111110111111000011	1
0000000000000011110	1	0100100001111110111	1	1111110111111000111	1
0000000000000100001	1	0100100011111011110	1	1111110111111100011	1
0000000000000100111	1	0100111001111111111	1	1111110111111100101	1
0000000000001000000	1	1000000000000000000	1	1111110111111100111	1
0000000000010000000	1	1000000010000101101	1	1111111000001000000	1
0000000000110011111	1	1101000001111001111	1	1111111001111000000	1
0000000001000000010	1	1101101011111000100	1	1111111001111000100	1
0000000001101000100	1	1101111011111111111	1	1111111001111100100	1
0000000001110000000	2	1101111101111000001	2	1111111011111000011	1
0000000001110010100	1	1101111101111110111	1	1111111011111000101	1
0000000001111010110	2	1110011111111111111	1	1111111100111110111	1
0000000001111111111	1	1111100000000000101	1	1111111101111100000	1
0000001001000000110	1	1111100000001000000	1	1111111101111111111	1
0000001001111000100	1	1111100001111000000	6	1111111110001101111	1
0000001011110000110	2	1111100001111000001	1	1111111111111000000	1
0000010000000001001	1	1111100001111000010	1	1111111111111000111	3
0000010000000001110	1	1111100001111000011	1	1111111111111010010	1
0000010001000000111	1	1111100001111000100	3	1111111111111010111	1
0000010001100110101	1	1111100001111000101	1	1111111111111100001	2
0000010001110001101	1	1111100001111010000	1	1111111111111100100	2
0000010001110011110	1	1111100001111010100	1	1111111111111100111	1
0000010001110100100	1	1111100001111100001	1	1111111111111101000	1
0000010001110100111	1	1111100101111000001	1	1111111111111101111	2
0000010001111001110	1	1111101000111000111	1	1111111111111110011	1
0000010111111000101	1	1111101001111000010	2	1111111111111110111	3
0000011001110010100	1	1111101001111001110	1	1111111111111111101	2
0000011011110010100	1	1111101101111100111	1	1111111111111111111	24
0000011101111111111	1	1111101111111111111	1		
0000011111000111111	1	1111110000111000110	1		

Thus, the dichotomous score 0 encoded the negative manifest response (not endorsed regulation) and negative latent response (not possessed regulation), and 1 the positive manifest response (endorsed regulation) and positive latent response (possessed regulation). Response patterns containing at least one balanced score of 3 or missing values were discarded. The sample sizes for the used subsets of motivations ranged from $N = 180$ up to 1, 168 cases for all motivation variables altogether or only the introjected subscale, respectively (Table 2). For all resulting data sets, refer to Appendix B.

TABLE 4H For $A = \{i_1, i_3, \ldots, i_5, d_5, j_1, \ldots, j_4, e_3, e_4, e_5\}$, with $N = 342$, $k = 12$ items, and $u = 130$ unique response patterns (Appendix B).

Pattern	Frequency	Pattern	Frequency	Pattern	Frequency	Pattern	Frequency
000000000000	32	000011000011	1	101101010100	1	111110001000	2
000000000001	1	000011000100	1	101111011111	1	111110001101	2
000000000010	8	000011000110	2	101111100001	2	111110010000	1
000000000100	24	000011010100	1	101111101111	1	111110011111	1
000000000110	29	000011011110	1	101111110000	1	111110101011	2
000000000111	3	000011100110	1	110011011111	1	111110110011	1
000000001001	1	000011100111	1	110011111111	1	111110110111	1
000000001111	2	000011101110	1	111000000000	1	111110111011	1
000000010111	1	000011101111	2	111011010110	1	111110111101	1
000000100000	1	000011110110	1	111100000000	19	111110111111	1
000000100111	1	000011111110	1	111100000001	1	111111000000	3
000001000000	1	000011111111	3	111100000010	1	111111000100	1
000001000100	3	000100000100	2	111100000011	1	111111001100	1
000001000110	3	000100001111	1	111100000100	5	111111001101	1
000001000111	2	000100010110	1	111100000101	3	111111010011	1
000001001001	1	000111000110	1	111100000110	2	111111010101	1
000001010110	4	000111001111	1	111100000111	1	111111010110	1
000001011110	1	001000000110	1	111100001001	1	111111011110	1
000001110000	1	001010000101	1	111100010000	1	111111011111	1
000001110010	1	001010001101	1	111100100000	2	111111101000	1
000010000001	1	001111001101	1	111100100001	1	111111101111	10
000010000100	1	011111111011	1	111100111101	1	111111110000	2
000010000101	2	100000000000	1	111101000000	1	111111110010	1
000010000110	3	100000000100	1	111101000010	2	111111110111	7
000010000111	1	100000000110	1	111101000110	1	111111111000	2
000010001100	2	100000010110	1	111101000111	1	111111111001	2
000010001101	1	100000011101	1	111101010000	1	111111111011	2
000010001110	2	100010000010	1	111101101111	1	111111111100	2
000010001111	2	100010111101	1	111101111111	1	111111111101	4
000010010000	1	100011110011	1	111110000000	7	111111111110	1
000010101111	1	101000000111	1	111110000001	1	111111111111	39
000010110101	1	101011111111	1	111110000100	2		
000011000010	1	101100000000	1	111110000110	1		

In general, IITA is supposed to provide more reliable results with larger sample sizes. In practice, however, a solution computed for a smaller sample size can still be an empirically useful model. This may be the case if the derived IITA hierarchy yields an acceptable and also simple description of a higher number of motivation variables, scaled altogether in one go. This may be especially so, if no other alternative model is available or difficult to get, or if an exploratory first model is needed that covers a larger number of variables to inspect for their multivariate relationships. Such a preliminary model may be further adjusted in subsequent analyses. A strategy for how to perform sequential exploratory IITA analyses with more and more refined subsets of motivation variables is outlined, actually for the first time in the literature on IITA, with the example of this section. Thus, we also contribute to the methodology of IITA. The results obtained are illuminating as they constructively add insight into a dispute in SDT literature (Chemolli and Gagné, 2014; Sheldon et al., 2017).

TABLE 4| For $B = \{i_1, ..., i_5, d_1, ..., d_4\}$, with $N = 577$, $k = 9$ items, and $u = 52$ unique response patterns (Appendix B).

Pattern	Frequency	Pattern	Frequency
000000000	105	011001111	1
000000001	1	011100000	1
000000010	1	011111111	2
000000011	1	100000000	2
000000100	1	100111111	1
000000110	1	101111111	1
000000111	1	110000000	1
000001000	10	110000001	1
000001001	1	110001111	1
000001100	12	110010011	1
000001101	2	110011111	1
000001110	31	110101111	2
000001111	22	110111111	10
000010000	1	111001111	4
000101000	1	111100001	1
000101001	1	111101111	1
001010001	1	111110000	11
010000000	5	111110001	8
010000111	1	111110111	11
010001000	1	111111001	2
010001100	1	111111010	1
010001110	1	111111011	1
010001111	11	111111100	1
010011111	5	111111101	1
010101110	1	111111110	1
010111110	1	111111111	289

A remark regarding notation is in order. The manifest items or indicators of the instruments and their measured or underlying motivations are denoted with the same symbols. For example, we use e to stand for a manifest test item of the external regulation subscale and the corresponding latent external regulation. In particular, the variable names of the IITA solutions, for example, presented in the plots, are understood to be denoting the underlying motivations, not indicator variables. In the sequel, the five indicator variables and their corresponding (possibly equal) gradations of intrinsic regulation are $i_1, ..., i_5$ ($o = 5$); for external regulation the manifest and latent variable names are $e_1, ..., e_5$ ($l = 5$); and for identified regulation $d_1, ..., d_5$ ($n = 5$) and introjected regulation $j_1, ..., j_4$ ($m = 4$). Thus, the motivation domains for the poles of the internalization continuum are $M_i = \{i_1, i_2, i_3, i_4, i_5\}$ and $M_e = \{e_1, e_2, e_3, e_4, e_5\}$, and for the intermediate regulations $M_d = \{d_1, d_2, d_3, d_4, d_5\}$ and $M_j = \{j_1, j_2, j_3, j_4\}$. We performed IITA analyses and derived sets and relations for the individual subscales (Figure 3). We also distinguished and ran the analyses separately in autonomous and controlled motivations $M_a = M_i \cup M_d$ and $M_c = M_e \cup M_j$, respectively (Figure 4). In addition, all motivations were jointly analyzed $M = M_i \cup M_e \cup M_d \cup M_j$ (Figure 5). Motivation models for further subsets of the domain were also derived (Figure 6).

Those motivations are interpreted as the gradations of the types of intrinsic regulation, external regulation, identified regulation, and introjected regulation, or of autonomous motivation and controlled motivation, which may or may not be a same gradation, depending on the solution obtained by IITA, in the following sense. Whether two motivations denominate the same gradation will be judged on the basis of parallel or equally informative motivations. Given the data-analytically derived motivation co-occurrence relation by IITA on M or M_x, where $x = i, e, d, j, a, c$, we can obtain the corresponding quasi-ordinal motivation space. Any (two or more) regulations in M or M_x, which are then contained in the same motivation states of this space, are called parallel or equally informative. In a sense, they carry the same information regarding the distribution of motivation in the population. Equivalently, in the other representation, parallel or equally informative regulations imply each other in the motivation co-occurrence relation. That is, for any two equally informative gradations, each one is a possessed motivation if and only if the other is. Equally informative gradations occur always jointly. We take this to mathematically define those parallel gradations or regulations of the solution that are "equal."

A remark on this definition of "equality" is in order. Obviously, if two regulations as true constructs are equal objects, they necessarily, by identity, must be parallel. However, the converse may not be the case. One could imagine equally informative gradations being different constructs. But this is only hypothetical. In reality, by definition, parallel gradations can only be observed jointly. They cannot be separated empirically if the model holds true. In this sense, experimentally, we can only know their summative effect on another substantive variable of a study, but we have no information about what effect an individual part or gradation may have. This is reminiscent of entanglement in quantum mechanics in physics (e.g., Jaeger, 2009; Duarte, 2019), where we may know everything about a system, but nothing about its parts. If we cannot separate the parallel gradations, but only observe their "summative motivation," which may be an aggregate "motivation" different from the postulated basic motivations, we can take any of those gradations to represent their same "summative motivation." If this "summative motivation" is composed of only genuine gradations of the same basic regulation type, we assume or say that this "summative motivation" and its constituting parallel gradations are a same "gradation" of that regulation type.

TABLE 4J For $C = \{d_5, j_1, \ldots, j_4, e_1, e_2, e_5\}$, with $N = 706$, $k = 8$ items, and $u = 114$ unique response patterns (Appendix B).

Pattern	Frequency	Pattern	Frequency	Pattern	Frequency	Pattern	Frequency
00000000	273	01010010	1	10100100	1	11011011	1
00000001	19	01010110	1	10101001	4	11011100	1
00000010	11	01011001	1	10101011	1	11011110	2
00000011	2	01011011	1	10101111	1	11011111	2
00000100	14	01011110	1	10110001	3	11100000	2
00000101	1	01101001	2	10110100	1	11100001	4
00000110	7	01110000	1	10111001	5	11100111	1
00000111	2	01110001	1	10111011	1	11101000	2
00001000	2	01111111	1	10111111	1	11101001	8
00001001	3	10000000	29	11000000	14	11101010	1
00001101	1	10000001	10	11000001	1	11101011	1
00001110	1	10000010	3	11000010	1	11101101	2
00001111	1	10000011	2	11000011	3	11101111	4
00010000	4	10000100	3	11000100	2	11110000	3
00010001	1	10000101	1	11000101	1	11110001	16
00010110	1	10000110	1	11000110	1	11110011	1
00010111	1	10000111	1	11001000	4	11110100	2
00011011	1	10001000	8	11001001	1	11110101	1
00011110	1	10001001	4	11001011	1	11110111	1
00100000	2	10001100	1	11001100	1	11111000	5
00100001	2	10001101	1	11001110	1	11111001	25
00100100	1	10010000	2	11001111	1	11111010	2
00110001	1	10010001	1	11010000	3	11111011	12
00111111	1	10010110	1	11010001	4	11111100	1
01000000	20	10011001	1	11010010	1	11111101	11
01000001	2	10011011	1	11010100	1	11111110	1
01000010	1	10100000	1	11010111	2	11111111	50
01000100	2	10100001	4	11011000	3		
01010000	15	10100010	1	11011001	1		

To sum up, for practical purposes, if by data analysis, we detect two parallel regulations, letting aside variability of the solution, they may or may not be truly equal gradations (objects), which does not matter empirically, as a result of their unidentifiability, but in any case, they can be considered to be "equal" or equivalent in that more general sense. That is, if we term the latter notion of "equality" the parallel equality, in contrast to true (object) equality, the true equality cases entail parallel equality, but parallel equality encompasses additional cases that are not true equality. It is this generalized notion of equality that we use in the sequel.

In Figure 3, we can see the motivation co-occurrence relations, with their spaces below, that were detected for the separate subscales. We may expect a linear order since the items of each subscale should be measuring the same latent regulation type, in cumulative and possibly equal gradations.

The corresponding quasi-ordinal motivation spaces are, by application of Birkhoff's theorem, $\mathcal{M}_i = \{\emptyset, \{i_2\}, \{i_1, i_2\}, \{i_1, i_2, i_4, i_5\}, M_i\}$, $\mathcal{M}_e = \{\emptyset, \{e_3\}, \{e_3, e_4\}, \{e_3, e_4, e_5\}, M_e\}$, $\mathcal{M}_d = \{\emptyset, \{d_1, d_2, d_3\}, \{d_1, d_2, d_3, d_4\}, M_d\}$, and $\mathcal{M}_j = \{\emptyset, \{j_1\}, \{j_1, j_2\}, \{j_1, j_2, j_4\}, M_j\}$. Except for introjected regulation, in each solution, we have equally informative regulations, namely, $\{i_4, i_5\}$, $\{e_1, e_2\}$, and $\{d_1, d_2, d_3\}$. Thus, by definition (of parallel equality), we assume that these regulations denote equal gradations for each type of intrinsic regulation, external regulation, and identified regulation, respectively, and that the other gradations are distinct.

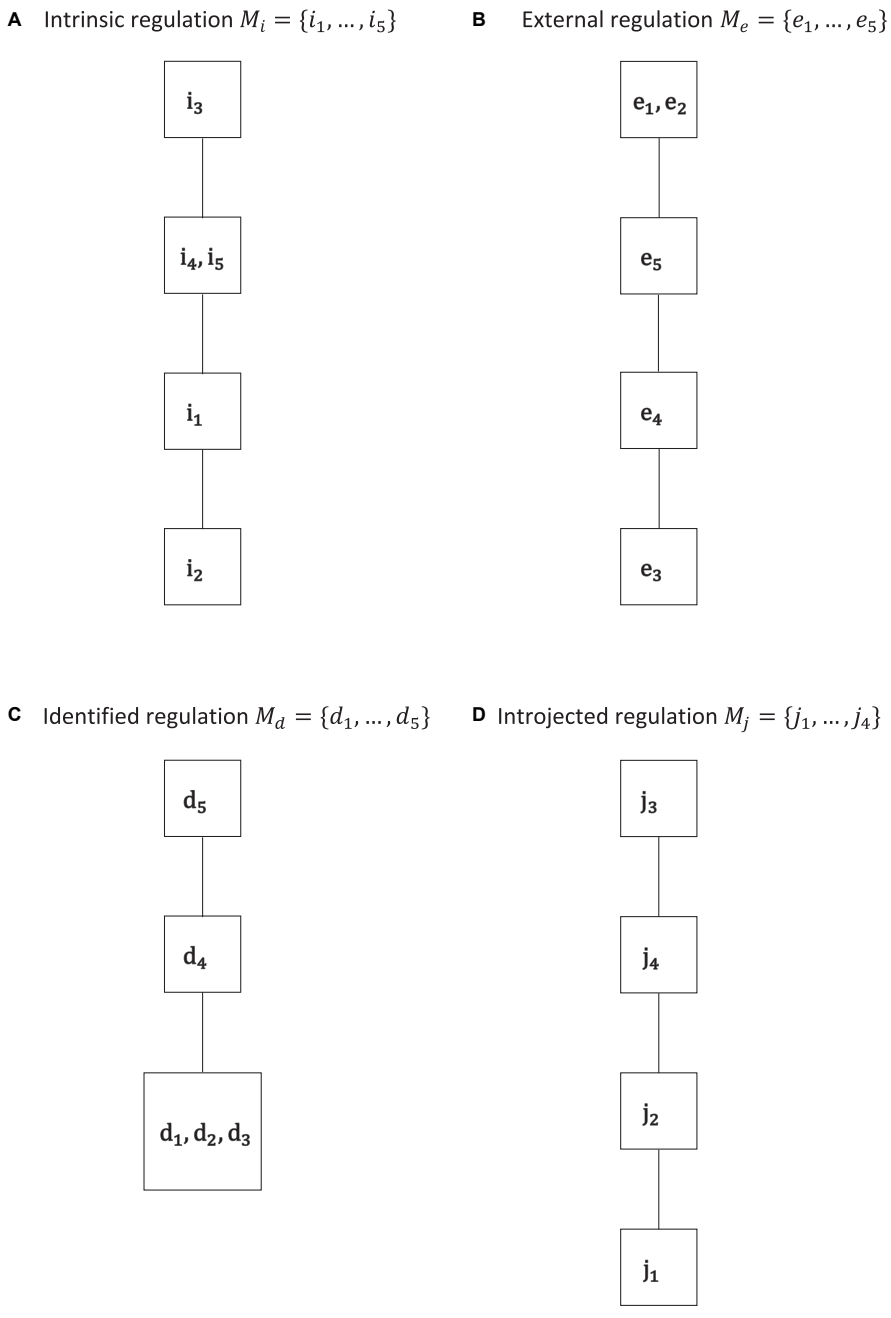


FIGURE 3
Motivation co-occurrence relations obtained by IITA of the empirical data set for the types of intrinsic regulation **(A)**, external regulation **(B)**, identified regulation **(C)**, and introjected regulation **(D)**. In each type, we see gradations that are cumulative, in the sense that they form linear orders or chains. Except for introjected regulation, there are equally informative gradations.

In SDT, researchers distinguish between autonomous (intrinsic and identified) motivations and controlled (external and introjected) motivations. This higher-order interpretation of the basic regulations is shown to be substantively important for the study of motivation and empirically adequate in motivation data (e.g., Deci and Ryan, 2008; Gagné et al., 2010, 2014). For example, evidence is reported by Gagné et al. (2010) that a second-order two-factor confirmatory factor analysis model can yield adequate fit, where the two higher-order factors group together the autonomous vs. controlled regulations as first-order factors. Consequently, in **Figure 4**, the relations and sets obtained by IITA in the autonomous and controlled motivations of the data set are reported.

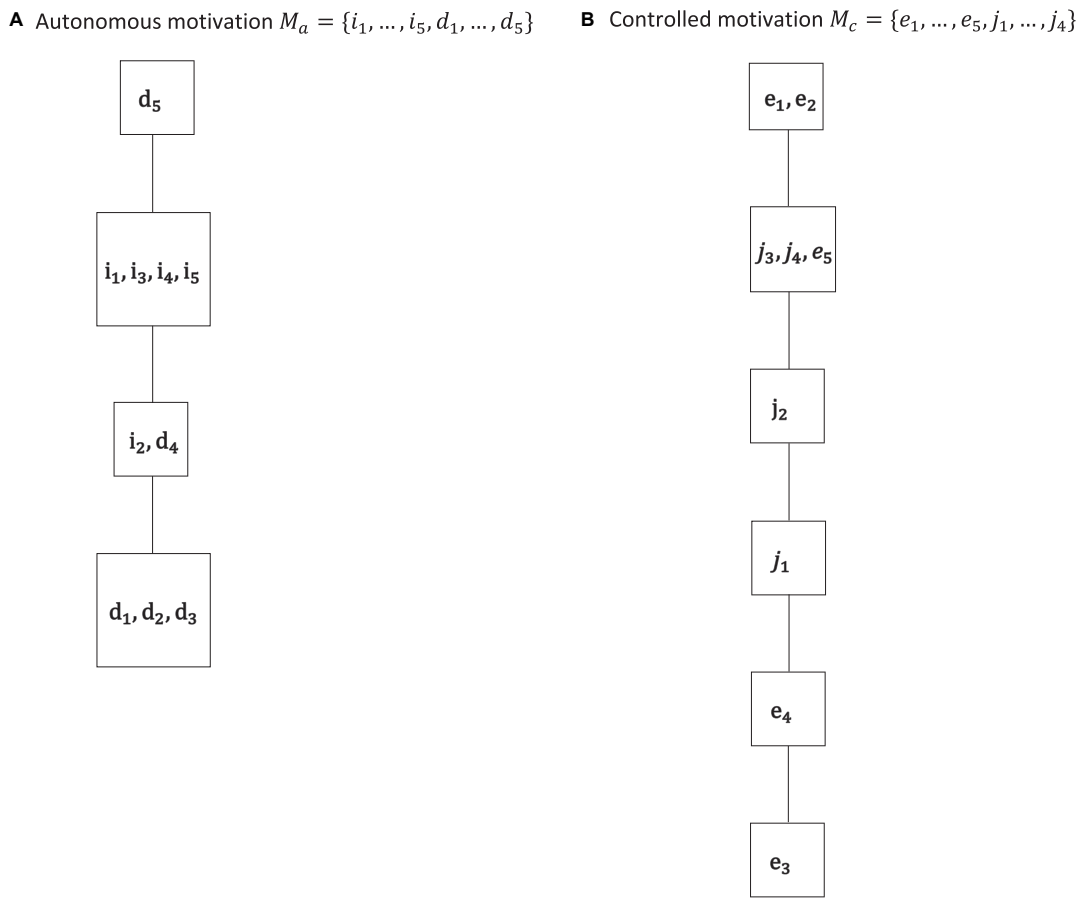


FIGURE 4
The IITA quasi-order solutions for autonomous motivation **(A)** and controlled motivation **(B)** of the empirical data set. By Birkhoff's theorem, the corresponding quasi-ordinal motivation spaces are $\mathcal{M}_a = \{\emptyset, \{d_1, ..., d_3\}, \{i_2, d_1, ..., d_4\}, \{i_1, ..., i_5, d_1, ..., d_4\}, M_a\}$, and $\mathcal{M}_c = \{\emptyset, \{e_3\}, \{e_3, e_4\}, \{e_3, e_4, j_1\}, \{e_3, e_4, j_1, j_2\}, \{e_3, ..., e_5, j_1, ..., j_4\}, M_c\}$, respectively. We see cumulative gradations of autonomous motivation and controlled motivation. This corroborates the higher-order interpretation of the basic motivations in these two non-basic autonomous and controlled motivations, which is also advocated in SDT literature.

Autonomous motivation and controlled motivation may be viewed as higher-order, aggregate types of motivation, different from SDT's postulated basic motivations. The cumulative (i.e., linearly ordered) gradations of autonomous motivation and controlled motivation obtained in the solutions corroborate that interpretation. In particular, the results indicate that the intermediate regulations are interweaved with their respective polar regulations in their common aggregate types, fairly arbitrary. The former can imply the latter and vice versa. They can be equally informative gradations of their underlying higher-order regulation types, even if they are different basic motivations. For example, intrinsic motivation i_2 and identified motivation d_4, basic motivations of different subscales, are equally informative, and by definition, they may be viewed as the same gradation of autonomous motivation. Or, the parallel regulations j_3, j_4, and e_5, if the model is true, can be viewed to be an equal gradation of the controlled motivation type. It is important to note, however, that the interpretation of the relations and their derivation from data are based on motivation possession or occurrence. Thus, possession of intrinsic motivation i_2 implies the possession of one, and thus all, of the identified motivations d_1, d_2, and d_3. External regulation e_1 occurs if and only if external regulation e_2 occurs, and in this case, all other regulations, external and introjected, must necessarily be also possessed motivations. Thus, if a person attains specific polar or intermediate regulations, we may infer that this person also possesses the motivations implied by those regulations.

In Figure 5, we ran the IITA algorithm on all motivations available in the data set for exploration, first and foremost. The sample size of $N = 180$ is small, relative to a large number of $k = 19$ motivation variables. That is okay since the aim was to look for possible interrelationships among the variables and to gain guidance on what motivation combinations to further investigate in narrow analyses.

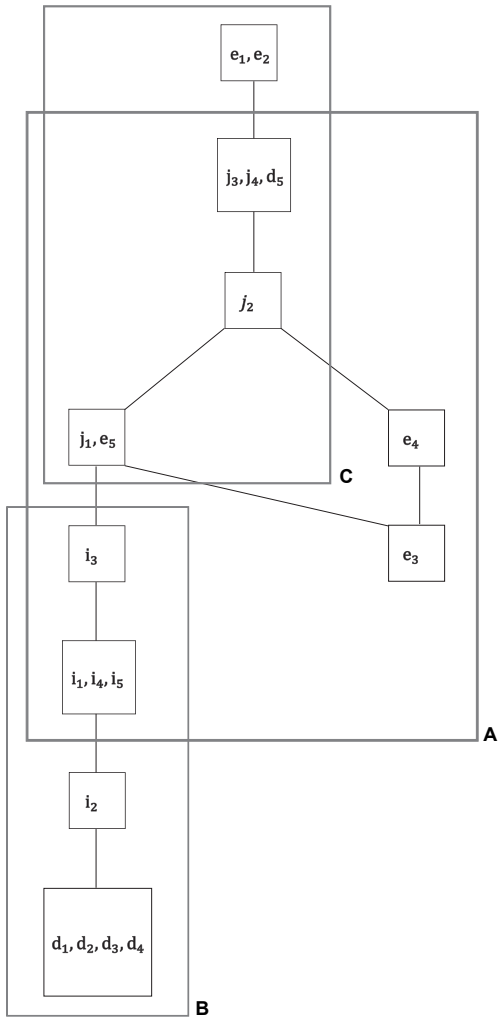


FIGURE 5
For exploratory analysis, all motivations of the empirical data set were jointly analyzed $M = M_i \cup M_e \cup M_d \cup M_j$. The sample size of $N = 180$ students for a total of $k = 19$ motivations was small for an overall IITA computation. Thus, based on the exploratory global solution (this figure), collections of motivations were delineated that were further analyzed more reliably, in lower numbers of jointly scaled motivations with larger sample sizes. In this example, the three subsets of motivations *A*, *B*, and *C* were regarded to be adequately sized and located within the overall graph for subsequent, more refined IITA analyses (**Figures 6A–C**).

In this example, we carved out the three subsets of motivations *A*, *B*, and *C* in **Figure 5**, which basically give a covering of the whole graph with overlapping motivations that will be used to mesh together the separate IITA solutions in those subsets. The choice of subsets made in **Figure 5** was also motivated by the following observations. Motivation e_5 seemed to be critical in the overall solution compared with what we obtained in **Figures 3B**, **4B**. Also, motivations d_5 and i_1, i_3, i_4, i_5 constituted linkages of autonomous motivation with controlled motivation, which is important information not contained in **Figures 3**, **4**. In addition, subsets *B* and *C* are chains, indicating that each of their motivations is more strongly interrelated, where *B* is a proper subset of the autonomous motivation scale and deviates from the relation obtained in **Figure 4A**. The binding subset between *B* and *C* is *A*, where the external motivations are not in accordance with the solution computed in controlled motivation in **Figure 4B**. Thus, these subsets with their extra or deviating structures may require further consideration.

The refined analyses in those subsets *A*, *B*, and *C* are reported in **Figure 6**.

The dashed boxes of two types indicate the knots where the three graphs are lumped together, respectively. For solutions *A* and *B* of **Figure 6**, there is only one option, to append solution *B* to solution *A*. For solutions *A* and *C* of **Figure 6**, the more refined solution *C* replaces the corresponding part of solution *A*. This also turned out to be the more preferential meshed solution in accompanying data analyses in even smaller subsets of the involved variables. Thus, the combined and final motivation co-occurrence relation in **Figure 7** was obtained.

The corresponding space of motivation states for this overall solution, as a graph, is shown in **Figure 8**.

From **Figures 7**, **8**, we can see how the different motivations are interrelated with each other in the interpretation of motivation possession. Deviations from a perfect chain structure are due to incomparability between (all) intrinsic regulations ($i_1, \ldots, i_5$) and (four) identified regulations (d_1, d_2, d_3, d_4) on the one hand and (two) external regulations (e_3, e_4) on the other, located in the bottom parts of these figures. Except for (one) identified motivation (d_5), which is intermingled with (two) introjected motivation (j_3, j_4), the hierarchy depicted in **Figure 7** can essentially be partitioned into the two chains of autonomous motivation (V) and controlled motivation (U). These parts are connected by an edge (w) between intrinsic motivation (i_1, i_3, i_4, i_5) and introjected motivation (j_1).

We have the following interpretation, in our example, regarding the discrimination or separability of controlled motivation and autonomous motivation. According to the hierarchical structure of their defining regulations, in **Figure 8**, controlled motivation and autonomous motivation cannot always be mutually exclusive, and thus separable. There are the motivation states, which only entail either controlled motivation or autonomous motivation, $\{e_3\}$ and $\{e_3, e_4\}$ or $\{d_1, d_2, d_3\}$, $\{i_2, d_1, d_2, d_3, d_4\}$, and $\{i_1, i_2, i_3, i_4, i_5, d_1, d_2, d_3, d_4\}$, respectively. However, the majority of the states imply autonomous as well as controlled motivations jointly. For example, the state $\{i_2, d_1, d_2, d_3, d_4, e_3\}$ is autonomous motivations i_2 and d_1, d_2, d_3, and d_4 as well as controlled motivation e_3. Thus, in a population of reference, depending on the distribution of the motivation states, we may end up sampling students either autonomously or controlled motivated

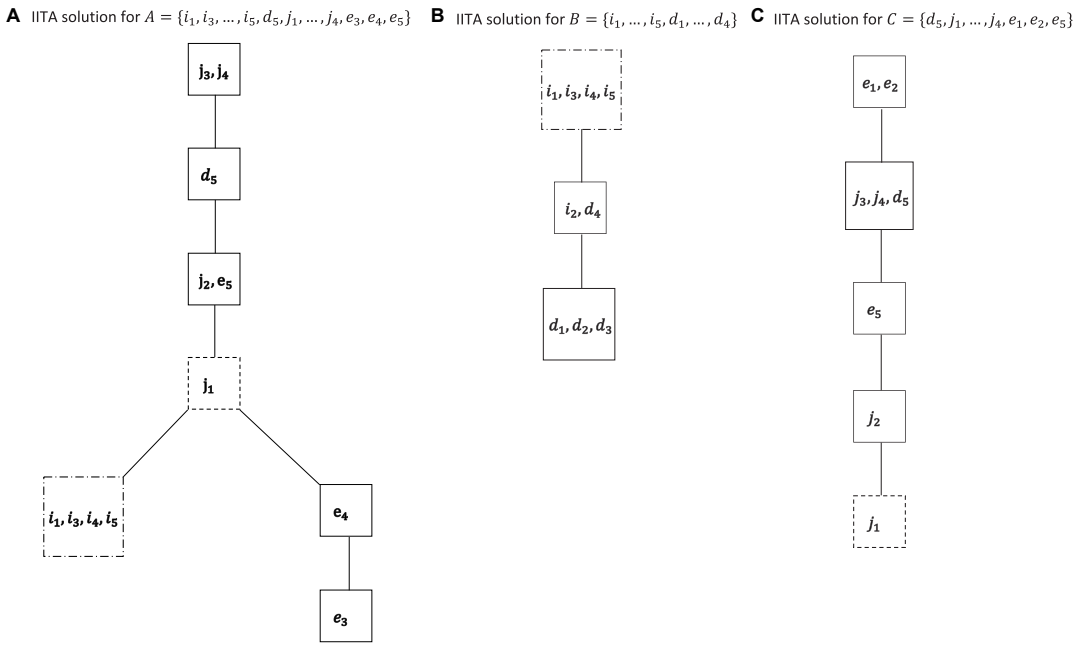


FIGURE 6
IITA computations for the motivations $A = \{i_1, i_3, \ldots, i_5, d_5, j_1, \ldots, j_4, e_3, e_4, e_5\}$ **(A)**, $B = \{i_1, \ldots, i_5, d_1, \ldots, d_4\}$ **(B)**, and $C = \{d_5, j_1, \ldots, j_4, e_1, e_2, e_5\}$ **(C)** of the empirical data set. The dashed boxes indicate the respective linkage knots for the three graphs (**Figure 7**).

(three or two states, respectively), autonomously as well as controlled motivated (eleven states), or none of them ($\emptyset$).

Importantly, the final solution can be utilized for the qualitative assessment of combinatorial dimensionality, in the following sense. We see that the derived multidimensional structure, that is, genuine quasi-order, is close to a unidimensional structure, meaning a chain. In particular, note that this is not the common definition of numerical dimensionality of the parameter or factor space of a statistical (e.g., item response or structural equation) model. Compared to the latter, the combinatorial view of dimensionality is more qualitative. In the solution, in **Figure 7**, omitting the external gradations e_3 and e_4, on $M\backslash\{e_3, e_4\}$, the restricted motivation co-occurrence relation is a single chain. Thus, the two-dimensional structure (i.e., linked two chains) of autonomous motivation V and controlled motivation U on the entire motivation domain becomes a unidimensional structure, if slightly pruned. That is, with the approach of this article, based on relations, we may see how close a multidimensional structure is to unidimensionality. In the empirical example it is very close, and where discrepancies in the structure may occur combinatorially. This may especially be so in relatively structured motivation data of validated SDT questionnaires.

In particular, this could dissolve a dispute of opinions put forth by Chemolli and Gagné (2014) and Sheldon et al. (2017). These authors advocated mutually exclusive views of a multidimensional vs. a unidimensional structure of motivation, respectively. In the empirical application of this article, at least, the two opinions only differ slightly, so both views basically seem to be justified in this example. In other applications, in the same manner, researchers could investigate the dimensionality of motivation. Presumably, if the theory holds true empirically, there should not be greater discrepancies between the unidimensional vs. multidimensional views of motivation, and this could be combinatorially quantified, similar to the example. However, this is only a conjecture, which needs to be tested critically in more popular scales; for example, in the Multidimensional Work Motivation Scale (MWMS), refer Gagné et al. (2010, 2014) and Trépanier et al. (2022).

To conclude, in **Table 2**, we summarize the frequency distributions of the dichotomous scores across motivation variables and the sample sizes underlying any of the IITA analyses of the empirical data sets. All analyzed data sets with individual binary entries can be found in **Appendix B**.

6. Usefulness of this approach for motivation research and limitation

We summarize the main conclusions from the results and the application of KST to SDT.

1. First and foremost, the problem of unidimensionality vs. multidimensionality of motivation, which has been disputed among SDT researchers (in particular, Chemolli and Gagné, 2014; Sheldon et al., 2017), can be more informatively assessed and resolved based on discrete KST combinatorial structures,

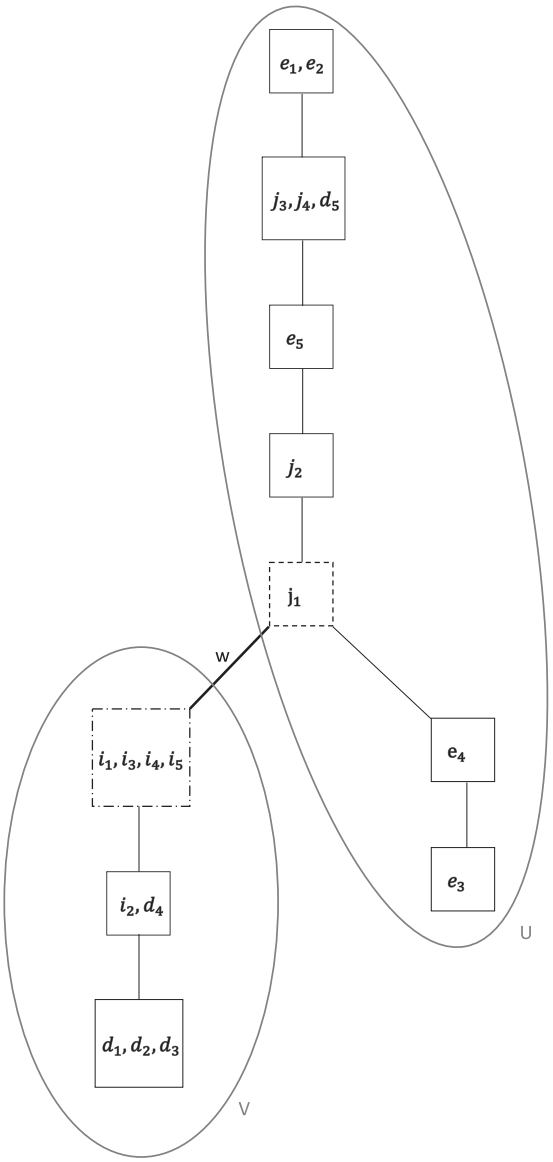


FIGURE 7
The concatenated final motivation co-occurrence relation $\sqsubseteq_f$ on the motivation domain of all regulations $M = \{i_1, ..., i_5, d_1, ..., d_5, j_1, ..., j_4, e_1, ..., e_5\}$, obtained by sequential exploratory IITA analyses of the data set. The graph is composed of two, connected by *w*, chains *U* (of essentially controlled motivation) and *V* (of autonomous motivation).

as compared to numerical dimensionality of parametric psychometric models. Take as an example, **Figure 7**. It is by no means obvious, or unique, how to derive this figure's hierarchy among the motivations by covariance structure models. You have too many fit and model selection indices (Tables 13.1 and 13.2 in West et al., 2012), and even if some of these indices may be indicative (but others generally are not) and you order motivations by their real unidimensional factor loadings, for instance, you may not be able to distinguish parallel or equally informative gradations nor account for the incomparable motivations (branching) located at the bottom part of the hierarchy. There may be workarounds though, more or less *ad hoc* and arbitrary solutions, which, however, may not be as straightforward and principled anymore as the natural combinatorial approach directly operating with orders.

2. In particular, we have contributed to the issue of the number of essential dimensions underlying the theory's posited motivations. We could only try one data set. Thus, we conjecture that in other empirical data sets of validated SDT instruments, if occurring, multidimensionality, essentially, is two dimensions only, which are also qualitatively close to unidimensional. The findings of this article corroborated the higher-order classification into autonomous motivation and controlled motivation, which is also advocated by other SDT researchers, the two branches of the hierarchy, but altogether the motivations were also close to a single chain structure. Whether the close proximity of one vs. two combinatorial dimensions may entail significant differences in real use cases is an interesting question. If necessary, based on the hierarchy of dependencies among the regulations, combinatorially, we could exploit the full information in the graph and differentiate between "mixed" dimensionalities, in the sense that certain motivations are chained, thus combinatorially unidimensional, whereas others are branched, thus combinatorially two (or more) dimensional.

3. In addition, this article adds insight into possible inconsistencies that may arise in applications of the theory in the way the different regulation types can be empirically distinguishable (e.g., integrated regulation seems to be difficult to identify). In particular, we conjecture that there may be equally informative or parallel gradations of the basic regulation types that may not be separable empirically, and thus remain unidentifiable or indistinguishable by experiments, most likely. Therefore, the generalized notion of equality, parallel equality, of the regulations of motivation can be adequate.

4. A key requirement for any statistical or mathematical model used to represent motivation qualitatively should be the flexibility that comes with representing structures. This is more the case with discrete structures than with numerical values. Numerical values such as factor loadings, factor scores, person abilities, or item difficulties, typical parameters of the structural equation or item response models, are more strongly aggregated numbers, implying their more restrictive natural linear orderings. The use of more general combinatorial structures such as surmise relations or even surmises functions (Doignon and Falmagne, 1999) allows for greater flexibility in the representing structures, for a more qualitative, thus diagnostic, conceptualization of motivation. Even a multidimensional numerical approach, which generally may be ambiguous and more difficult to interpret compared with a unidimensional model, can only yield partially ordered structures, by component-wise comparisons. In particular,

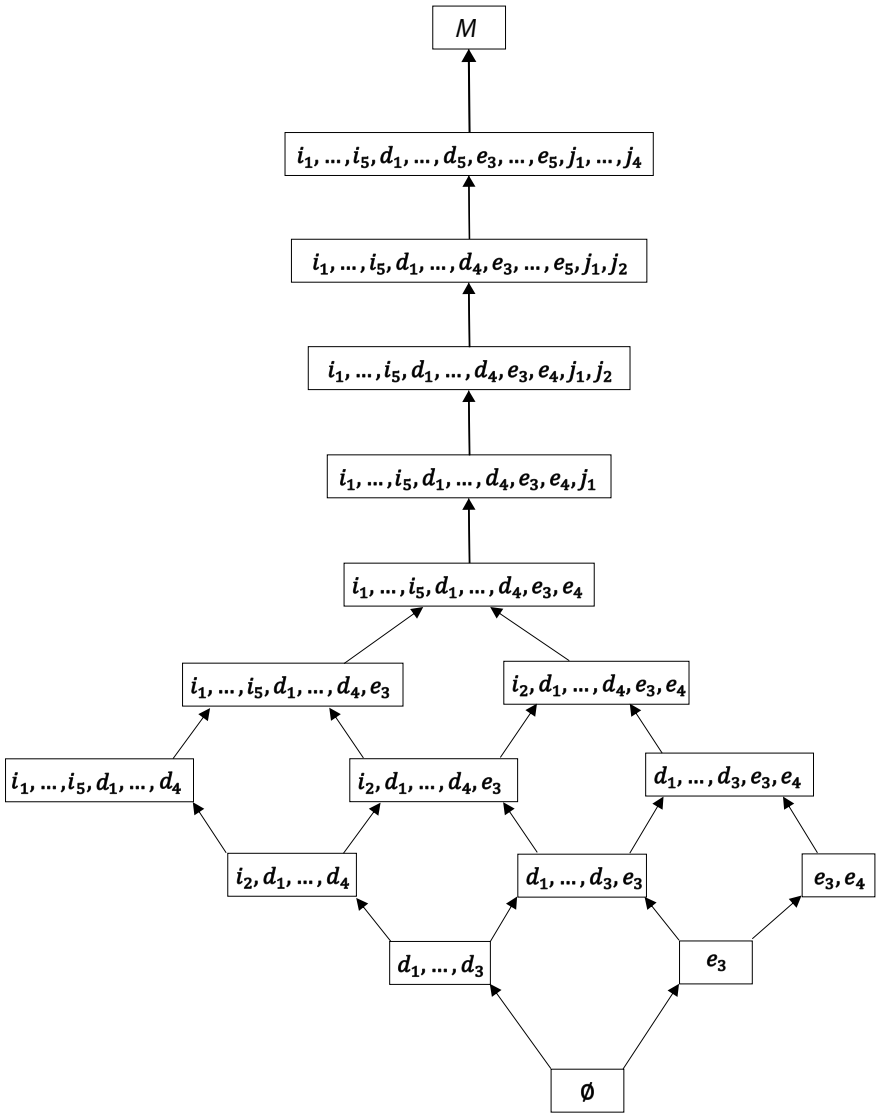


FIGURE 8
The quasi-ordinal motivation space $\mathcal{M}_f$ corresponding to the final IITA solution $\sqsubseteq_f$. The graphical representation is read from bottom to top, where a sequence of arcs linking a motivation state to a state located top of it means that the former is a subset of the latter. The bottom part of the graph indicates a branched multidimensional structure, the top part a linear unidimensional ("mixed" dimensionality).

partial orders do not differentiate parallel or equally informative motivations. Due to the use of more general KST structures, you can in principle describe more data sets in practice.

5. On top of that, in contrast to the commonly used psychometric approaches, the representing structures in KST are mathematically linked by theorems such as Birkhoff's, offering additional flexibility in the choices of equivalent representations, depending on the targeted use cases. For example, in the Rasch model (Van der Linden and Hambleton, 1997; Chemolli and Gagné, 2014), ability and difficulty estimates provide essentially separate representations, at the levels of persons and items, respectively. There is no direct and interpretable connection between these real-valued point estimates. In contrast, the more differentiated set representation (motivation structure) used for the persons is mathematically connected, by Birkhoff's theorem (**Theorem 1**), to the fine-grained order representation (co-occurrence relation) used for the motivations. The two levels, people and regulations, are combinatorially interrelated, which offers more ways for qualitatively representing and interpreting motivation-related results.

6. A cornerstone of KST is adaptive testing (for applications in education, see Falmagne et al., 2013). Originally, KST was developed and is predestinated for this purpose. Thus, the KST approach to SDT, as advocated by us, can provide the necessary framework to develop adaptive assessment procedures for motivation testing in SDT. To our knowledge, adaptive

testing has not been addressed at all in the SDT literature. To motivate, consider **Figure 7**. Under this co-occurrence relation, for example, if we test that a student is intrinsically motivated, possessing motivation i_2, we do not need to further test that student for the identified regulations d_1, d_2, d_3, and d_4, as possessing the latter is necessary, implied by possession of the former. Or, if a student is tested not to be externally motivated, not possessing regulation e_5, we can infer that this student must also not possess the motivations j_3, j_4, d_5, e_1, and e_2. Thus, the assessment can skip testing for those motivations. These are only simple illustrative examples. Dependency hierarchies accompanied by adaptivity of this sort can build the basis for efficient computerized adaptive assessment procedures for testing and also training motivation in various SDT application domains (e.g., to automate and efficiently test, or train, work motivation among employees of a bigger company).

7. Related to the preceding point, the KST approach to SDT also has the advantage that it can allow for dynamical motivation systems for the study of motivational behavior in time, particularly how motivation can progress in an orderly fashion, or perhaps become altogether unpredictable or even chaotic. The provision of a dynamical (including longitudinal) self-determination theory could be accomplished based on stochastic (learning) paths (Falmagne, 1993). The general idea underlying such an extension of SDT has been conceptually exemplified in section "Sets and relations among motivations," with **Figure 1**.

8. Mathematical, not statistical, modeling is possible. Mathematical models of motivation, such as discrete combinatorial structures (not only of KST), may help to understand, from a purely theoretical viewpoint, the logical foundations of the theory. In particular, the order-theoretic and algebraic properties of self-determination may be derived and studied mathematically (Ünlü, 2022), thus providing principled definitions of the central concept of self-determination. This may improve on the theory since in SDT, self-determination is commonly "defined" in a more or less *ad hoc* manner, based on such descriptive scoring protocols as the relative autonomy index (Grolnick and Ryan, 1987), an index that was criticized (Chemolli and Gagné, 2014; Ünlü, 2016). Or, the important internalization continuum of SDT can be mathematically defined by orders (**Figure 2B**).

9. Sets and relations derived for different motivations, for example, in the domains of work, sport, or learning, could be combinatorially compared, at a qualitative level. This could aim at finding structural invariants of motivation across different application contexts of the theory of self-determination.

The KST approach is relatively universal. It is generally applicable in any context in which implications between, even abstract, information units are of interest. For example, information units can be the geometry questions of a mathematical literacy test, where we are interested in whether the mastery of a geometry question implies the mastery of other questions of the test. In our context, the information units are regulations, among which the implications in their interpretation of motivation possession are considered. In principle, the application presented in this article is generalizable to other (psychological) theories (with appropriate operationalizations), if the system of interest, its defining units of information, the implications among those units, and their interpretation are delineated as the objects of the study. Not-so-obvious applications of this approach to other fields include system failure analysis (e.g., of a nuclear plant), where the failure of a component of the system may imply the failure of other components. Or, in medical diagnosis, the system is the patient, the information units are represented by the symptoms, and a physician examines whether the presence of certain symptoms in the patient imply that of others.

The present article has an obvious limitation, in that only one scale and data set were analyzed, which can be found in **Appendices A, B**, respectively. In future studies, more popular scales such as the MWMS (Trépanier et al., 2022), Sport Motivation Scale (SMS; Pelletier et al., 2013), and Academic Motivation Scale (AMS; Vallerand et al., 1992) with corresponding data sets could be examined using the KST and IITA approaches. This is an important direction for subsequent work. On the one hand, it remains to be seen if in those application domains with their questionnaires and data sets similar results can be obtained. On the other hand, it would be interesting to see how domain-specific motivation spaces and co-occurrence relations obtained in different application contexts of SDT structurally compare with each other. This could give (qualitative) information about the properties, similarities, and differences, of work motivation, sport motivation, and academic motivation, for example. To accomplish such a program, at this point, we have a personal recommendation addressed to the field of SDT. In the future, SDT researchers could consider providing their most pertinent data sets in a publicly accessible database. That would greatly facilitate research of the sort reported in this article.

7. Conclusion

Chemolli and Gagné (2014) advocated the use of qualitative motivation in self-determination theory (SDT). The approach presented here to motivation, an application of knowledge space theory (KST), is inherently qualitative, based on sets and relations in motivation regulations. In particular, this methodology allows us to treat each regulation as a separate variable of the motivation domain (cf., Koestner et al., 1996), and to represent every person by her or his total motivational profile, the motivation state, of all regulations within the person (cf., Vansteenkiste et al., 2009). This approach is flexible and general. It incorporates combinatorial unidimensionality

as well as multidimensionality of the possible motivation structures in a unified and natural manner by the use of linear orders (chains) and genuine quasi-orders, respectively. In the empirical application, we have seen that there are unidimensional substructures, such as that of autonomous motivation (**Figure 4A**) and controlled motivation (**Figure 4B**). We have also seen that the overall motivation structure was basically branched into autonomous motivation and controlled motivation and was two-dimensional in that sense (**Figure 7**). However, that motivation structure could be slightly pruned to become a chain, and thus unidimensional. Basically, two combinatorial dimensions reducible to one were observed.

In essence, these results are in accordance with, both, the seemingly contrary and exclusionary opinions expressed in the articles by Chemolli and Gagné (2014) and Sheldon et al. (2017), thereby bringing together and uniting their views. The conjecture is that in empirical data sets of reliable and valid SDT instruments if occurring, multidimensionality is essentially two dimensions that are qualitatively close to unidimensional. In future applications of this approach to other scales (e.g., Trépanier et al., 2022), one could check if this may be generally so. But in any case, one could study how close a multidimensional relational structure is to unidimensionality, and how to prune the structure accordingly, if possible.

The presented qualitative conceptualization of motivation based on KST can contribute to SDT in novel ways, practically and theoretically (cf. also section "Usefulness of this approach for motivation research and limitation"). Especially, adaptive assessment and training of motivation, ideally computerized, or the dynamical or longitudinal analysis of motivation progression in time, could be accomplished based on stochastic paths in empirically valid motivation spaces of the feasible motivation states (**Figures 1**, **8**). This article paves the way for these and other fruitful and interdisciplinary contributions to the study of motivation from the viewpoint of knowledge axiomatization and assessment in education. We have only considered the basic, yet powerful, concepts and their interpretations in motivation. More work in this direction is needed, especially by applied SDT researchers, to relate the combinatorial structures of motivation to behaviors and experimental outcomes.

In the end, we would like to describe this contribution as a cross-disciplinary methodology, an application of knowledge modeling in education and mathematical psychology, KST, to the study of motivation in, amongst others, social and personality psychology, SDT. As such, the KST approach is, most probably, not familiar to applied researchers in SDT, but we think it is worth the effort. To our knowledge, the qualitative representation and the analysis of motivation based on discrete combinatorial structures, as proposed in this article, are interesting new views on the quantitative treatment of motivation in the theory of self-determination.

Data availability statement

The original contributions presented in this study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s), nor the minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

The author thanks Dr. Erol Koçdemir for his invaluable, always competent support.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Albert, D., and Lukas, J. (eds). (1999). *Knowledge spaces: theories, empirical research, and applications*. Mahwah, NJ: Lawrence Erlbaum.

Assor, A., Vansteenkiste, M., and Kaplan, A. (2009). Identified versus introjected approach and introjected avoidance motivations in school and in sports: the limited benefits of self-worth strivings. *J. Educ. Psychol.* 101, 482–497. doi: 10.1037/a0014236

Birkhoff, G. (1937). Rings of sets. *Duke Math. J.* 3, 443–454. doi: 10.1215/S0012-7094-37-00334-X

Chemolli, E., and Gagné, M. (2014). Evidence against the continuum structure underlying motivation measures derived from self-determination theory. *Psychol. Assess.* 26, 575–585. doi: 10.1037/a0036212

Conesa, P. J., Onandia-Hinchado, I., Duñabeitia, J. A., and Moreno, M. Á. (2022). Basic psychological needs in the classroom: a literature review in elementary and middle school students. *Learn. Motiv.* 79:101819. doi: 10.1016/j.lmot.2022.101819

Davey, B. A., and Priestley, H. A. (2002). *Introduction to lattices and order*. Cambridge: Cambridge University Press. doi: P10.1017/CBO9780511809088

Deci, E. L., Hodges, R., Pierson, L., and Tomassone, J. (1992). Autonomy and competence as motivational factors in students with learning disabilities and emotional handicaps. *J. Learn. Disabil.* 25, 457–471. doi: 10.1177/002221949202500706

Deci, E. L., and Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY. Plenum. doi: 10.1007/978-1-4899-2271-7

Deci, E. L., and Ryan, R. M. (2000). The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. *Psychol. Inq.* 11, 227–268. doi: 10.1207/S15327965PLI1104_01

Deci, E. L., and Ryan, R. M. (2008). Facilitating optimal motivation and psychological well-being across life's domains. *Can. Psychol.* 49, 14–23. doi: 10.1037/0708-5591.49.1.14

Doignon, J.-P., and Falmagne, J.-Cl. (1985). Spaces for the assessment of knowledge. *Int. J. Man-Mach. Stud.* 23, 175–196. doi: 10.1016/S0020-7373(85)80031-6

Doignon, J.-P., and Falmagne, J.-Cl. (1999). *Knowledge spaces*. Berlin: Springer. doi: 10.1007/978-3-642-58625-5

Duarte, F. J. (2019). *Fundamentals of quantum entanglement*. Bristol: IOP Publishing. doi: 10.1088/2053-2563/ab2b33

Falmagne, J.-Cl. (1993). Stochastic learning paths in a knowledge structure. *J. Math. Psychol.* 37, 489–512. doi: 10.1006/jmps.1993.1031

Falmagne, J.-Cl., Albert, D., Doble, C., Eppstein, D., and Hu, X. (eds). (2013). *Knowledge spaces: applications in education*. Heidelberg: Springer. doi: 10.1007/978-3-642-35329-1

Falmagne, J.-Cl., and Doignon, J.-P. (2011). *Learning spaces*. Berlin: Springer. doi: 10.1007/978-3-642-01039-2

Gagné, M., Forest, J., Gilbert, M.-H., Aube, C., Morin, E., and Malorni, A. (2010). The Motivation at Work Scale: validation evidence in two languages. *Educ. Psychol. Meas.* 70, 628–646. doi: 10.1186/s13054-016-1208-6

Gagné, M., Forest, J., Vansteenkiste, M., Crevier-Braud, L., Van den Broeck, A., Aspeli, A. K., et al. (2014). The Multidimensional Work Motivation Scale: validation evidence in seven languages and nine countries. *Eur. J. Work Organ. Psychol.* 24, 178–196. doi: 10.1080/1359432X.2013.877892

Gagné, M., Parker, S. K., Griffin, M. A., Dunlop, P. D., Knight, C., Klonek, F. E., et al. (2022). Understanding and shaping the future of work with self-determination theory. *Nat. Rev. Psychol.* 1, 378–392. doi: 10.1038/s44159-022-00056-w

Grolnick, W. S., and Ryan, R. M. (1987). Autonomy in children's learning: an experimental and individual difference investigation. *J. Pers. Soc. Psychol.* 52, 890–898. doi: 10.1037//0022-3514.52.5.890

Jaeger, G. (2009). *Entanglement, information, and the interpretation of quantum mechanics*. Berlin: Springer. doi: 10.1007/978-3-540-92128-8

Koestner, R., Losier, G. F., Vallerand, R. J., and Carducci, D. (1996). Identified and introjected forms of political internalization: extending self-determination theory. *J. Pers. Soc. Psychol.* 70, 1025–1036. doi: 10.1037//0022-3514.70.5.1025

Müller, F. H., Hanfstingl, B., and Andreitz, I. (2007). *Skalen zur motivationalen Regulation beim Lernen von Schülerinnen und Schülern: adaptierte und ergänzte Version des Academic Self-Regulation Questionnaire (SRQ-A) nach Ryan & Connell [Scales of motivational regulation for student learning: adapted and supplemented version of the Academic Self-Regulation Questionnaire (SRQ-A) by Ryan & Connell] (Transl. A. Ünlü). In Institut für Unterrichts- und Schulentwicklung (ed.), Wissenschaftliche Beiträge [Scientific contributions] (Transl. A. Ünlü)*. Klagenfurt: Alpen-Adria-Universität, 1–17.

Pelletier, L. G., Rocchi, M. A., Vallerand, R. J., Deci, E. L., and Ryan, R. M. (2013). Validation of the revised Sport Motivation Scale (SMS-II). *Psychol. Sport Exerc.* 14, 329–341. doi: 10.1016/j.psychsport.2012.12.002

Roth, G., Assor, A., Niemiec, C. P., Ryan, R. M., and Deci, E. L. (2009). The emotional and academic consequences of parental conditional regard: comparing conditional positive regard, conditional negative regard, and autonomy support as parenting practices. *Dev. Psychol.* 45, 1119–1142. doi: 10.1037/a0015272

Ryan, R. M. (ed.) (2019). *The Oxford handbook of human motivation*. New York, NY: Oxford University Press. doi: 10.1093/oxfordhb/9780190666453.001.0001

Ryan, R. M., and Connell, J. P. (1989). Perceived locus of causality and internalization: examining reasons for acting in two domains. *J. Pers. Soc. Psychol.* 57, 749–761. doi: 10.1037//0022-3514.57.5.749

Ryan, R. M., and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* 55, 68–78. doi: 10.1037/0003-066X.55.1.68

Ryan, R. M., and Deci, E. L. (2002). "Overview of self-determination theory: an organismic dialectical perspective," in *Handbook of self-determination research*, eds E. L. Deci and R. M. Ryan (Rochester, NY: University of Rochester Press), 3–33. doi: 10.1111/bjhp.12054

Sargin, A., and Ünlü, A. (2009). Inductive item tree analysis: corrections, improvements, and comparisons. *Math. Soc. Sci.* 58, 376–392. doi: 10.1016/j.mathsocsci.2009.06.001

Schrepp, M. (1999a). Extracting knowledge structures from observed data. *Br. J. Math. Stat. Psychol.* 52, 213–224. doi: 10.1348/000711099159071

Schrepp, M. (1999b). On the empirical construction of implications between bi-valued test items. *Math. Soc. Sci.* 38, 361–375. doi: 10.1016/S0165-4896(99)00025-6

Schrepp, M. (2003). A method for the analysis of hierarchical dependencies between items of a questionnaire. *Methods Psychol. Res. Online* 19, 43–79.

Schrepp, M. (2005). About the connection between knowledge structures and latent class models. *Methodology* 1, 93–103. doi: 10.1027/1614-2241.1.3.93

Schrepp, M. (2006). ITA 2.0: a program for classical and inductive item tree analysis. *J. Stat. Softw.* 16:10. doi: 10.18637/jss.v016.i10

Sheldon, K. M., Osin, E. N., Gordeeva, T. O., Suchkov, D. D., and Sychev, O. A. (2017). Evaluating the dimensionality of self-determination theory's relative autonomy continuum. *Pers. Soc. Psychol. Bull.* 43, 1215–1238. doi: 10.1177/0146167217711915

Stefanutti, L. (2006). A logistic approach to knowledge structures. *J. Math. Psychol.* 50, 545–561. doi: 10.1016/j.jmp.2006.07.003

The R Core Team (2022). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Trépanier, S.-G., Peterson, C., Gagné, M., Fernet, C., Levesque-Côté, J., and Howard, J. L. (2022). Revisiting the Multidimensional Work Motivation Scale (MWMS). *Eur. J. Work Organ. Psychol.* doi: 10.1080/1359432X.2022.2116315

Ünlü, A. (2006). Estimation of careless error and lucky guess probabilities for dichotomous test items: a psychometric application of a biometric latent class model with random effects. *J. Math. Psychol.* 50, 309–328. doi: 10.1016/j.jmp.2005.10.002

Ünlü, A. (2007). Nonparametric item response theory axioms and properties under nonlinearity and their exemplification with knowledge space theory. *J. Math. Psychol.* 51, 383–400. doi: 10.1016/j.jmp.2007.07.002

Ünlü, A. (2011). A note on the connection between knowledge structures and latent class models. *Methodology* 7, 63–67. doi: 10.1027/1614-2241/a000023

Ünlü, A. (2016). Adjusting potentially confounded scoring protocols for motivation aggregation in organismic integration theory: an exemplification with the relative autonomy or self-determination index. *Front. Educ. Psychol.* 7:272. doi: 10.3389/fpsyg.2016.00272

Ünlü, A. (2022). Mathematical self-determination theory. doi: 10.31234/osf.io/jz76k

Ünlü, A., and Sargin, A. (2010). DAKS: an R package for data analysis methods in knowledge space theory. *J. Stat. Softw.* 37:2. doi: 10.18637/jss.v037.i02

Ünlü, A., and Schrepp, M. (2021). Generalized inductive item tree analysis. *J. Math. Psychol.* 103:102547. doi: 10.1016/j.jmp.2021.102547

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Brière, N. M., Senécal, C., and Vallières, E. F. (1992). The Academic Motivation Scale: a measure of intrinsic, extrinsic, and amotivation in education. *Educ. Psychol. Meas.* 52, 1003–1017. doi: 10.1177/0013164492052004025

Van der Linden, W. J., and Hambleton, R. K. (eds). (1997). *Handbook of modern item response theory*. New York, NY: Springer. doi: 10.1007/978-1-4757-2691-6

Van Leeuwe, J. F. J. (1974). Item tree analysis. *Ned. Tijdschr. Psychol.* 29, 475–484.

Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., and Lens, W. (2009). Motivational profiles from a self-determination perspective: the quality of motivation matters. *J. Educ. Psychol.* 101, 671–688. doi: 10.1037/a0015083

West, S. G., Taylor, A. B., and Wu, W. (2012). "Model fit and model selection in structural equation modeling," in *Handbook of structural equation modeling*, ed. R. H. Hoyle (New York, NY: Guilford Press), 209–231.

Appendix

Appendix A. German modification Academic Self-Regulation Questionnaire (SRQ-A)

The SRQ-A questionnaire, in its original two versions, can be retrieved from https://selfdeterminationtheory.org/self-regulation-questionnaires/ (as of 30 October 2022). Validations of this scale were presented by Ryan and Connell (1989) and Deci et al. (1992). The adapted and supplemented version of the SRQ-A in German by Müller et al. (2007), which is used for the empirical application, can be retrieved from https://ius.aau.at/wp-content/uploads/2016/01/mui_fragebogen.pdf (as of 30 October 2022). We have decided not to translate this scale into English, for the following reason. The German modification seems to be more than a mere translation of the original SRQ-A. According to the authors, only nine of the seventeen items are reused, just translated, items of the original SRQ-A. Understandably, it has only been validated in the German language (Müller et al., 2007). Thus, to avoid the risk of any distortion of its validated properties by an unverified translation (e.g., from German to English), we reproduce this scale in its original form in **Table 3**. Other works could study possible translations of this German modification of the SRQ-A into different languages (such as English), to validate or invalidate it in further settings.

Appendix B. Data sets

The binary data sets used for the empirical application are available in **Tables 4A–4J**, as frequency tables containing the observed response patterns with their respective absolute frequencies, for intrinsic regulation (**Table 4A**), external regulation (**Table 4B**), identified regulation (**Table 4C**), introjected regulation (**Table 4D**), autonomous motivation (**Table 4E**), controlled motivation (**Table 4F**), all regulations (**Table 4G**), and for the subsets of motivations $A = \{i_1, i_3, ..., i_5, d_5, j_1, ..., j_4, e_3, e_4, e_5\}$ (**Table 4H**), $B = \{i_1, ..., i_5, d_1, ..., d_4\}$ (**Table 4I**), and $C = \{d_5, j_1, ..., j_4, e_1, e_2, e_5\}$ (**Table 4J**). Note that the sequence of how the response patterns are laid out in a data set according to their absolute frequencies is irrelevant for computations. The data sets made available in this appendix were analyzed by IITA with free software (Schrepp, 2006; Ünlü and Sargin, 2010).

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 22 February 2023
DOI 10.3389/fpsyg.2023.1060963

OPEN ACCESS

EDITED BY
Fco. Pablo Holgado-Tello,
National University of Distance Education (UNED), Spain

REVIEWED BY
Giorgia Bussu,
Uppsala University,
Sweden
Fausta Micanti,
University of Naples Federico II, Italy

*CORRESPONDENCE
Hashem Salarzadeh Jenatabadi
✉ jenatabadi@um.edu.my
Nur Anisah Mohamed
✉ nuranisah_mohamed@um.edu.my

SPECIALTY SECTION
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

RECEIVED 04 October 2022
ACCEPTED 19 January 2023
PUBLISHED 22 February 2023

CITATION
Mohamed NA, Alanzi ARA, Azizan NA, Azizan SA, Samsudin N and Jenatabadi HS (2023) Evaluation of depression and obesity indices based on applications of ANOVA, regression, structural equation modeling and Taguchi algorithm process. *Front. Psychol.* 14:1060963. doi: 10.3389/fpsyg.2023.1060963

Evaluation of depression and obesity indices based on applications of ANOVA, regression, structural equation modeling and Taguchi algorithm process

Nur Anisah Mohamed[1]*, Ayed R. A. Alanzi[2], Noor Azlinna Azizan[3], Suzana Ariff Azizan[4], Nadia Samsudin[4,5] and Hashem Salarzadeh Jenatabadi[4]*

[1]Faculty of Science, Institute of Mathematical Sciences, Universiti Malaya, Kuala Lumpur, Malaysia, [2]Department of Mathematics, College of Science and Arts in Gurayat, Jouf University, Gurayat, Saudi Arabia, [3]College of Business Administration, Prince Sultan University, Riyadh, Saudi Arabia, [4]Department of Science and Technology Studies, Faculty of Science, Universiti Malaya, Kuala Lumpur, Malaysia, [5]Faculty of Social Sciences and Liberal Arts, UCSI University, Kuala Lumpur, Malaysia

Introduction: Depression and obesity are the main threat among women which have been considered by many research scholars in psychology studies. In their analysis for measuring and estimating obesity and depression they were involving statistical functions.

Methods: Regression, Analysis of Variance (ANOVA), and in the last two decades Structural Equation Modeling are the most familiar statistical methods among research scholars. Taguchi algorism process is one the statistical methods which mostly have been applying in engineering studies. In this study we are looking at two main objectives. The first one is to introduce Taguchi algorism process and apply it in a case study in psychology area. The second objective is challenging among four statistical techniques include ANOVA, regression, SEM, and Taguchi technique in a same data. To achieve those aims we involved depression and obesity indices with other familiar indicators contain socioeconomic, screen time, sleep time, and usage fitness and nutrition mobile applications.

Results and discussion: Outputs proved that Taguchi technique is able to analyze some correlations which are not achieved by applying ANOVA, regression, and SEM. Moreover, SEM has a special capability to estimate some hidden correlations which are not possible to evaluate them by using ANOVA, regression, and even Taguchi method. In the last, we found that some correlations are significant by SEM, however, in the same data with regression those correlation were not significant. This paper could be a warning for psychology research scholars to be more careful with involving statistical methods for measuring and estimating of their research variables.

KEYWORDS

Taguchi, obesity and depression, mHealth apps, women's obesity, experimental design

1. Introduction

Over time, there has been a consistent rise in the number of people receiving a depression diagnosis. The patient's ability to work, financial situation, and interpersonal relationships are all impacted by this mental disease (Muharam et al., 2018). Passive behaviors such as disinterest, guilt-ridden thoughts, low self-esteem, lack of sleep, poor appetite, perpetual sadness, or signs of weariness can all be indicators of depression (El-Gilany et al., 2018; Zeng et al., 2018). Being depressed on a

day-to-day basis results in a major handicap that can cause mental and behavioral difficulties (Archana et al., 2017). It is very likely that this condition will have an effect on the patient's physical well-being, which will ultimately lead to an elevated risk of morbidity and mortality (Elamoshy et al., 2018; Jia et al., 2018; Muharam et al., 2018). According to the World Health Organization (WHO), more than 300 million people worldwide experienced symptoms of depression in World Health Organization (2017). However, earlier research found that women, rather than men, were more likely to suffer from depression (Jin et al., 2016). It was shown that hormonal changes, such as those that occur during puberty, pregnancy, and menopause, were the most significant contributors to depression in women. Particularly after giving birth, a woman needs to have extra care and obtain the right kind of health care priorities because any unpleasant act can cause depression at this stage, which will be devastating to the entire family (Correa et al., 2017). In addition, a woman needs to obtain the right kind of health care priorities because any unpleasant act can cause depression at this stage.

On the other hand, the epidemic of overweight and obesity is quickly spreading, especially to younger age groups; as a result, it has become a serious concern for healthcare systems due to the tremendous economic and psychosocial load it places on those systems (Rippe, 2021). In 2017, it was estimated that 13% of the adult population globally was obese, and this number has consistently risen over the past few decades in both developed and developing countries (650 million people) (World Health Organization, 2017). Several types of cancer, heart disease, and diabetes type 2 are just a few of the many chronic illnesses for which obesity is a major risk factor.

On the other side, technology can help people reach their health goals. The Global Observatory for eHealth defines mobile health (mHealth) as the "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices" (World Health Organization, 2011) has become a major focus in the delivery of health care in recent years because of its potential impact on a wide range of health outcomes (Boulos et al., 2014). With the use of mobile technology, psychosocial and health behavior therapies are now accessible to a wider range of patients, in their natural environments, and in real time (Heron and Smyth, 2010). Mobile applications, sometimes known as "apps," enable remote access to health services by linking patients with health experts all over the world in a safe, confidential, and secure manner, with quick results (Qudah and Luetsch, 2019). There are now a great number of smartphone apps that may be downloaded to help manage the symptoms of anxiety and depression.

Healthcare professionals are turning to mobile health (mHealth) technologies (Yuan et al., 2015), such as nutrition and fitness apps (Bt Wan Mohamed Radzi et al., 2020; Taylor et al., 2022) to support women in managing obesity and depression from home. The nutritional apps engage women because they are convenient for monitoring daily food intake for healthier eating. Meanwhile, fitness apps help women chart their weight loss progress and BMI levels while following various exercises (Cho et al., 2015). Women also benefit from quick and easy access to health information and use mHealth apps to communicate with health professionals and peers.

There is some encouraging evidence of their usefulness, but this still has to be validated because the majority of them have not been planned or studied with the level of rigor required (Rathbone and Prescott, 2017). Moreover, inadequate scientific coverage, inaccurate weight-related information, lack of important evidence-based features, lack of involvement of health-care experts in the development process, overlooking behavior change techniques, and not undergoing rigorous scientific testing are some of the issues that plague the majority of commercial mobile apps for weight loss and management (Bardus et al., 2016).

1.1. Gaps in previous psychology studies

There are three gaps in previous psychology studies regarding measuring depression and obesity. Some previous studies just focused on estimating either obesity or depression. There are few studies also considered depression as an input for estimating obesity, or they involved obesity as an input for estimating depression. However, based on some former studies we found that there is a correlation between depression and obesity. The first gap of these types of studies is that lack of analysis for estimating both obesity and depression as two dependent variables in a single model. We full-fill this gap with introducing a single model with two dependent variables (depression and BMI) based on SEM technique. In terms statistical methods, ANOVA, simple correlation, regression, and SEM are the most familiar statistical methods among research scholar in psychology studies. These statistical modeling or even mathematical modeling techniques including neuro-fuzzy inference systems can determine the impact of significant independent (input) variables on dependent variable/s.

They are unable to demonstrate, however, which levels or groups of independent variables result in higher, lower, or nominal dependent variable (output) rate levels. Therefore, the question of what level or category of independent variables results in higher or lower dependent variables cannot be addressed by statistical or mathematical modeling techniques. Based on our research variables we could turn this question to, what level of usage nutrition and fitness apps, sleep time, screen time, and demographic leads women to have high level of depression or BMI? Therefore, Taguchi method can answer this question. The second gap of former in phycology studies that, based on best on knowledge there is lack of studies to answer this question. In other words, Taguchi technique, been applied in various engineering studies but there is scarce evidence for applying this pattern in psychology studies. The final gap is that, while they frequently used fitness and nutrition apps separately, it was uncommon to find apps that effectively treated both obesity and depression using a single model or pattern.

1.2. Objectives of the study

Based on above matters, this study is trying to achieve the following objectives:

First objective: to find out the impact of nutrition apps and physical apps usage on women's obesity and depression with application of ANOVA. This is a common analysis which have been done in previous studies.

Second objective: to estimate the effectiveness of nutrition apps and physical apps (with other research variables) on women's depression and obesity based on regression modeling. In this part of the study, we will have separates models for estimating BMI and depression.

Third objective: to estimate the effectiveness of nutrition apps and physical apps (with other research variables) on women's depression and obesity in a single model based on SEM.

Fourth objective: to recognize the combination of levels of both types of nutrition and physical apps (with other research variables) leads

to higher obesity and depression. For this objective of the study, we applied Taguchi algorithm process.

Fifth objective: To compare outputs of data analysis with different above statistical techniques.

2. Materials and methods

2.1. Taguchi method structure

In the Taguchi method, standard deviation and variation need to be measured for the expected value. The observed values were spread out from the expected value by a high standard deviation due to noise factors. The lower standard deviation indicates that the observed values are near the expected value due to noise factors. Observed and noise factor values can be controlled by the Signal-to-Noise (SN) ratio. The SN ratio effects noise factors on performance features and quantifies the variability (Salarzadeh Jenatabadi et al., 2016). The formula to calculate the signal to noise ratio (Manigandan et al., 2020) is given in Table 1.

The Taguchi method process analysis is as follows, which refers to the previous literature (Salarzadeh Jenatabadi et al., 2016):

Step 1: Identifying the Indicators.
Step 2: Calculate the weight of each indicator.
Step 3: Choosing an appropriate experimental design based on the Taguchi method.
Step 4: Identifying the Optimal Levels of the Indicators.
Step 5: Data Analysis.
Step 6: Assessing the Factors in the Columns of the Orthogonal Array.
Step 7: Introducing suitable patterns based on the optimal levels.

2.2. Participants

The sample size estimation was calculated using Krejcie and Morgan's method (Chuan and Penyelidikan, 2006). The formula below is commonly used to establish the total sample size requirement (Krejcie and Morgan, 1970):

$$s = \frac{\chi^2 NP(1-P)}{d^2(N-1)} + \chi^2 P(1-P) \quad (1)$$

In the above equation, s is the required sample size; χ^2 is the table value of chi-square for one degree of freedom at the desired confidence level; N is the population size; P is the population proportion; and d is the degree of accuracy expressed as a proportion (0.05). This research subject focused on Malaysian women who lived in urban and populated cities, i.e., Kuala Lumpur, Selangor, Penang, and Johor. The data was distributed after the MCO *via* online questionnaires by sending the questionnaire links through WhatsApp to 878 respondents.

Note: Approaching to the participant were based on two data sets (a) University Malaya grant [title: New Framework and Statistical Approaches for Health Index Studies: Case Study in Malaysia (GPF066B-2018)] and (b) our previous studies data set (Bt Wan Mohamed Radzi et al., 2020, 2021).

TABLE 1 Signal-to-Noise (SN) application.

Optimisation type	Calculation of SN
Lower is better	$SN = -10\log\frac{1}{n}\sum_{i=1}^{n} y_i^2$
Larger is better	$SN = -10\log\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}$

Where, n, y_i, and i shows the repetitions in the experiments. To be concise, y is the response factor (BMI and level of depression) and n is the number of observations.

2.3. Measurement

The research variables were classified into five sections: demographic information, lifestyle, BMI, depression level, and mHealth app frequency. The demographic details were measured in terms of four criteria as follows:

- Age: (a) 21–25 years old (value 1), (b) 26–30 years old (value 2), (c) 31–35 years old (value 3), and (d) over 35 years old (value 4).
- Education: (a) less than a high school diploma (value 1); (b) a high school diploma (value 2); (c) a diploma (value 3); (d) a bachelor's degree (value 4); (e) a master's or PhD degree (value 5).
- Job Experience: (a) no job experience (value 1); (b) less than 3 years (value 2); (c) 3–6 years (value 3); (d) 6–10 years (value 4); and (e) more than 10 years (value 5).
- Income: (a) less than RM 2,000 (value 1); (b) between RM 2,000 and RM 3,000 (value 2); (c) between RM 3,000 and RM 4,000 (value 3); (d) between RM 4,000 and RM 5,000 (value 4); (e) greater than RM 5,000 (value 5).

Lifestyle was measured based on Nakayama et al. (2001) and Khajeheian et al. (2018) research. The indicators included average working hours per day; physical activity per week; average screen time use per day (e.g., TV, smartphone, tablet); and average sleeping hours per night. These indicators were grouped as follows:

- Working hours: (a) none; (b) less than 7 h; (c) 7–8 h; (d) 8–9 h; (e) more than 9 h.
- Physical activity: (a) none; (b) once; (c) twice; (d) three times; (e) four times; (f) more than four times.
- Screen time: (a) less than 1 h; (b) 1–2 h; (c) 2–3 h; (d) 3–4 h; (e) more than 4 h.
- Sleeping hours: (a) less than 6 h; (b) 6–7 h; (c) 7–8 h; (d) 8–9 h; (e) more than 9 h.

Individuals' BMI ranges were calculated using the standardized formula: (Weight in kilograms)/(Height in meters)/2 (Carlson et al., 2016). Respondents gave their weight and height, so we can group their BMI according to the following categories (Fu et al., 2018):

BMI classification: (a) underweight (18.5 kg/m^2); (b) normal (18.5–23.9 kg/m^2); (c) overweight (24.0–27.9 kg/m^2); (d) obese (28.0 kg/m^2).

The Center for Epidemiologic Studies Depression Scale (CES-D), created by Radloff (1977) was used to measure depression. Twenty items make up the CES-D questionnaire, each of which is graded on a 4-point Likert scale. Higher scores showed that the depression was more severe.

The frequency of respondents using mHealth apps was measured based on previous studies (Rasche et al., 2018). In this study, fitness and nutrition apps usage were indicated as follows:

The frequency with which mHealth apps are used: (a) Every day; (b) every couple of days; (c) weekly; (d) monthly; (e) never.

3. Results

3.1. ANOVA, regression, and SEM

Before Taguchi analysis, we applied ANOVA and regression as the main common statistical approaches. ANOVA is used to evaluate the total sum of squares, the sum of squares due to BMI, and the sum of squares due to depression level. Estimating significant factors on BMI and depression involves regression modeling. From the outputs of Table 2, it can be seen that there are some significant differences between the outputs of using nutrition and fitness apps.

Tables 3, 4 show the regression outputs for BMI and depression. Both tables show that using nutrition apps among women does not significantly impact BMI and depression. Moreover, using fitness apps helped Malaysian women reduce their BMI and depression. However, if we use SEM and include both fitness and nutrition apps in a single model to estimate depression and BMI, we can see different results. Figure 1 shows there is a significant correlation between the usage of fitness apps and nutrition apps, and both fitness and nutrition apps have significant effects on both depression and BMI.

3.2. Taguchi method analysis

Based on the Taguchi method, data analysis was performed on the primary dataset extracted for the Taguchi experiment. In this study, all the variables were taken into consideration for the Taguchi experimental analysis. However, three indicators, including physical activity and average working hours, were eliminated from the analysis for the lifestyle variable. According to the previous studies, the average screen time used and sleeping hours were retained in the analysis (Khajeheian et al., 2018). For the demographic variable, we calculated the average level of the respondents' backgrounds and distributed them into five categories, as indicated in Table 2. Each variable has five levels. Hence, the L_{25} (5^5) Taguchi experimental design was utilized. Table 5 contains the Taguchi coding structure for data analysis using the MINITAB software.

Figures 1, 2 and Tables 6, 7 express the Taguchi method outputs from MINITAB software for BMI and depression levels, respectively.

Figures 1, 2 show the Taguchi output for women's obesity and women's depression, respectively. Figure 1 illustrates that the highest BMI occurred in women who never use the fitness apps or use them once a month, with a 4-h average screen time per day and more than 9 h of sleep. However, according to Figure 2, the highest depression level among women can be observed for respondents who never use fitness apps, spend more than 4 h per day on screen, and have an average sleeping time of fewer than 6 h or more than 9 h per day.

4. Discussion

4.1. Discussion based on Taguchi method

In this study, we designed the Taguchi method based on demographic details, screen time, sleep amount, F-App, and N-App. We chose "Larger is Better" (See Table 1) in Taguchi's design. We want to know which combination of research variable levels causes higher BMI and depression levels among women. Figures 1, 2 illustrate the MINITAB software output according to the Taguchi method design for BMI and depression levels, respectively. Based on Figures 1, 2, the demographic and N-App are not significant for both outputs. The patterns shown in these two variables are near the dotted lines. Therefore, it can be interpreted that the demographic and N-App do not significantly affect BMI and depression levels. In other words, for women who have higher BMI and depression levels, their demographics, and N-App do not significantly impact their BMI and depression levels.

Figures 1, 2 show that the F-App has a negative slope, and the diagram has high variation. Therefore, F-App is effective in reducing BMI and depression levels. For better understanding, we grouped this variable (F-App) into groups: group 1 = never; group 2 = monthly; group 3 = weekly; group 4 = 2–3 times per week, and group 5 = more than three times per week. Figure 1 shows that the outputs for groups 1, 2, and 3 are the same in reducing depression. However, there is a significant decrease between groups 3 and 4, and a slight decrease from group 4 to

TABLE 2 ANOVA outputs.

			Sum of squares	df	Mean square	*F*	Significant
N-App	Depression level	Between groups	137.037	4	34.259	1.939	0.102
		Within groups	15423.820	873	17.668		
		Total	15560.858	877			
	BMI	Between groups	169.461	4	42.365	1.737	0.140
		Within groups	21289.792	873	24.387		
		Total	21459.254	877			
F-App	Depression level		Sum of squares	df	Mean square	*F*	Significant
		Between groups	360.638	4	90.159	5.178	0.000
		Within groups	15200.220	873	17.411		
		Total	15560.858	877			
	BMI	Between groups	2612.257	4	653.064	40.250	0.000
		Within groups	18846.997	873	21.589		
		Total	21459.254	877			

group 5. As a result, if the F-App matches these groups; never, monthly, and weekly, the outputs regarding reducing depression levels are the same, and second, there is no significant effect on reducing depression levels. Females who were using the F-App 2–3 times per week showed better progress in reducing depression. Figure 2 shows similar outputs for groups 1 and 2. However, there is a significant difference between group 2 and group 3 and also between group 3 and group 4. Note that there are no significant differences between groups 4 and 5. As a result, females who use F-App weekly, particularly 2–3 times per week, have a higher likelihood of reducing their BMI over time.

The third diagram (from left) of Figures 1, 2 shows the impact of screen time on depression level and BMI, respectively. The diagram has a positive slope and high variation. We grouped this variable into groups: group 1 = less than 1 h; group 2 = 1–2 h; group 3 = 2–3 h; group 4 = 3–4 h; and group 5 = more than 4 h per day. The diagrams of screen time from Figures 1, 2 show that BMI and depression levels increase drastically with increasing screen time. However, respondents who spend less than 2 h (group 1 and group 2) on screen time have low depression symptoms. If they spend more time on-screen, the findings indicate that they have a higher depression level. This result is supported by Kim et al. (2017)'s previous studies in which smartphone addiction positively correlates with depression. The findings of this study suggest that more screen time increases respondents' levels of BMI. Respondents who sleep more per night would have a higher BMI as well. This output is supported by previous studies (Maddahi et al., 2020; Salarzadeh Jenatabadi et al., 2020).

The last diagrams of Figures 1, 2 show how sleep amount matters in measuring women's BMI and depression, respectively. When the women have an average of 7–8 h or 8–9 h of sleep, it will not change their BMI levels. To have a normal BMI range, sleeping between 7 and 9 h is not effective at all. The Taguchi output for depression levels is different compared to the BMI level output. Respondents who use fitness apps every 2–3 days or daily seem to have lower depression levels. The use of fitness apps regularly decreased the level of depression among women. The Taguchi output in Figure 2 for sleeping hours shows exciting patterns. Respondents who sleep less than 6 h or more than 9 h every day have the highest level of depression (i.e., likely depression). The impact of these two levels of sleep on depression levels is the same. Sleep quality might be one of the main possibilities for respondents to suffer from depression. Besides, previous studies also claim that depression is linked to women's lifestyle choices, e.g., sleep quality (Yang et al., 2019). Berk et al. (2013) abridged that claim by stressing that poor lifestyle choices cause depression.

TABLE 3 Regression outputs (dependent variable: BMI).

Variables	Coefficients	Standard error	*t*-Stat	*p*-Value
(Constant)	28.386	1.195	23.761	0.000
F-App	−1.624	0.142	−11.465	0.000
N-App	−0.004	0.146	−0.030	0.976
Screen Time	0.370	0.168	2.200	0.028
Sleeping	0.398	0.153	2.604	0.009
Demographic	0.104	0.215	0.485	0.627

TABLE 4 Regression outputs (dependent variable: Depression).

Variables	Coefficients	Standard error	*t*-Stat	*p*-value
(Constant)	14.411	1.077	13.386	0.000
F-App	−0.550	0.128	−4.307	0.000
N-App	−0.008	0.131	−0.611	0.541
Screen Time	0.313	0.152	2.059	0.039
Sleeping	0.404	0.138	2.927	0.034
Demographic	0.213	0.194	1.101	0.271

4.2. Discussion based on SEM

We find a strong and consistent link between obesity and depression in this sample of Malaysian women. Additionally, we noticed a stepwise

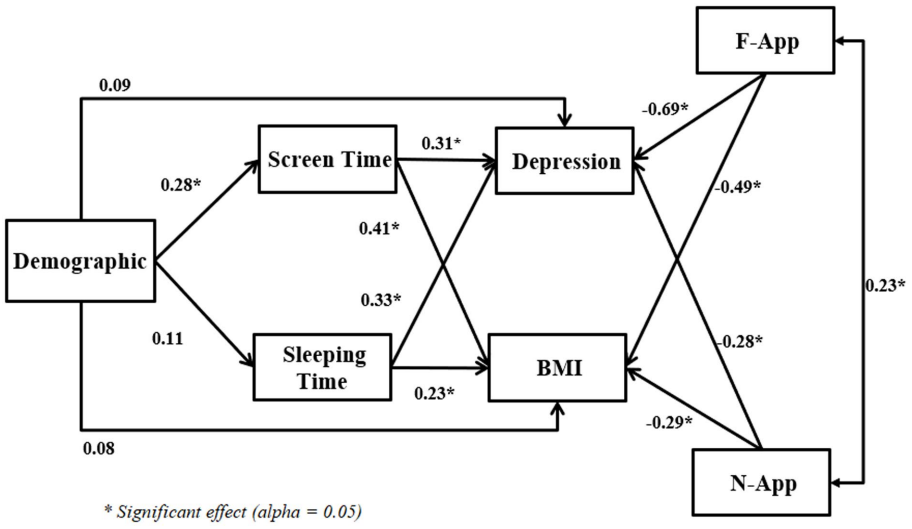


FIGURE 1
SEM outputs (dependent variables: Depression and BMI).

10.3389/fpsyg.2023.1060963

increase in both directions: increasing body mass index was strongly associated with higher risk of depressive disorder and increasing severity of depressive symptoms was strongly associated with higher risk of obesity. We discover that the link between depression and obesity was not just present in cases of more severe obesity. Accounting for potential confounders like demographic, screen time, sleep time, and use of mobile applications for nutrition and fitness had a negligible impact on this association. The correlation could not be explained by the specific effects of obesity on somatic depression symptoms. The demographic groups where the obesity-depression association is strongest may have been the focus of our survey. Despite the fact that we did not find differences in the relationship between obesity and depression across demographic groups, earlier studies have suggested that this connection may differ based on ethnicity, education, age, and monthly income (Gavin et al., 2010; Assari et al., 2014; Lincoln, 2019).

We investigate two likely mediators of the link between obesity and depression: screen time and sleep time. In our study, after controlling for the link between sleep duration and obesity, depression was found to be independently associated with less sleep. Of course, both hypotheses—that obesity causes depression or that depression causes obesity—are consistent with this observation. In the first scenario, depression might cause less sleep, which would then result in weight gain. In the latter case, fewer hours of sleep brought on by obesity may be a factor in low mood. In either case, though, our data imply that the amount of sleep that a person gets may play a role in mediating this relationship. After taking into account the connection between screen time and BMI, screen time was also linked to depression. This result supports the idea that obesity causes depression by stigmatizing the condition and reducing screen time.

TABLE 5 Taguchi method coding.

Level	Coding	Level	Coding
Demographic (Average)		N-App	
Very Low	Code "1"	Never	Code "1"
Low	Code "2"	Monthly	Code "2"
Moderate	Code "3"	Weekly	Code "3"
High	Code "4"	Every 2–3 days	Code "4"
Very High	Code "5"	Daily	Code "5"
F-App		Screen time	
Never	Code "1"	Less than 1 h	Code "1"
Monthly	Code "2"	1–2 h	Code "2"
Weekly	Code "3"	2–3 h	Code "3"
Every 2–3 days	Code "4"	3–4 h	Code "4"
Daily	Code "5"	More than 4 h	Code "5"
Sleep Amount			
Less than 6 h	Code "1"		
6–7 h	Code "2"		
7–8 h	Code "3"		
8–9 h	Code "4"		
More than 9 h	Code "5"		

4.3. Comparison of Taguchi outputs with ANOVA, regression analysis, and SEM process

Can fitness and nutrition apps help women control their BMI and depression during the COVID-19 breakdown? From Tables 3–5 and Figures 1, 2, we can deduce that using nutrition apps does not help women to control BMI and depression levels during the COVID-19 outbreak. However, using fitness apps can increase the frequency of usage and may lower BMI and depression levels among women. From Tables 4, 5, the ANOVA and regression methods show us which variables effectively identify BMI and depression levels. It is worth noting that the Taguchi method gave information that ANOVA and regression could not present. Significantly, the Taguchi outputs of Figures 1, 2 show which combinations of research variables based on their level effectively identify women with higher BMI and depression.

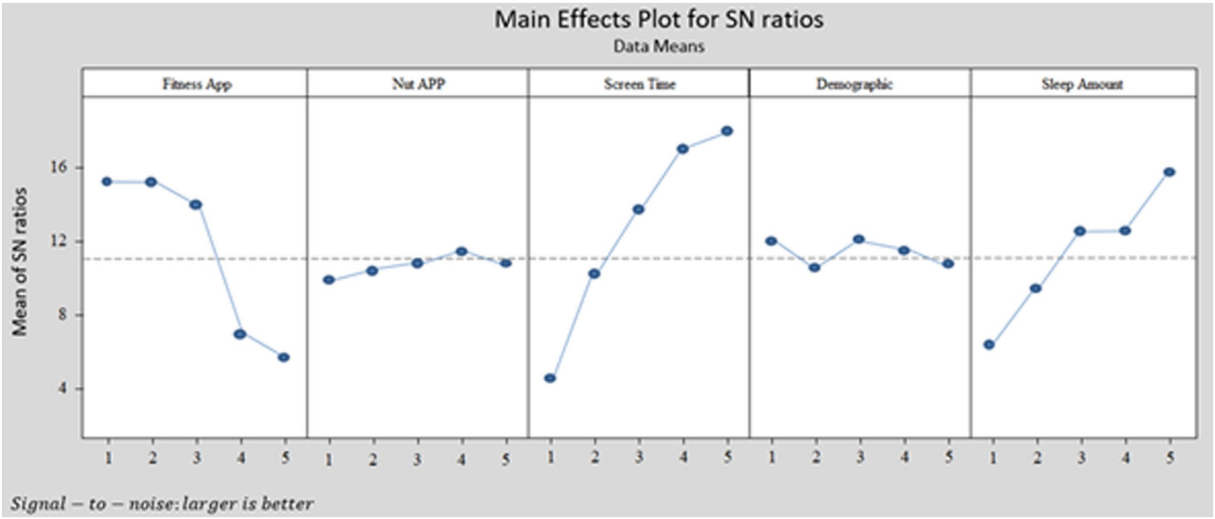


FIGURE 2
Taguchi output for BMI pattern with MINITAB software.

TABLE 6 Taguchi output for BMI.

A	B	C	D	E	BMI1	BMI2	BMI3	BMI4	BMI5	BMI6	SNRA	MEAN
1	1	1	1	1	23.2	18.4	30.5	24.2	17.2	22.2	26.653	22.637
1	2	2	2	2	31.2	32.1	20.0	20.8	30.8	23.4	27.915	26.369
1	3	3	3	3	25.9	19.8	26.4	24.0	31.3	29.3	28.045	26.103
1	4	4	4	4	24.5	27.1	27.0	30.6	30.1	17.1	27.788	26.086
1	5	5	5	5	23.6	26.2	24.4	21.6	29.7	28.4	28.019	25.624
2	1	2	3	4	27.4	24.6	20.0	27.1	24.0	19.9	27.319	23.814
2	2	3	4	5	26.5	20.0	21.3	20.3	17.0	30.8	26.624	22.667
2	3	4	5	1	17.8	22.0	32.8	30.1	20.1	26.8	27.321	24.946
2	4	5	1	2	20.2	28.6	19.7	19.1	17.8	32.0	26.608	22.913
2	5	1	2	3	18.3	22.0	29.1	29.4	27.6	25.7	27.692	25.361
3	1	3	5	2	22.2	23.8	22.6	20.1	28.0	17.0	26.654	22.275
3	2	4	1	3	30.7	29.3	24.6	31.4	23.4	32.8	28.939	28.686
3	3	5	2	4	30.6	19.3	26.7	32.6	19.3	28.3	27.775	26.133
3	4	1	3	5	31.4	24.5	24.5	21.2	27.7	31.7	28.309	26.856
3	5	2	4	1	24.4	31.6	32.7	17.8	19.9	27.9	27.528	25.705
4	1	4	2	5	25.4	29.5	26.1	25.0	23.2	23.6	28.038	25.457
4	2	5	3	1	30.9	31.6	31.8	27.9	29.5	23.6	29.17	29.227
4	3	1	4	2	25.9	31.4	23.8	17.1	30.6	27.5	27.755	26.063
4	4	2	5	3	25.5	18.5	17.8	24.8	20.6	32.6	26.8	23.299
4	5	3	1	4	26.8	26.0	19.8	18.1	27.8	31.7	27.462	25.046
5	1	5	4	3	27.1	32.8	26.1	28.0	24.5	24.7	28.571	27.201
5	2	1	5	4	18.4	26.4	21.8	22.1	26.0	29.5	27.302	24.032
5	3	2	1	5	23.6	24.5	21.5	31.8	29.6	33.0	28.405	27.335
5	4	3	2	1	21.4	20.5	17.2	25.9	21.0	24.9	26.537	21.823
5	5	4	3	2	27.5	27.5	21.0	18.6	28.9	31.2	27.764	25.774

However, Figure 1 shows that, with considering both F-APP and N-App for estimating both depression and BMI in a same model, we will have different outputs compare Taguchi method, regression, and ANOVA.

4.4. Contributions of the study

The novel contribution of this paper to theory are multifield. First, and the main contribution, the study adds its valuable contribution the effectiveness of different ways of estimating correlation among research variables with depression and BMI. With considering correlation or effectiveness of F-App and N-App on BMI and depression, ANOVA, regression, and Taguchi method were useful. However, these methods are not able to evaluate both F-App and N-App with other research variables on both BMI and depression in a single model. As you can see in data analysis part for ANOVA analysis we have to analyze the effectiveness of F-App and N-App separately and for every of dependent variables need to present different analysis (see Table 2). This weakness of the analysis can be seen in regression (see Tables 3, 4) and Taguchi (see Figures 2, 3). Therefore, for analyzing one dependent variable analysis, these methods might be applicable. However, estimating two or more than dependent variables in a single model or analysis is applicable by involving SEM.

Second, this study adds valuables results to the mHealth literature. The study revealed demographic does not have significant impact on both depression and BMI based on regression, Taguchi method, and even SEM. However, with SEM technique, we realized demographic has an indirect significant effect on depression through screen time. The sudden and unexpected changes way of analysis, we are able to find hidden significant relationships whish have not been considered before.

Third, another novel contribution of the study is that we found out there is a strong correlation between the usage of F-App and N-App based on SEM technique which has nobody have done before.

4.5. Study restrictions

4.5.1. This research also has a few limitations

Self-reported: The questionnaires were self-reported by respondents, especially the weight and height measurements, to determine their BMI. The CES-D and frequency of using mHealth were also self-reported. The validity of this method has been verified in previous studies (Sanchez et al., 2010; Banting et al., 2014). So, we considered the research data acceptable to be analyzed.

The CES-D questionnaire: It should be noted that in this study, we only analyzed the depression symptoms of the respondents. As

10.3389/fpsyg.2023.1060963

TABLE 7 Taguchi output for depression.

A	B	C	D	E	Dep1	Dep2	Dep3	Dep4	Dep5	Dep6	SNRA	MEAN
1	1	1	1	1	26	23	29	6	25	22	22.218	21.833
1	2	2	2	2	6	10	21	34	7	33	19.841	18.5
1	3	3	3	3	35	25	17	23	34	35	28.028	28.167
1	4	4	4	4	18	6	24	34	24	5	19.019	18.5
1	5	5	5	5	21	18	21	29	20	10	24.496	19.833
2	1	2	3	4	12	11	7	8	12	27	20.032	12.833
2	2	3	4	5	10	8	15	18	32	15	21.918	16.333
2	3	4	5	1	11	16	24	16	24	25	24.529	19.333
2	4	5	1	2	21	34	31	15	31	30	27.46	27
2	5	1	2	3	28	12	6	27	21	17	21.437	18.5
3	1	3	5	2	35	15	10	15	24	19	23.941	19.667
3	2	4	1	3	22	19	23	17	34	24	26.714	23.167
3	3	5	2	4	30	15	11	32	25	14	24.457	21.167
3	4	1	3	5	26	13	31	29	25	22	26.546	24.333
3	5	2	4	1	33	33	27	20	35	27	28.807	29.167
4	1	4	2	5	18	12	6	27	6	34	19.468	17.167
4	2	5	3	1	31	16	12	11	25	26	24.12	20.167
4	3	1	4	2	14	27	11	5	10	8	18.731	12.5
4	4	2	5	3	34	27	27	24	15	26	27.263	25.5
4	5	3	1	4	16	25	23	10	13	13	23.123	16.667
5	1	5	4	3	12	11	12	12	11	6	19.643	10.667
5	2	1	5	4	15	30	15	21	11	30	24.428	20.333
5	3	2	1	5	32	23	21	7	33	23	23.256	23.167
5	4	3	2	1	9	28	33	25	30	8	22.613	22.167
5	5	4	3	2	32	8	9	26	31	30	22.652	22.667

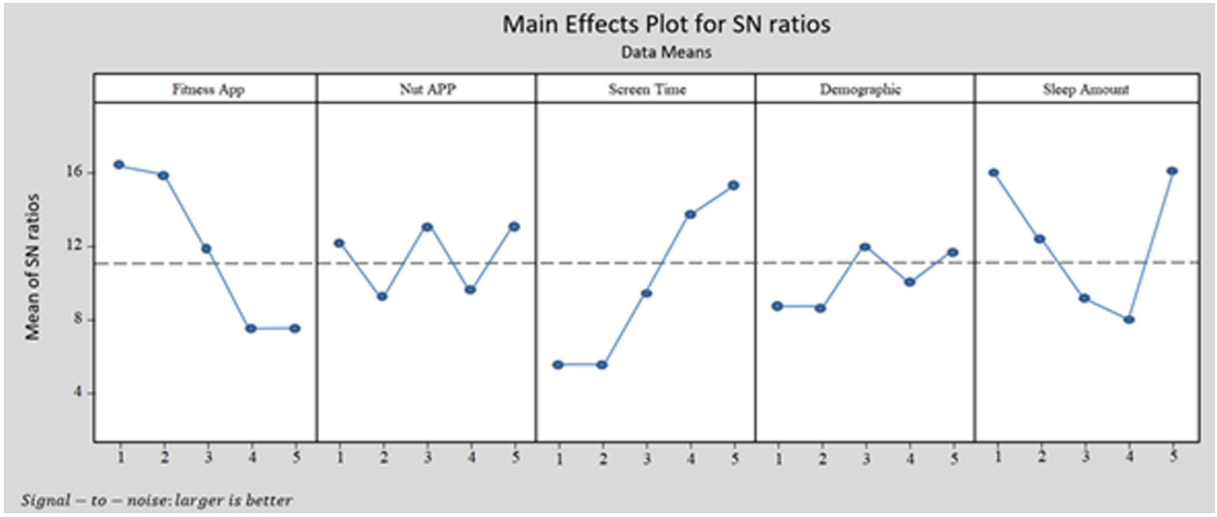


FIGURE 3
Taguchi output for depression pattern with MINITAB software.

we calculated the score, we grouped the respondents according to their depression levels, as described in the method section. The questionnaire is not a substitute for clinical diagnosis.

Additional treatments: Some of the respondents might also receive other additional treatments such as personal trainers or psychologists, which could probably contribute to the BMI and depression levels.

Future research should look into how these indicators affect female obesity and depression.

5. Conclusion

We conclude that throughout the pandemic in Malaysia, using fitness apps consistently was more effective than using nutrition apps to establish a better quality of life among women. As we live in the era of ICT, the availability of thousands of mHealth apps would help people organize their daily lives. For example, fitness apps help promote suitable physical activity, diet control, weight management, stress relief, and sleep monitoring (Higgins, 2016). Apart from that, we believe that poor sleep is also correlated with women's depression and obesity development, which has been proven in previous studies (Saucedo et al., 2021) and extensive screen time usage. The Taguchi method, which we used in this study, gave public health researchers a way to look at the levels of obesity and depression in women. The last and the main conclusion of this studies, with application of SEM, research scholar are able to find out significant different correlations among research variables compare to familiar mythologies include ANOVA, regression, and Taguchi methods.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by University of Malaya Research Ethics Committee (UMREC). The patients/participants provided their written informed consent to participate in this study.

Author contributions

NM, AA, and HJ contributed to conception and design of the study. NM, AA, and HJ organized the database and performed the statistical analysis. NS, NA, and SA wrote the first draft of the manuscript. NM, AA, NA, SA, NS, and HJ wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This research is funded by Universiti Malaya Research Grant (Grant No. GPF083B-2020) and Prince Sultan University.

Acknowledgments

The authors would like to thank all participants for their cooperation during the research.

Conflict of interest

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.

Publisher's note

References

Archana, P. S., Das, S., Philip, S., Philip, R. R., Joseph, J., Punnoose, V. P., et al. (2017). Prevalence of depression among middle aged women in the rural area of Kerala. *Asian J. Psychiatr.* 29, 154–159. doi: 10.1016/j.ajp.2017.05.016

Assari, S., Das, S., Philip, S., Philip, R. R., Joseph, J., Punnoose, V. P., et al. (2014). Association between obesity and depression among American blacks: role of ethnicity and gender. *J. Racial Ethn. Health Disparities* 1, 36–44. doi: 10.1007/s40615-014-0007-5

Banting, L. K., Gibson-Helm, M., Polman, R., Teede, H. J., and Stepto, N. K. (2014). Physical activity and mental health in women with polycystic ovary syndrome. *BMC Womens Health* 14:51. doi: 10.1186/1472-6874-14-51

Bardus, M., van Beurden, S. B., Smith, J. R., and Abraham, C. (2016). A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. *Int. J. Behav. Nutr. Physical Act.* 13, 1–9. doi: 10.1186/s12966-016-0359-9

Berk, M., Sarris, J., Coulson, C. E., and Jacka, F. N. (2013). Lifestyle management of unipolar depression. *Acta Psychiatr. Scand.* 127, 38–54. doi: 10.1111/acps.12124

Boulos, M. N. K., Brewer, A. C., Karimkhani, C., Buller, D. B., and Dellavalle, R. P. (2014). Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online J. Public Health Inform.* 5:229. doi: 10.5210/ojphi.v5i3.4814

Bt Wan Mohamed Radzi, C. W. J., Salarzadeh Jenatabadi, H., and Samsudin, N. (2020). mHealth apps assessment among postpartum women with obesity and depression. *Healthcare* 8:72. doi: 10.3390/healthcare8020072

Bt Wan Mohamed Radzi, C. W. J., Salarzadeh Jenatabadi, H., and Samsudin, N. (2021). Postpartum depression symptoms in survey-based research: a structural equation analysis. *BMC Public Health* 21, 1–12. doi: 10.1186/s12889-020-09999-2

Carlson, J. A., Remigio-Baker, R. A., Anderson, C. A. M., Adams, M. A., Norman, G. J., Kerr, J., et al. (2016). Walking mediates associations between neighborhood activity supportiveness and BMI in the Women's Health Initiative San Diego cohort. *Health Place* 38, 48–53. doi: 10.1016/j.healthplace.2016.01.001

Cho, J., Lee, H. E., Kim, S. J., and Park, D. (2015). Effects of body image on college students' attitudes toward diet/fitness apps on smartphones. *Cyberpsychol. Behav. Soc. Netw.* 18, 41–45. doi: 10.1089/cyber.2014.0383

Chuan, C. L., and Penyelidikan, J. (2006). Sample size estimation using Krejcie and Morgan and Cohen statistical power analysis: a comparison. *Jurnal Penyelidikan IPBL* 7, 78–86.

Correa, M. S. M., de Oliveira Feliciano, K. V., Pedrosa, E. N., and de Souza, A. I. (2017). Postpartum follow-up of women's health. *Cad. Saude Publica* 33:e00136215. doi: 10.1590/0102-311X00136215

Elamoshy, R., Bird, Y., Thorpe, L., and Moraros, J. (2018). Risk of depression and suicidality among diabetic patients: a systematic review and meta-analysis. *J. Clin. Med.* 7:445. doi: 10.3390/jcm7110445

El-Gilany, A. H., Elkhawaga, G. O., and Sarraf, B. B. (2018). Depression and its associated factors among elderly: a community-based study in Egypt. *Arch. Gerontol. Geriatr.* 77, 103–107. doi: 10.1016/j.archger.2018.04.011

Fu, L.-Y., Wang, X. X., Wu, X., Li, B., Huang, L. L., Li, B. B., et al. (2018). Association between obesity and sickness in the past two weeks among middle-aged and elderly women: a cross-sectional study in southern China. *PLoS One* 13:e0203034. doi: 10.1371/journal.pone.0203034

Gavin, A. R., Simon, G. E., and Ludman, E. J. (2010). The association between obesity, depression, and educational attainment in women: the mediating role of body

image dissatisfaction. *J. Psychosom. Res.* 69, 573–581. doi: 10.1016/j.jpsychores.2010.05.001

Heron, K. E., and Smyth, J. M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *Br. J. Health Psychol.* 15, 1–39. doi: 10.1348/135910709X466063

Higgins, J. P. (2016). Smartphone applications for patients' health and fitness. *Am. J. Med.* 129, 11–19. doi: 10.1016/j.amjmed.2015.05.038

Jia, H., Zack, M. M., Gottesman, I. I., and Thompson, W. W. (2018). Associations of smoking, physical inactivity, heavy drinking, and obesity with quality-adjusted life expectancy among US adults with depression. *Value Health* 21, 364–371. doi: 10.1016/j.jval.2017.08.002

Jin, Q. G., Mori, E., and Sakajo, A. (2016). Risk factors, cross-cultural stressors and postpartum depression among immigrant Chinese women in Japan. *Int. J. Nurs. Pract.* 22, 38–47. doi: 10.1111/ijn.12438

Khajeheian, D., Colabi, A., Ahmad Kharman Shah, N., Bt Wan Mohamed Radzi, C., and Jenatabadi, H. (2018). Effect of social media on child obesity: application of structural equation modeling with the Taguchi method. *Int. J. Environ. Res. Public Health* 15, 1–22. doi: 10.3390/ijerph15071343

Kim, E., Cho, I., and Kim, E. J. (2017). Structural equation model of smartphone addiction based on adult attachment theory: mediating effects of loneliness and depression. *Asian Nurs. Res.* 11, 92–97. doi: 10.1016/j.anr.2017.05.002

Krejcie, R. V., and Morgan, D. W. (1970). Determining sample size for research activities. *Educ. Psychol. Meas.* 30, 607–610. doi: 10.1177/001316447003000308

Lincoln, K. D. (2019). Social stress, obesity, and depression among women: clarifying the role of physical activity. *Ethn. Health* 24, 662–678. doi: 10.1080/13557858.2017.1346190

Maddahi, N. S., Yarizadeh, H., Setayesh, L., Nasir, Y., Alizadeh, S., and Mirzaei, K. (2020). Association between dietary energy density with mental health and sleep quality in women with overweight/obesity. *BMC. Res. Notes* 13, 1–6. doi: 10.1186/s13104-020-05025-1

Manigandan, S., Atabani, A. E., Ponnusamy, V. K., Pugazhendhi, A., Gunasekar, P., and Prakash, S. (2020). Effect of hydrogen and multiwall carbon nanotubes blends on combustion performance and emission of diesel engine using Taguchi approach. *Fuel* 276:118120. doi: 10.1016/j.fuel.2020.118120

Muharam, R., Setiawan, M. W., Ikhsan, M., Rizkinya, H. E., and Sumapraja, K. (2018). Depression and its link to other symptoms in menopausal transition. *Middle East Fertil. Soc. J.* 23, 27–30. doi: 10.1016/j.mefs.2017.08.003

Nakayama, K., Yamaguchi, K., Maruyama, S., and Morimoto, K. (2001). The relationship of lifestyle factors, personal character, and mental health status of employees of a major Japanese electrical manufacturer. *Environ. Health Prev. Med.* 5, 144–149. doi: 10.1007/BF02918290

Qudah, B., and Luetsch, K. (2019). The influence of mobile health applications on patient-healthcare provider relationships: a systematic, narrative review. *Patient Educ. Couns.* 102, 1080–1089. doi: 10.1016/j.pec.2019.01.021

Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Appl. Psychol. Meas.* 1, 385–401. doi: 10.1177/014662167700100306

Rasche, P., Wille, M., Bröhl, C., Theis, S., Schäfer, K., Knobe, M., et al. (2018). Prevalence of health app use among older adults in Germany: national survey. *JMIR Mhealth Uhealth* 6:e26. doi: 10.2196/mhealth.8619

Rathbone, A. L., and Prescott, J. (2017). The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. *J. Med. Internet Res.* 19:e7740. doi: 10.2196/jmir.7740

Rippe, J. M. (2021). "The epidemiology of adult obesity" in *Obesity Prevention and Treatment*, eds. James M. Rippe and John P. Foreyt (Boca Raton: CRC Press), 21–27.

Salarzadeh Jenatabadi, H., Babashamsi, P., and Yusoff, N. I. M. (2016). The combination of a fuzzy analytical hierarchy process and the Taguchi method to evaluate the Malaysian users' willingness to pay for public transportation. *Symmetry* 8:90. doi: 10.3390/sym8090090

Salarzadeh Jenatabadi, H., Bt Wan Mohamed Radzi, C. W. J., and Samsudin, N. (2020). Associations of body mass index with demographics, lifestyle, food intake, and mental health among postpartum women: a structural equation approach. *Int. J. Environ. Res. Public Health* 17:5201. doi: 10.3390/ijerph17145201

Sanchez, S. E., Williams, M. A., Pacora, P. N., Ananth, C. V., Qiu, C., Aurora, S. K., et al. (2010). Risk of placental abruption in relation to migraines and headaches. *BMC Womens Health* 10:30. doi: 10.1186/1472-6874-10-30

Saucedo, M., Esteves-Pereira, A. P., Pencolé, L., Rigouzzo, A., Proust, A., Bouvier-Colle, M. H., et al. (2021). Understanding maternal mortality in women with obesity and the role of care they receive: a national case-control study. *Int. J. Obes.* 45, 258–265. doi: 10.1038/s41366-020-00691-4

Taylor, G., Bylund, C. L., Kastrinos, A., Alpert, J. M., Puig, A., Krajewski, J. M. T., et al. (2022). Practicing mindfulness through mHealth applications: emerging adults health-enhancing and inhibiting experiences. *Int. J. Environ. Res. Public Health* 19:2619. doi: 10.3390/ijerph19052619

World Health Organization (2011). mHealth New Horizons for Health Through Mobile Technologies. Available at: https://www.who.int/about/

World Health Organization (2017). Depression: Let's Talk. Available at: https://www.who.int/mental_health/management/depression/en/ (Accessed December 20, 2018).

Yang, C.-L., Schnepp, J., and Tucker, R. M. (2019). Increased hunger, food cravings, food reward, and portion size selection after sleep curtailment in women without obesity. *Nutrients* 11:663. doi: 10.3390/nu11030663

Yuan, S., Ma, W., Kanthawala, S., and Peng, W. (2015). Keep using my health apps: discover users' perception of health and fitness apps with the UTAUT2 model. *Telemed. e-Health* 21, 735–741. doi: 10.1089/tmj.2014.0148

Zeng, N., Pope, Z., Lee, J., and Gao, Z. (2018). Virtual reality exercise for anxiety and depression: a preliminary review of current research in an emerging field. *J. Clin. Med.* 7:42. doi: 10.3390/jcm7030042

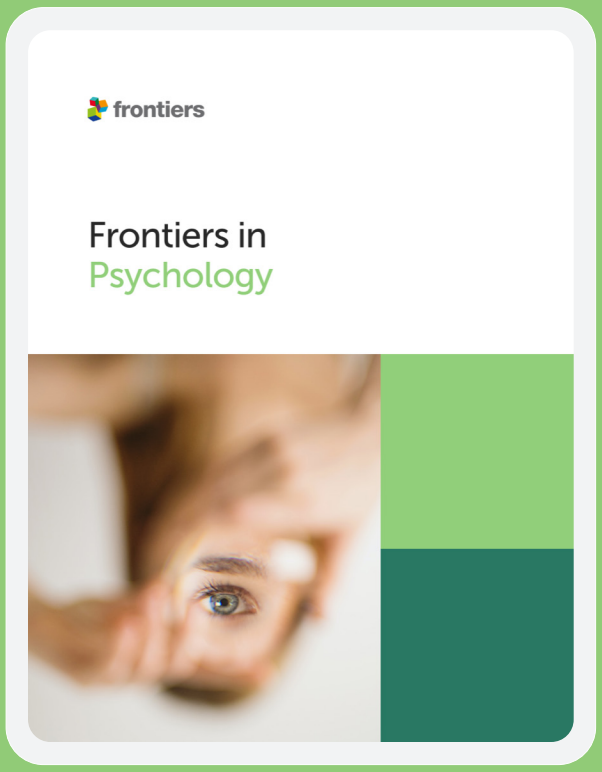
Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact