NEW CHALLENGES IN GLOBALIZED SOCIETIES: CROSS-CULTURAL STUDIES AND TEST ADAPTATION

EDITED BY: Fco. Pablo Holgado-Tello, Salvador Chacón-Moscoso, Susana Sanduvete-Chaves and José Antonio Lozano Lozano

PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-293-4
DOI 10.3389/978-2-88976-293-4

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

NEW CHALLENGES IN GLOBALIZED SOCIETIES: CROSS-CULTURAL STUDIES AND TEST ADAPTATION

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

Citation: Holgado-Tello, F. P., Chacón-Moscoso, S., Sanduvete-Chaves, S., Lozano, J. A. L., eds. (2022). New Challenges in Globalized Societies: Cross-cultural Studies and Test Adaptation. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88976-293-4

Table of Contents

EDITORIAL
published: 10 May 2022
doi: 10.3389/fpsyg.2022.900535

Editorial: New Challenges in Globalized Societies: Cross-Cultural Studies and Test Adaptation

Fco. Pablo Holgado-Tello[1]*, *Salvador Chacón-Moscoso*[2], *Susana Sanduvete-Chaves*[3] and *José Antonio Lozano-Lozano*[2]

[1] National University of Distance Education (UNED), Madrid, Spain, [2] Universidad Autónoma de Chile, Santiago, Chile, [3] Universidad de Sevilla, Seville, Spain

Keywords: psychometrics, cross-cultural studies, tests adaptation, validity, analytical strategies

Editorial on the Research Topic

New Challenges in Globalized Societies: Cross-Cultural Studies and Test Adaptation

Tests have become one of the basic tools to carry large-scale psychological and educational evaluations. In globalized societies, having common elements of evaluation allows the comparison and measurement of participants from diverse cultures and backgrounds, with assurances that the instruments are free of bias, and consequently, that the assessments are objective. However, obtaining such instruments involves overcoming serious drawbacks that arise in their development. Among other aspects, these disadvantages are mainly due to the cultural origin of the populations to be measured with the same instrument.

In practice, this question involves considering all aspects, from the connotative and denotative meaning of each word to the very interpretation of the test scores. That is why it is not enough to translate literally the items of an already validated test in one population and apply them directly in a different language and culture. Instead, in addition to guarantee certain linguistic equivalences or construct, as well as developing new standards, it is necessary to obtain empirical evidence of validity again, in the broad sense of the term, in the new population.

The relevance of this issue was revealed in 1985, when the American Educational Research Association (AERA), together with the American Psychological Association (APA) and the National Council on Measurement in Education (NCME) published the Standards for Educational and Psychological Testing. It provides a theoretical framework to assist in the process of test adaptation and, although today they have been modified and completed with the new guidelines developed by the International Test Commission (ITC), it served to warn the possible sources of error that may arise during the adaptation process (Barbero et al., 2008). The ITC regularly publishes and updates the standards for the adaptation of tests. In the latest 2017 version, we can find a checklist of the fundamental aspects to consider (Hernández et al., 2020). However, although the importance and relevance of the problem were made explicit more than 30 years ago, a study by Rios and Sireci (2014) shows that most publications that propose adaptations do not follow ITC standards.

In this Research Topic, we would like to address the importance of test adaptations, which could imply to investigate the psychometric properties on different samples without the need of items translation, or studies that translate the test to a different language with a different sociocultural background. The relevance of this problematic is so high that authors from four different continents (Asia, Europa, America and Africa) have submitted their manuscripts contributing to reach an interesting Research Topic formed by different applications and techniques implied in test construction and adaptation. Within the most commonly used techniques in the general research about test construction and adaptation, it is significant the Confirmatory Factor Analysis and those derived from it, such as multigroup analysis, exploratory structural equation modeling (ESEM), or bifactor models.

OPEN ACCESS

Edited and reviewed by:
Laura Galiana,
University of Valencia, Spain

Correspondence:
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 March 2022*
Accepted: *04 April 2022*
Published: *10 May 2022*

Citation:
Holgado-Tello FP, Chacón-Moscoso S, Sanduvete-Chaves S and Lozano-Lozano JA (2022) Editorial: New Challenges in Globalized Societies: Cross-Cultural Studies and Test Adaptation. Front. Psychol. 13:900535. doi: 10.3389/fpsyg.2022.900535

Within the adaptation process, and in line with the fast changes that characterize societies nowadays, new challenges are appearing for psychometricians. In this sense, Psychometrics is undergoing its own developments, and as Ruhe and Zumbo (2009) point out, they affect and revitalize the framework of their classical test theories. Measurement in general, and Psychometrics in particular, are significantly affected by the emergence of modern technologies and methodologies such as information technology and the Internet which, among other aspects, allow for example the presentation of multimedia items, tele-assessment, or experience sampling methodology (Myin-Germeys and Kuppens, 2022). These new contexts and assessment formats also imply, to a greater or lesser degree, certain adaptation process that guarantees that the interpretations of the scores are free of artifactual bias.

The concrete contributions to this Research Topic present, in the context of on-line evaluation, empirical validity evidence of a short version of an instrument designed to assess the precompetitive psychological state of the athletes (Diaz-Tendero et al.). Additionally, differences between amateur and professional athletes were found based on on-line assessment. As mentioned before, this kind of assessment is highly facilitated by the irruption of methodologies as Internet and smart phones, which are going to overcome some of the traditional criticisms about the assessment using tests. Assessing the athletes just before the competition increases notably the ecological validity.

Precisely in the assessment process, one of the challenges that psychometricians must face is how to measure and obtain data in daily life in the real world in real time. In this sense, Fuster-RuizdeApodaca et al., step by step, according to the guidelines for test construction, have developed an useful and relevant measurement instrument to be applied in routine specialist clinical care to identify, in real time and in the real world, people living with HIV that could be suffering health-related issues referred to stigma, emotional distress, sexuality, social support, material deprivation, sleep, cognitive problems or physical symptoms.

To generalize the use of specific instruments, Zhang and Bian confirmed the measurement invariance of the two-factor structure of a test to measure emotion regulation across male and female college students in China. Additionally, Pina et al. did not found differences in math performance across gender in Chile and Spain.

To obtain validity evidence of the use of tests in specific countries, Goessmann et al. created a scale to measure intimate partner violence for the specific casuistic found in women in Iraq. Salamon et al. obtained a factor structure in a scale to measure work engagement in Hungary, consisted in a global factor and three co-existing specific factors of vigor, dedication, and absorption. Samfira and Maricutoiu assessed the psychometric properties of an inventory to measure perfectionism in Romanian teachers. Additionally, Hill et al. obtained validity evidence of an inventory to measure personality in South Africa.

Finally, in the context of test adaptation across cultures, Lacko et al. obtained validity evidence in the Czech version of a questionnaire to measure individualism/collectivism. Klocek et al. studied the psychometric properties of the Czech version of a scale to measure group cohesiveness. Sahagún-Morales et al. adapted two tools to measure potential child abuse and protective factors to the Mexican population. Furthermore, Xiong et al. adapted a survey to measure safety cognition capability to the Chinese population.

The editors greatly appreciate the contributions received from the authors in this Research Topic. We hope it will be of interest to readers, and that potential researchers will find motivation for the development of future work.

AUTHOR CONTRIBUTIONS

FH-T, SC-M, SS-C, and J-LL contributed to documenting, designing, drafting, writing the manuscript, and revised it for important theoretical and intellectual content. All authors provided final approval of the version to be published and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This work was supported by the Fondo Nacional de Desarrollo Científico y Tecnológico FONDECYT Regular, ANID, Chilean Government (1190945), the ERDF government program for Andalusia 2014-2020, Spain (US-1263096), the VI Research and Transfer Plan-VI-PPITUS, Universidad de Sevilla, Spain (VIPP PRECOMPETI 2020/1333), and the PID2020-115486GB-I00 grant funded by MCIN/AEI/10.1309/501100011033.

ACKNOWLEDGMENTS

The editors greatly appreciate the contributions received from the authors in this Research Topic.

REFERENCES

Barbero, I., Vila, E., and Holgado-Tello, F. P. (2008). Tests adaptation in cross-cultural comparative studies. *Acción Psicológica* 5, 7–16. doi: 10.5944/ap.5.2.454

Hernández, A., Hidalgo, M. D., Hambleton, R. K., and Gómez-Benito, J. (2020). International test commission guidelines for test adaptation: A criterion checklist. *Psicothema* 32, 390–398. doi: 10.7334/psicothema2019.306

Myin-Germeys, I., and Kuppens, P. (2022). *Experience Sampling Methodology: A Step-By-Step Guide To Designing, Conducting, and Analyzing ESM Studies*. Leuven: Independently published.

Rios, J. A., and Sireci, S. G. (2014). Guidelines versus practices in cross-lingual assessment: A disconcerting disconnect. *Int. J. Test*. 14, 289–312. doi: 10.1080/15305058.2014.924006

Ruhe, V., and Zumbo, B. D. (2009). *Evaluation in Distance Education and E-Learning: the Unfolding Model*. New York, NY: Guilford Press.

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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 October 2020
doi: 10.3389/fpsyg.2020.566828

Psychometric Properties of the Psychological State Test for Athletes (TEP)

Patricia Díaz-Tendero[1], M. Carmen Pérez-Llantada[2] and Andrés López de la Llave[2]*

[1] Health Psychology, National Distance Education University (UNED), Madrid, Spain, [2] Department of Behavioral Sciences and Methodology, Faculty of Psychology, National Distance Education University (UNED), Madrid, Spain

This study has two objectives: to validate an adapted online version of the Psychological State Test (TEP, in its Spanish acronym); and to assess differences in pre-competitive psychological state profiles between amateur and professional athletes in team sports. The TEP questionnaire is an instrument which is used to assess, in a quick and simple fashion, the psychological state of athletes prior to competing. Its psychometric properties were evaluated by means of an analysis of internal consistency, an Exploratory Factor Analysis and a Confirmatory Factor Analysis. The EFA's results showed a factorial structure consisting of two principal factors and reliability coefficients, both globally and at the factor level, which can be considered acceptable (global $\alpha = 0.81$; Factor 1 $\alpha = 0.85$; and Factor 2 $\alpha = 0.73$). The CFA confirmed the model proposed by the EFA so that the items were distributed around these factors, giving rise to one factor which includes variables that have a positive relationship with performance, and another with variables that negatively affect performance. Meanwhile, regarding the difference between the pre-competitive psychological state of amateurs and professionals, professional athletes presented higher levels of Motivation ($p = 0.5$ and $d = -0.23$). It is concluded that the TEP is a suitable tool for the evaluation of pre-competitive psychological states. However, in future research, this study should be complemented by analyzing the TEP's predictive validity in terms of the performance of athletes and/or teams, as well as its use as a tool available to athletes and coaches.

Keywords: anxiety, emotional state, motivation, performance, pre-competition, psychological evaluation

OPEN ACCESS

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

Reviewed by:
Antonio Granero-Gallegos, University of Almería, Spain
José Manuel Aguilar Parra, University of Almería, Spain
África Borges, University of La Laguna, Spain

Correspondence:
Patricia Díaz-Tendero
info@patriciadiaztendero.com

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

Received: *28 May 2020*
Accepted: *16 September 2020*
Published: *15 October 2020*

Citation:
Díaz-Tendero P, Pérez-Llantada MC and López de la Llave A (2020) Psychometric Properties of the Psychological State Test for Athletes (TEP). Front. Psychol. 11:566828. doi: 10.3389/fpsyg.2020.566828

INTRODUCTION

Various studies have shown how the emotional states presented by athletes before competing are determining factors in regard to their performance (Cerin, 2003; Hanin, 1997, 2000a,b). According to Lazarus (2000), the function of emotions is to facilitate adaptation to different environmental conditions and, by extension, when referring to athletes, to facilitate their performance. Moreover, this author considers that the positive or negative influence of emotions on performance will depend on the threat-challenge balance which the athlete perceives in the situation they are about to face, as well as the resources they possess in order to handle this situation. From a different perspective, but in agreement with the importance of the athletes' personal assessment, Hanin (1997) presented his model: "Individual Zone of Optimal Functioning" (IZOF), in which he proposes that emotional states prior to competition can affect athletes in different ways. Thus, each athlete would have an Individual Zone of Optimal Functioning (IZOF), defined by a set of emotions which may vary in terms of their intensity (high, moderate and low), and which are functional or not depending on the athlete.

Hanin (2000a) also proposed the IZOF-Emotion model, based on which he developed the "Emotional State Profile" (ESP-40) in which he identifies four emotional categories based on a list of 40 adjectives which helped to define the pre-competitive emotional state of athletes. These four categories of emotional states are: positive emotions which improve performance (P+), negative emotions which improve performance (N+), positive emotions which impair performance (P-), and negative emotions which impair performance (N-). Feeling happy and vigorous are considered by Hanin to be emotional states which improve performance (P+), while feeling unhappy and sad are identified as emotions which impair performance (N-).

The psychological state of the athletes as a construct shall be defined by these psychological variables together with the emotional state prior to competition, which has evidenced to have an influence on sports performance. Those psychological variables most studied in the scientific literature for their effect on sports performance are: self-confidence, motivation, stress, arousal levels, anxiety, and mood. To summarize, some data on these variables and their relationship to performance have been included. Self-confidence was used as a synonym for the term self-efficacy. Bandura (1977, 1997) described self-efficacy as the belief that one can master a situation and produce positive results. Campbell and Pritchard (1976) considered motivation as the factor that induces people to make the decision to start an activity, to put forth a certain amount of effort and to persist in it for a certain period of time. In relation to motivation, Clancy et al. (2016) conducted a review on competitive sport motivation between 1995 and 2016. In this study, the large number of studies that tested the importance of the motivation variable in influencing athletes' performance were taken into account. Concentration, understood as the maintenance of the attentional conditions for the duration of the sporting activity, promotes the athletes' processing of information (Boutcher, 2008). The arousal level variable refers to the level of physical and psychological activation of athletes while practicing sport (Malmo, 1959). The search for the optimal level of arousal, as well as the effects of high and low arousal level on performance, has been widely studied by many authors (e.g., Weinberg and Gould, 2007; Yerkes and Dodson, 1908). Stress refers to the perception of pressure by athletes when confronted with the demands of a situation, and they must adapt their responses under conditions in which failure can bring about serious consequences (McGrath, 1970). Finally, there is the anxiety variable. Nuñez and Garcia-Mas (2017) conducted a systematic review of the relationship of anxiety in sport. Although they concluded that no sufficient empirical and/or experimental evidence existed to explain the relationship between anxiety and sporting performance, the authors clarified that *"as a result of the volume of studies carried out on this subject, at the "popular" level anxiety is a problem that affects performance, given the large number of studies focused on dealing with this problem."* One reason for the possible cause of these difficulties of finding evidence of the relationship between anxiety and performance is the disparity in theoretical frameworks and the difficulty of systemising the concept of "performance." In this sense, the Catastrophe Theory (Hardy, 1990) states that when cognitive anxiety is high, the increase in the activation only improves performance to a certain point, and from this point it would produce a dramatic decline in performance ("catastrophe"), rather than a gradual decrease. Therefore, the activation may produce different effects on performance based on the individual's level of cognitive anxiety. In this sense and based on the Jones (1991) on the Directionality Theory, Ponseti et al. (2016) found in a study of swimmers that competitive anxiety has a blocking and debilitating effect on sporting performance. The authors concluded that the most important component is the cognitive factor, associated in turn with concern about performance. Authors who have studied the psychological state of athletes in the days and hours before a competition have demonstrated that, during the days, hours and even minutes leading up to competition, it is optimal that athletes present the physical and psychological state which enables them to achieve the best possible performance according to their sporting circumstances (e.g., Buceta, 1998; Cerin, 2003; Hanin, 1997, 2000a,b). Martens et al. (1990a) determined that cognitive anxiety and somatic anxiety functioned differently depending on the time interval before the sporting event. They concluded that both types of anxiety had different effects from two days before the competition up to 24 h afterwards. Meanwhile, Buceta et al. (2003) carried out research to determine the psychological state of public marathon runners between 65 and 12 h before the event. To do so, they used the CSAI-2 questionnaire (Martens et al., 1990a) which considers variables such as somatic anxiety, cognitive anxiety and self-confidence. They found that the psychological profile of these marathon runners before the race was defined by medium-low scores for somatic anxiety and cognitive anxiety, and a medium-high score for self-confidence. This suggested that, in general, public marathon runners, in the days l eading up to the event, managed to keep stress related to the race at an acceptable level and perceived that they could achieve their objectives.

Over the past few years, a number of studies have emerged that have investigated the momentary mood states of Brazilian football players in pre-competition situations (e.g., Silva, 2017; Souza, 2014). They found that football players have a common profile in the pre-competition moments determined by interest, happiness and hope. They also confirmed that the mood state prior to competition differed between players depending on their position in the game, so players in defence and forward positions presented different psychological profiles (Bueno and Souza, 2019). The authors drew on The Present Mood States List (PMSL), proposed by Engelmann (1986, 2002) to evaluate mood states resulting from wide empirical research performed in Brazil for these studies.

There are other several tools which have been used to assess the psychological state of athletes before competing. One of the most widely used is the Profile of Mood States (POMS) from McNair et al. (1971). The POMS is a test which consists of a list of multidimensional adjectives. In psychology, tests based on adjective lists are generally used to measure feelings, affects and mood states (Ávila and Giménez de la Peña, 1991). Based on the POMS, Morgan (1980a,b) Morgan and Jhonson (1977, 1978) and (Morgan and Pollock, 1977) identified the "Iceberg Profile," which refers to the layout of the scores in graphic form (which resembled the shape of an iceberg) by which the variables: Pressure, Depression, Anger, Fatigue, and Confusion are below the population average, and Vigor above it. In addition, the "Iceberg Profile" was regarded as a predictor of athletic performance before competition (e.g., Nagle et al., 1975). Limitations of this instrument have been pointed out, among which we can highlight those by Beedie (2005): (a) its factors have a predominantly negative orientation, (b) some items are associated with constructs not related to mood states, and (c) there are difficulties as regards distinguishing between emotions and mood states when using the POMS with athletes. Schacham (1983) found that individuals under conditions of stress or pain could take between 15 and 20 min to complete this questionnaire, which could be significantly limiting in pre-competitive sports contexts. In recent years, a number of tools based on the POMS have emerged, such as the Interactive Profile of Mood States in Sports (PIED in the Spanish acronym), created by Barrios (2011). This scale includes six lists of adjectives which correspond to each of the POMS scales, and in which athletes must indicate the intensity with which they are perceived on a scale ranging from 0 ("not at all") to 4 ("very much"). Another such tool is the POMS-VIC, developed by De la Vega et al. (2014). This is an adaptation of the version of the POMS developed by Andrade et al. (2008), in which three scales were used: mood state intensity, mood state valence and mood state control.

As an alternative to the POMS, another tool to assess pre-competitive psychological state which we have previously discussed is the reduced version of the CSAI-2. The CSAI-2R (Andrade et al., 2007) adapted from Martens et al. (1990a), consists of 17 items and has been widely used to assess levels of somatic anxiety, cognitive anxiety and self-confidence in the moments leading up to competition. The main limitation of this instrument is that, compared to other scales or instruments which give a broader profile of the athlete in these pre-competitive moments, the CSAI-2R only provides information on the anxiety (cognitive and somatic) and self-confidence variables.

A different approach which resolves many of the limitations found so far in regard to assessing precompetitive psychological state, is that developed by Buceta (2010). The Psychological State Test (TEP) was created with the objective of assessing the psychological state of athletes in an overall manner. Athletes can complete the test quickly (in under a minute) and at any time. It has mainly been used to better understand a team's collective disposition (it was initially used with soccer players) and, based on this information, to advise coaches and/or help athletes individually. The TEP is based on the PODIUM questionnaire for marathon runners, which was created in order to help these athletes adapt their psychological state prior to racing (Buceta et al., 2003; Larumbe et al., 2018). The PODIUM questionnaire consists of 20 items grouped into 6 psychological factors or variables: Somatic Anxiety (α = 0.83), Cognitive Anxiety (α = 0.77), Motivation (α = 0.86), Self-confidence (α = 0.72), Physical Perception (α = 0.90), and Social Support (α = 0.74). The response format used was Visual Analog Scales (VAS), from which two opposite adjectives were presented, upon which runners were asked to mark their responses on a 10-centimeter line according to how they felt at the time. Accordingly, the TEP (Buceta, 2010) consists of nine similar visual scales, each of which refers to one item in the questionnaire (9 items in total) and each consisting of two opposing adjectives. Each scale/item refers to a psychological variable related to sports performance (Weariness, General Tiredness, Positive arousal, Motivation, Self-Confidence, Concentration, Negative arousal, Anxiety, and Hostility), and the result is an overall profile of the athlete's psychological state. Athletes answer by placing an x on the one-hundred-millimeter line which separates both adjectives, depending on how they feel at that moment (self-report). Response coding is obtained by measuring the position of the athlete's mark on the line, considering each millimeter as a unit starting at 0 and ending at 100, and allowing the athlete's score on the item to be recognized. The small number of items, as well as the simplicity of the response format, mean that this test is suitable for use in the moments prior to competing, when anxiety levels can limit the athlete's capacity for self-observation. These limitations in self-observation have been explained mainly by the role of somatic anxiety in these pre-competition moments, with the anxiety felt most intensely in the hours and minutes leading up to the event (Martens et al., 1990a; Buceta et al., 2003).

As previously seen in relation to the different psychological profiles presented by Brazilian football players using the PMSL scale (Bueno and Souza, 2019), one of the possible applications of TEP could be related to the ability to assess whether different groups of athletes (based on gender, age, level of dedication, etc.) present distinct psychological profiles. There are various studies which compare the function of psychological variables related to the performance of different groups of athletes. In terms of comparisons between the influence of psychological variables on amateur (or non-professional) and professional (or elite) athletes, some examples include: Modroño and Guillén's (2006) study with regard to the motivation variable and the difference between amateurs, professionals, and non-professionals; and Kerr and Pos (1994) study which demonstrated a difference in the psychological mood experience between high level and low level competitive gymnasts, both in the training setting as well as the competition setting.

Although the TEP is a tool which is frequently used by sports psychologists in the applied field, its psychometric characteristics are unknown. The main objective of this paper is to study the psychometric characteristics of the Psychological State Test, through an Exploratory Factor Analysis and a Confirmatory Factor Analysis, in addition to the calculation of its Reliability indices (which will allow us to obtain an approximation as to the instrument's validity). On the other hand, in order to verify the TEP test, once its factorial structure was confirmed, we chose to propose a secondary objective: the comparison between the pre-competition profiles of amateur athletes and professional athletes in team sports. The aim is to verify whether there are statistically significant differences between the two groups in the moments leading up to competition in relation to the psychological variables assessed by the TEP and, if so, to examine what these differences consist of and whether they are in line with those found in previous studies which have compared psychological variables related to performance among these two groups.

MATERIALS AND METHODS

Instruments

An adaptation of the TEP in an online format was used. A pilot study was carried out with 20 athletes (selected based on the inclusion criteria established). The aim of this was to assess whether the adjectives used to describe the scales of the TEP were understood according to the psychological variable that was intended to be measured, and that words habitually used in the sporting context of the athletes were being used. As a result of this first study, it was found that participants were having trouble differentiating between the following adjectives used in the scales: Positive Arousal, Negative Arousal and Anxiety. One possible cause of this ambiguity regarding the conceptualization of the variables of the initial version of the TEP may be warranted by the author's description of them. According to Buceta (2010) the variable "positive arousal" emanates from motivation and "negative arousal" from stress. Furthermore, he stated in his text that, with regard to the relationship of these two variables with optimal level of arousal, the following can be taken into account: positive arousal with optimal levels of general arousal and negative arousal with an excessive level of general arousal. We thus consulted with five experts in Sports Psychology (each of whom had more than 10 years of practical experience) in order to assess how the scales which had led to ambiguity could be reconceptualized. This resulted in three changes: the "Positive arousal" scale was reconceptualized as "General arousal," the "Negative arousal" scale was reconceptualized as "Stress" and the scale initially referred to as "Anxiety" was renamed "Cognitive Anxiety." The use of a "cognitive anxiety" scale that was more specific than the "anxiety" variable in the initial version of the questionnaire was also supported by the conclusions of Martens et al. (1990a), in which they highlighted that the "cognitive anxiety" variable over "somatic anxiety" indicated levels that were higher and more stable throughout the days and hours prior to competition. Furthermore, a number of the adjectives were modified by means of selecting those which were repeated most among the group of participants in the pilot study and verified by the group of experts. The latter agreed that the "General fatigue" scale should be changed to read "Rest." It was expected that the positive description of the variable would be met with less resistance from the athletes. Finally, since our interest was in the evaluation of teams, we considered it important to include a scale that referred to the athletes' perception of team cohesion, given that numerous studies have related high levels of group cohesiveness with a greater perception of collective efficacy within teams (e.g., Heuze et al., 2006; Leo Marcos et al., 2011; Spink, 1990). The results of the pilot study led to the version of the TEP used in this paper, which consisted of 10 items: Rest, Self-Confidence, Motivation, Concentration, Hostility, Mood State, General arousal, Stress, Cognitive Anxiety, and Team cohesion (**Table 1**).

Participants

The procedure used to obtain the sample was "snowball" probabilistic sampling. To this end, we began by contacting professionals related to different sporting disciplines that work in institutions and/or teams. They helped us to recruit participants that fit the profile outlined from among their acquaintances and these, in turn, helped is to find other potential participants among their acquaintances. The inclusion criteria were (a) that they were involved in a team sport, (b) were over 16 years old, and (c) that their mother tongue was Spanish.

The total number of participants was 309 men and women aged between 16 and 53 years old (M: 22.5; and SD: 7.2). The sample was divided randomly into two groups. Therefore, part of the sample was used for Exploratory Factorial Analysis (sub-sample A) and the other for Confirmatory Factorial Analysis (sub-sample B). **Table 2** presents the specific data associated with each sub-sample (EFA and CFA).

Procedure

All the participants were given the online version of the TEP. We composed a brief explanatory message which we distributed through mobile messaging applications and which included a link to the website where the TEP was hosted. Accessing this site brought participants to: a presentation and explanation of the research objectives, information related to data protection based on the Spanish Organic Law concerning the Protection of Personal Data and Guarantee of Digital Right (Boletín Oficial del Estado, 2020); and a brief questionnaire concerning demographic aspects and details of interest for the classification of the sample. Finally, in order to participate in the study, the athletes had to accept all its terms and conditions. Having done so, the participants gained access to the TEP. After completing it they received (via the email address they had given us) their individualized profile along with the results of their psychological state and a brief explanation to assist interpretation.

Data Analysis

With regard to the main objective of this study, conducting the study of the psychometric properties of the TEP, the internal structure of the test has been studied. This was conducted via a cross validation process (Lasa et al., 2008).

With sub-sample A, an Exploratory Factorial Analysis was carried out using the Principal Component extraction method. This method was chosen in hopes of maximizing the degree of variance explained by the variables, in this way ensuring a factorial solution that is as representative as possible. The application scenarios were verified using the Kaiser-Meyer-Olkin Measure of Sampling Adequacy and the Bartlett Sphericity Test was also carried out on sub-sample A. In order to facilitate the interpretation of the significance of the selected factors, a Varimax rotation with Kaiser normalization was conducted (Kaiser, 1958). This implements an orthogonal rotation of the

TABLE 1 | Comparison between the variables in the original version of the TEP and those in our adaptation.

TEP 2010		Adaptación TEP 2019	
Variables	**Adjetivos opuestos**	**Variables**	**Adjetivos opuestos**
Cansancio general	Cansado/a – Fresco/a	Descanso	Cansado/a – Con energía
Autoconfianza	Con Confianza – Sin Confianza	Autoconfianza	Con Confianza – Sin Confianza
Motivación	Motivado/a – Desmotivado/a	Motivación	Motivado – Desmotivado
Concentración	Centrado/a – Disperso/a	Concentración	Centrado/a – Disperso/a
Hostilidad	Calmado/a – Enfadado/a	Hostilidad	Calmado/a – Enfadado/a
Desánimo	Contento/a – Triste	Estado de ánimo	Contento/a – Triste
Activación positiva	Activado/a – No activado/a	Activación general	Activado/a – No activado/a
Activación negativa	Tenso - Relajado	Estrés	Con presión – Sin presión
Ansiedad	Nervioso/a – Tranquilo/a	Ansiedad cognitiva	Preocupado/a – Tranquilo/a
		Cohesión	Desconectado/a del equipo – Integrado/a
TEP 2010		**Adaptation TEP 2019**	
Variables	**Opposing adjectives**	**Variables**	**Opposing adjectives**
General tiredness	Tired – Fresh	Rest	Tired – Energetic
Self-confidence	Confident - Not Confident	Self-confidence	Confident - Not Confident
Motivation	Motivated - Unmotivated	Motivation	Motivated – Unmotivated
Concentration	Focused – Scattered	Concentration	Focused – Scattered
Hostility	Calm – Angry	Hostility	Calm – Angry
Weariness	Happy – Sad	Mood state	Happy – Sad
Positive arousal	Activated - Not activated	General arousal	Activated - Not activated
Negative arousal	Tense – Relaxed	Stress	Under pressure – Not under pressure
Anxiety	Nervous – At ease	Cognitive anxiety	Worried – At ease
		Team cohesion	Disconnected from team - Integrated

First, we show the Spanish original adjectives, and after these, the English translation.

TABLE 2 | Sample characteristics.

	Sub-Sample A	Sub-Sample B
Sample size (n)	199 participants (M age = 23.29 / S.D. = 7.62)	110 participants (M age = 22.47 / S.D. = 6.56)
Women	33 participants (16.59%) (M age = 25.90 / S.D. = 8.95)	11 participants (10%) (M age = 23.54 / S.D. = 5.53)
Men	166 participants (83.41%) (M age = 22.50 / S.D. = 7.02)	99 participants (90%) (M age = 22.35 / S.D. = 6.65)
Years of practice	M = 13,98 / S.D. = 6.41	M = 14.68 / S.D. = 6.25
Sports	Soccer = 86.93% / Basketball = 12.06% Baseball = 0.5% / Handball = 0.5%	Soccer = 94.54% /Field Hockey = 2.72% / Volleyball = 1.81% / Indoor Soccer = 0.90%
Professionals vs. Amateurs	Professionals = 30.20% Amateurs = 69, 8%	Professionals = 25.45% Amateurs = 77, 27%

factorial axes based on the independence at a theoretical level as has been mentioned previously with respect to each of the factors. In addition, internal consistency was analyzed based on the Cronbach reliability index. For statistical analyses, we used the IBM SPSS Statistics 25 Software.

With sub-sample B, a Confirmatory Factorial Analysis (CFA) was carried out to estimate the parameters and evaluate the fit of the model resulting from the EFA. In this statistical test, the null hypothesis established that the proposed theoretic model is adjusted to the model resulting from the EFA data. If the null hypothesis is rejected, the proposed theoretic model is not adjusted to the model resulting from the EFA data. The Robust Maximum Likelihood method (RML) was used. The RML method can be applied when the variables observed are of a continuous nature and the data does not follow a normal distribution. In comparison to other estimation methods used in the CFA with ordinal variables, the RLM method (together with the Robust Unweighted Least Squares method, RULS) has demonstrated a better performance with fewer Type I error values (Holgado-Tello et al., 2018). We have used the LISREL 9.2 Software to carry out this analysis.

Related to the second objective, an analysis of the Student's *T* test for independent samples with an abnormality correction (Z) was carried out to evaluate the existence of significant statistical differences in relation to the level of dedication variable. Additionally, a MANOVA was carried out to find out if significant differences between the TEP psychological variables, depending on the level of dedication and the gender of the participants as well, as the interaction between both variables exist, considering each of the 9 TEP variables, as dependent variables; and the level of dedication (amateurs and professionals) and the gender (women and men) as independent variables.

RESULTS

Sample Normality Analysis and Kurtosis

The Kolmogorov-Smirnov Normality test indicates that the sample does not follow a normal distribution (p = 0.000 and d.f. = 200, in all items), so we opted to use non-parametric tests. In relation to Kurtosis or Asymmetry, all items obtained values which were considered adequate (<1.5 and -1.5), except for the item referring to Unity, which presented greater asymmetry (2.28). See **Table 3**.

Internal Consistency Analysis

Lastly, internal consistency was analyzed using Cronbach's Alpha Coefficient. The result showed a value of α = 0.823, taking into account all the variables of the TEP. As regards the individual factors, Factor 1 showed an internal consistency of α = 0.851 and Factor 2 of α = 0.726.

Exploratory Factor Analysis (EFA)

In the initial EFA, it was found that the Hostility variable shared part of its variance with the two factors proposed by the EFA (Factor 1 = 0.60 and Factor 2 = 0.36), which indicated that its factorial structure was not clear. It was therefore decided that the Hostility variable be removed. It was also found that the results of the weights of the other variables as regards the factors do not present significant changes and the percentage of variance explained increased from 59.5% to 61.4% (see **Table 4**).

Under these conditions, the result of the KMO (Kaiser-Meyer-Olkin) index was 0.843, suggesting that the data were considerably interrelated (≥0.84). Meanwhile, the results of the Bartlett Sphericity Test confirmed the applicability of the Factorial Analysis (Chi-square = 753.933; d.f. = 45, and p = 0.000). Two factors were obtained which explained 61.3% of the variance. **Table 5** shows the matrix of rotated components in which the clustering of the variables around the two factors can be observed. The sedimentation graphic can also be seen in **Figure 1**.

The results demonstrated a factorial structure consisting of two principal factors and reliability coefficients, both globally and at the factor level. The items were distributed around these factors, giving rise to one factor which includes variables which have a positive relationship with performance (Self-confidence, Motivation, Concentration, Rest, Team cohesion, and Mood state), and another with variables which negatively affect performance (General arousal, Stress, and Cognitive Anxiety). The relationship, positive or negative, between each variable and performance is in line with the findings of other studies in assessing this effect in athletes (as detailed in the Discussion section).

TABLE 3 | Results of normality test and Kurtosis.

	Kolmogorov-Smirnov			Kurtosis	Standard error
	Statistics	gl	Sig.		
Self-confidence	0.153	200	0.000	0.79	0.34
Motivation	0.163	200	0.000	0.68	0.34
Team cohesion	0.232	200	0.000	2.28	0.34
Concentration	0.133	200	0.000	0.12	0.34
Mood state	0.169	200	0.000	0.26	0.34
Rest	0.173	200	0.000	−0.73	0.34
General arousal	0.092	200	0.000	−0.99	0.34
Cognitive anxiety	0.099	200	0.000	−0.55	0.34
Hostility	0.125	200	0.000	0.48	0.34
Stress	0.117	200	0.000	−1.05	0.34

Confirmatory Factor Analysis (CFA)

According to the results obtained through the EFA, a first-rate factorial model, consisting of two factors, is proposed. The values obtained through the confirmatory factor analysis on the second sub-sample (B) indicated an appropriate model fit. Thus, with a confidence level of 99% the proposed model will be accepted according to the chi-squared test. The main global indices of fit goodness are: χ^2 Satorra-Bentler (d.f. = 26; p = 0.0232) = 37.99; RMSEA = 0.08 with a confidence interval of 90% between 0.032 and 0.122; RMR = 0.085 and GFI = 0.91. The following incremental indices were also analyzed: CFI = 0.95 and NNFI = 0.93. The model was specified as shown in **Figure 2**.

The results obtained confirm the two-factor structure proposed by the EFA. Thus, we could deduce that it is possible to get a TEP profile which would define the psychological state which makes it easier for athletes to use all their resources to deal with the sporting situation they face. This profile would involve high levels of factor 1 variable (performance enhancing variables; motivation, mood, rest, concentration, self-confidence and team cohesion); together with low levels of factor 2 variables (performance limiting variables: cognitive anxiety, stress and general arousal). These results would be in line with those described by Mouloud and El-Kadder (2016), who after conducting a bibliographic review in relation to the psychological characteristics of elite athletes, found that performance was enhanced when athletes present high levels of self-efficacy (self-confidence) and motivation, and low levels of cognitive anxiety.

Comparison of Profiles of Amateur and Professional Athletes

We used the Student's T test to carry out this analysis on the independent samples. We initially standardized the scale of the data to Z-scores to correct the sample's abnormality, and from these scores, we carried out the Student's T test and found the size of the effect by analyzing Cohen's d. The results indicated significant differences between both groups (amateurs vs. professionals) in the motivation variable (p = 0.05 and d = −0.2). The group of professional athletes obtained higher scores, which means that professional athletes presented higher levels of motivation than amateur athletes in the moments leading up to competition. The results show small size effect (≤0.2) in motivation following the Cohen's guidelines for d (Cohen, 1988). Data obtained from this analysis of each of the TEP variables is presented in **Table 6**.

TABLE 4 | Total variance explained by the TEP items.

Comp.	Initial self-values			Sums of the squared saturations of the extraction			Sums of Squared saturations of rotation		
	Total	% de Variance	% Accumulated	Total	% de Variance	% Accumulated	Total	% de Variance	% Accumulated
1	3.773	41.921	41.921	3.773	41.921	41.921	3.553	39.479	39.479
2	1.753	19.475	61396	1.753	19.475	61.396	1.972	21.916	61.396
3	791	8.792	70.188						
4	627	6.968	77.156						
5	524	5.819	82.975						
6	497	5.523	88.497						
7	404	4.487	92.985						
8	361	4.008	96.993						
9	271	3.007	100.00						

TABLE 5 | Matrix of rotated components and communality.

Variables	Components		h2
	1	2	
Motivation	**0.827 (F1)**	0.020	0.685
Self-confidence	**0.804 (F1)**	0.117	0.660
Mood state	**0.777 (F1)**	0.210	0.648
Concentration	**0.726 (F1)**	−0.42	0.529
Team cohesion	**0.692 (F1)**	0.094	0.488
Rest	**0.695 (F1)**	0.086	0.490
General arousal	0.011	**0.835 (F2)**	0.697
Cognitive anxiety	0.359	**0.759 (F2)**	0.705
Stress	−0.025	**0.789 (F2)**	0.624

The bold type indicates the higher values in each factor. h2 = Communalities.

In the MANOVA test, the multivariate contrasts obtained using Pillai's Trace indicate that there is no difference in the interaction of dependent variables in relation to the interaction between gender + the level of dedication (F[9,297] = 1.82, p = 0.066), nor in relation to the gender variable alone (F[9,297] = 1.08, p = 0.375) nor to the level of dedication (F[9,297] = 1.85, p = 0.059). However, differences in the dependent variables were found separately: in the self-confidence variable in relation to gender (F[1,305] = 3.95, p = 0.048), in the motivation variable in relation to the level of dedication (F[1,305] = 9.57, p = 0.002); and in the motivation (F[1,305] = 5.93, p = 0.015) and rest variables (F[1,305] = 7.06, p = 0.008) in the interaction between gender and level of dedication. In general, men present higher levels of self-confidence than women, with a difference of 6.75 points on the VAS scale (95%CI = -13.43, -0.069); for the level of dedication the levels of motivation are higher in professional athletes, with a difference of 11.63 points on the VAS scale (95%CI = -19.02, -4.23); and the analysis of interaction between both variables found that the level of motivation was higher in professional female athletes, 20.79 points on the VAS scale (95%IC = 7.72, 33.85; p = 0.002); and the level of rest in amateur male athletes was higher, 7.44 points on

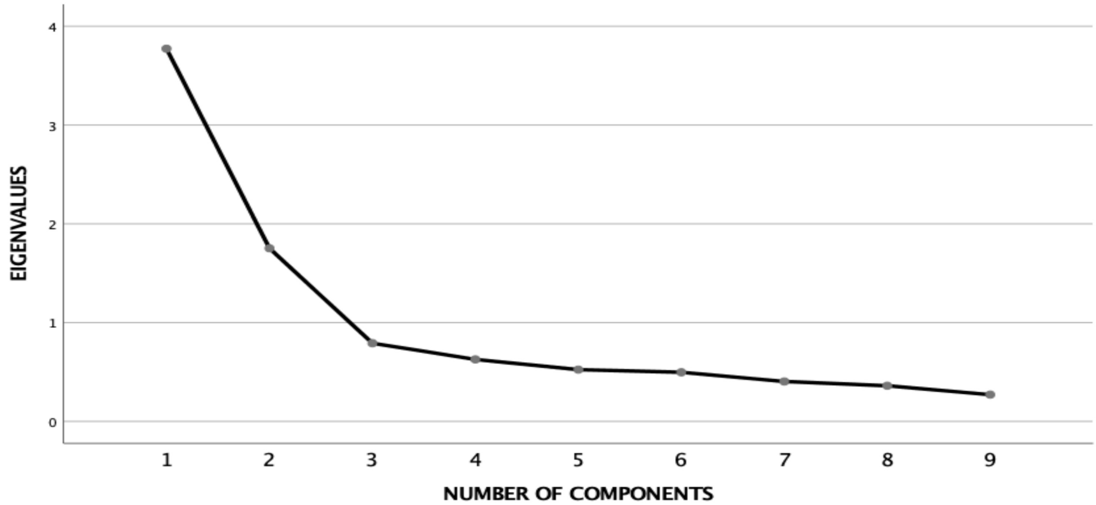

FIGURE 1 | Sedimentation graph.

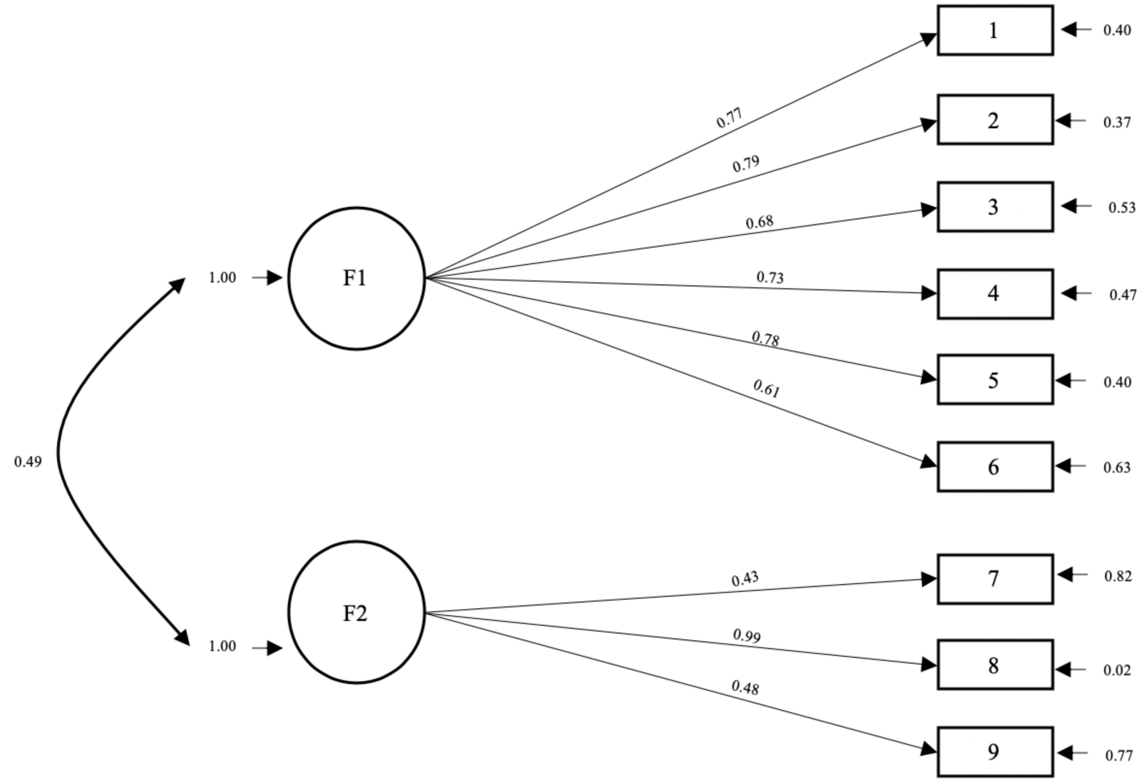

FIGURE 2 | Overview of the confirmatory factorial analysis results.

the VAS scale (95%IC = 0.25, 14.62; p = 0.043). No other differences were found.

DISCUSSION

This paper aims to check the characteristics of the new version of the Psychological State Test (TEP) developed by Buceta (2010) for team sports in Spanish. The main objective of the TEP is to evaluate the "pre-competitive psychological state" construct, for which the psychological variables that have been most studied and that have the greatest consensus in terms of their relationship and influence on sports performance were taken into account. This new version of the TEP was a response to the problems encountered by athletes in interpreting some of the adjectives used in the initial TEP. Accordingly, a pilot study was carried out with the collaboration of sports psychology experts in which several adjectives were modified so as to be more in line with the athletes' usual vocabulary. The reconceptualization of three of the scales from the original TEP was also carried out. These were General arousal, Stress and Cognitive Anxiety.

Regarding internal consistency, the TEP was found to be reliable ($\alpha > 0.80$). As a general criterion, George and Mallery (2003) consider a coefficient α greater than.80 as "good". In regard to each individual factor, Factor 1 would also have a good internal consistency (>0.80) and Factor 2 would have an acceptable internal consistency (in the classification established by these authors) of >0.70. The result of the Exploratory Factorial Analysis specified a structure of two main factors. Factor 1 is composed of six scales with weights between 0.69 and 0.83 and Factor 2 is composed of three scales, with weights between 0.76 and 0.83. This factorial solution explains 61.4% of the variance explained. This two-factor model was supported by results obtained through a Confirmatory Factorial Analysis.

Factor 1 would consist of the scales: Self-confidence, Motivation, Concentration, Rest, Team cohesion, and Mood state. These scales could be considered within the performance facilitators. Based on the related scientific literature the high scores on these scales could be interpreted as positive regarding psychological state profile prior to competition in sports teams.

Some examples are as follows: athletes with high levels of self-confidence/self-efficacy tend to be involved for the greatest length of time, to have a higher level of effort and to persist in order to achieve their goals (Dosil, 2008). Meanwhile, motivation is considered essential in order for athletes to acquire the commitment, perseverance and tolerance of frustration which competition demands. Therefore, it is considered that it should be high in athletes in order to assist performance (Buceta, 2003). High levels of concentration facilitate the implementation of strategies and resources for dealing with competitive events. According to Dosil (2008) concentration is key for athletes attaining their optimum performance as well as for facilitated learning. In relation to the level of perceived Rest, high levels

TABLE 6 | Summary of data comparison by group, Amateurs vs. Professionals.

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	**Sig.**	**t**	**df**	***Sig. (2-tailed)* p**	***Mean Difference (d Cohen)***	**Std. Error Difference**	**95% Confidence Interval of the Difference**	
									Lower	**Upper**
SELF-CONFIDENCE (Z score)	Equal variances assumed	0.92	0.33	−1.02	307	0.30	−0.13	0.13	−0.38	0.12
	Equal variances not assumed			−0.97	145.03	0.33	−0.13	0.13	−0.39	0.13
MOTIVATION (Z score)	Equal variances assumed	0.82	0.36	−1.89	307	**0.05**	−0.24	0.12	−0.48	0.01
	Equal variances not assumed			−1.95	172.57	**0.05**	−0.24	0.12	−0.47	0.00
TEAM COHESION(Z score)	Equal variances assumed	1.07	0.30	−0.49	307	0.62	−0.06	0.13	−0.31	0.18
	Equal variances not assumed			−0.53	187.89	0.59	−0.06	0.12	−0.29	0.17
CONCENTRATION (Z Score)	Equal variances assumed	0.89	0.34	−0.86	307	0.38	−0.11	0.13	−0.36	0.14
	Equal variances not assumed			−.83	148.24	0.40	−0.11	0.13	−0.37	0.15
MOOD STATE (Z Score)	Equal variances assumed	1.59	0.20	1.14	307	0.25	0.14	0.12	−0.10	0.39
	Equal variances not assumed			1.08	143.93	0.27	0.14	0.13	−0.12	0.40
REST (Z Score)	Equal variances assumed	1.82	0.17	0.90	307	0.36	0.11	0.13	−0.13	0.36
	Equal variances not assumed			0.86	147.44	0.38	0.11	0.13	−0.15	0.37
GENERAL AROUSAL (Z Score)	Equal variances assumed	0.00	0.99	−1.00	307	0.31	−0.13	0.13	−0.37	0.12
	Equal variances not assumed			−0.99	155.95	0.32	−0.13	0.13	−0.38	0.12
COGNITIVE ANXIETY (Z Score)	Equal variances assumed	0.33	0.56	−0.00	307	0.99		0.13	−0.25	0.25
	Equal variances not assumed			−0.00	153.04	0.99	−0.00	0.13	−0.25	0.25
STRESS (Z Score)	Equal variances assumed	0.05	0.81	−0.47	307	0.63	−0.06	0.13	−0.31	0.19
	Equal variances not assumed			−0.47	156.59	0.63	−0.06	0.13	−0.31	0.19

in this variable can stimulate the athletes' participation in the sporting activity. Studies conducted with the POMS determine that when athletes presented high scores on the Fatigue scale, this was related to a reduction in physical capacity and the athletes' perception of personal effectiveness (Terry, 1997). Hanin (2000b), meanwhile, found that soccer players considered feeling motivated, confident and alert (attention) as states which assisted performance. In the same study, the players identified feeling tired and insecure as states which impaired their performance. Lastly, in relation to the variable of Team cohesion, Carron et al. (2002) carried out a meta-analysis of 46 studies which looked at the association between unity and sporting success. The results confirmed the positive relationship (from "moderate" to "significant") between these variables. The more cohesive teams tended to have more success and the successful teams were more likely to develop a sense of unity.

On the other hand, "Factor 2" would be composed of the variables General arousal, Cognitive Anxiety, and Stress. In this case, we could consider that high scores on these scales can restrict the implementation of the athletes' resources and skills and therefore adversely affect their performance based on the related scientific literature.

Some examples of these studies are as follows: in relation to General arousal, the results obtained in our study are in line with the inverted "U" theory of Yerkes and Dodson (1908). These authors postulate that a higher level of performance could be attributed to medium levels of arousal, with the highest and lowest levels assisting performance the least. On the other hand, Cognitive anxiety (which refers to the degree to which a person worries or has negative thoughts) and Stress are considered as non-functional in relation to sporting performance (Weinberg and Gould, 2007; McGrath, 1970). In the Multidimensional Anxiety Theory, the authors Martens et al. (1990b) argued that anxiety may have an impact on attention, concentration and athletes' decision making. In terms of correlation, the authors stated that cognitive anxiety has a negative linear relationship with performance. In other words, the higher the levels of cognitive anxiety, the worse the performance. The ideal profile to benefit performance in this regard would involve athletes presenting medium levels of General arousal and low levels of Cognitive Anxiety and Stress.

We can conclude the existence of this optimal psychological profile that appears to facilitate sports performance, defined by high levels of Factor 1 variables and moderate and low levels of the variables grouped around factor 2. This follows the findings of Larumbe (2006) in his studies on the PODIUM in which he describes a "positive psychological disposition" among marathon runners characterized by high levels of self-confidence and motivation and with controlled arousal levels and anxiety.

The Hostility scale was not clearly associated with any of the resulting factors. Lane and Terry (2005) explained why the factors of stress and hostility are associated with good performance in some studies and not in others. According to these authors, depressive mood determines the functional impact of stress and hostility on performance. Without the presence of depressive symptoms, pressure and hostility contribute to increasing the athletes' determination. However, with depressive symptoms, stress and hostility did not benefit performance. We can conclude that hostility can mobilize athletes and lead to greater perseverance and willingness to compete with all their available resources. On the other hand, high levels of hostility may be related to greater difficulty controlling general arousal and therefore to issues regarding maintaining focus and being precise with their movements or technical actions, and/or lead to greater impulsivity in making decisions. The apparent need (based on the result of the factor analysis) to assess the effect of individual hostility on each athlete, leads us to conclude that it is not a good scale to use to assess the collective disposition of teams, and so it was removed from the instrument.

The second objective of this study was to look at whether there were significant differences as regards the psychological profile of amateur athletes and professional athletes. In this regard, we can conclude that professional athletes presented higher levels of motivation. In line with our findings, Modroño and Guillén (2006) did find significant differences in motivation levels between competitors and non-competitors, with the levels of extrinsic motivation found to be higher in competitors. It is important to understand that it is in the competition where windsurfers can win cash or material prizes, which may explain this increased level of extrinsic motivation; hence, it can be compared to the prizes or remuneration awarded to professionals in team sports, as is the case with our sample. One would expect, in such cases, that motivation levels would therefore be higher than those of amateur athletes. In this sense, Carpenter and Yates (1997) carried out a study on amateur and semi-professional football players, the authors found that semi-professional football players, when compared to amateurs, considered the financial and status enhancements of their sport to be the main reason for playing. Halldorsson et al. (2012), reported that elite athletes report higher levels of motivation and commitment than non-elites. Mallett and Hanrahan (2004) found that Olympic and World Championship level athletes exhibit self-determined forms of motivation, and are achievement oriented, highly driven, and self-believing. If we consider the gender variable, our results found that men presented higher levels of self-confidence than women, in line with the findings of a study on recreational runners carried out by Larumbe-Zabala et al. (2019).

With regard to the possibilities of practical applications of the TEP, several studies related to the advice to coaches and technical bodies in the design and management of the pre-match talks are worth highlighting. Vargas-Tonsing and Bartholomew (2006) studied the effect that pre-match talks had on the athletes' perception of self-efficacy. Their findings could not confirm that the various pre-match talks analyzed resulted in any significant effect on the participants' levels of self-efficacy. They concluded that in order for pre-match talks to have a positive effect on performance (to improve the players' levels of self-efficacy), the coaches had to be aware of the emotional intensity of the athletes prior to the talk in order to avoid generating states of over-arousal or anxiety. They concluded that it was very important for the coach to be aware of the players' prior emotional state in order to thus tailor their talk and achieve beneficial effects which stimulate the appropriate arousal levels.

In this regard, the TEP can be a useful tool for advising coaches on the psychological state of their teams prior to competition. One practical application in this regard was that presented by Díaz-Tendero et al. (2018) involving a field hockey team. The results showed that the percentage of times the coach used the information provided by the TEP in his pre-game talk was 92%. The evaluation obtained by the technical team after each game regarding the usefulness of this information was an average of 7.8 points (on a scale of 0 to 10), and the degree to which, according to the players, the "team profile" matched their perception of the team was an average of 8.1 points (on a scale of 0 to 10). In this case, the method used to complete the TEP was text messaging (SMS) via mobile phones. However, this system had many practical limitations with regard to completing the questionnaire and the delay in receiving the results from the players and the coach/coaching staff. To address these limitations, we used the online version of the TEP in this study.

In recent years, the number of apps and online resources which support psychological intervention tools has grown exponentially. Concepts such as cognitive ergonomics (which encompasses the psychological aspects of people's interaction with technology) and usability (a discipline which studies the processes involved in people's interaction with interactive products in order to facilitate their use), are key when it comes to designing and evaluating technological tools to assist interventions or evaluations in the applied field of sports psychology. In future research projects, these aspects should be looked at more specifically as regards the development of an application for smartphones, and other mobile devices, which would support the TEP. This app could make it easier for athletes to receive notifications in order to complete the questionnaire and to store the results of these measurements of the psychological state of each player. It could also provide coaching staff with an access profile: in this section, the psychological profiles of the team would be stored for analysis and for the potential integration of this information, along with that from other areas related to sporting performance.

The TEP has proved to be a reliable tool for assessing pre-competitive psychological states in team sports. Unlike other tools that attempt to evaluate the same construct, the TEP provides information on a larger number of variables by broadening the profile that can be obtained from athletes compared to, for example, the CSAI-2R (Andrade et al., 2007), which only provides information on three variables (somatic anxiety, cognitive anxiety and self-confidence). Another advantage that the TEP presents over other tools is the minimal time required for completion, which facilitates the precision of self-observation that athletes need to respond to this type of test in those pre-competitive times when higher levels of anxiety (both somatic and cognitive) are detected (Martens et al., 1990b). Furthermore, an innovative contribution to the online version of TEP used for this study is the automation of the immediate correction and feedback that athletes receive. This is a great advantage in that it allows sports psychologists to work with the athletes on their mental preparation prior to a competition, as well as facilitating self-regulation by the athletes themselves.

However, this study presents a number of limitations that should be taken into account. The sampling was done using the snowball method and not a simple random sampling which could give more solidity to the data. On the other hand, the sample used in the two factorial analyses (EFA and CFA) was collected at the same time, but it is more appropriate to collect the data consecutively. To support the conclusions regarding the differences between amateurs and professionals, it would be advisable to expand the sample of professionals in order to have matching numbers of participants from both groups.

For future research projects, we think it is important to evaluate the predictive validity of the TEP so that it can be a useful tool when predicting behaviors related to athletes' performance. In this line, another aspect to conduct further research on is the usefulness of coaching programs for the psychological management of teams based on the psychological state profiles provided by TEP.

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 Commission of the PhD program in Health Psychology and Bioethics Commission of National University of Distance Education (UNED). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

PD-T, MP-L, and AL performed conceptualization, methodology, and investigation. AL performed formal analysis and wrote, reviewed, and edited the manuscript. PD-T and MP-L performed resources and data curation. PD-T wrote original draft. MP-L performed supervision. All authors contributed to the article and approved the submitted version.

REFERENCES

Andrade, E., Arce, C., Armental, J., Rodríguez, M., and De Francisco, C. (2008). Indicadores del estado de ánimo en deportistas adolescentes según el modelo multidimensional del POMS. *Psicothema* 20, 630–635.

Andrade, E., Lois, G., and Arce, C. (2007). Propiedades Psicométricas de la versión española del inventario de ansiedad competitiva CSAI.2R en deportistas. *Psicothema* 19, 150–155.

Ávila, A., and Giménez de la Peña, A. (1991). Los adjetivos en tareas de evaluación psicológica: propiedades y valor. *Rev. Psicol. Gen. Apl.* 44, 465–475.

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

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York, NY: Freeman.

Barrios, R. (2011). *Elaboración de Un Instrumento Para Evaluar Estados de Ánimo en Deportistas de Alto Rendimiento*. Thesis. La Habana: UCCFD Manuel Fajardo.

Beedie, C. J. (2005). "It's the POMS, it measures mood–doesn't it," in *Promoting Health and Performance for life: International Society of Sport Psychology (ISSP)*, ed. T. Morris (Sydney, NSW: Sydney Convention and Exhibition Centre).

Boletín Oficial del Estado (2020). *Ley Orgánica 3/2018, de 5 de Diciembre, de Protección de Datos Personales y Garantía de Los Derechos Digitales*. Available online at: https://www.boe.es/eli/es/lo/2018/12/05/3 (accessed on 2 February 2020)

Boutcher, S. (2008). "Attentional processes and sport performance," in *Advances in Sport Psychology*, ed. T. H. Horn (Champaign: Human Kinetics), 325–338.

Buceta, J. (2003). "Evaluación psicológica en el deporte," in *Evaluación Psicológica Aplicada a Diferentes Contextos*, ed. V. Del Barrio (Madrid: Universidad Nacional de Educación a Distancia), 309–349.

Buceta, J. (2010). "El test de estado psicológico (TEP) para la evaluación de deportistas," in *Experiencias en Psicología del Deporte*, eds J. M. Buceta and E. Larumbe (Madrid: Dykinson), 47–60.

Buceta, J., López de la Llave, A., Pérez-Llantada, M., Vallejo, M., and Pino, M. (2003). Estado Psicológico de los corredores populares de maratón en los días anteriores a la prueba. *Psicothema* 15, 273–277.

Buceta, J. M. (1998). *Psicología del Entrenamiento Deportivo*. Madrid: Dykinson.

Bueno, J. L. O., and Souza, V. C. (2019). "Mood states as motivational determinats od football performance," in *Football Psychology: From Theory to Practice*, eds E. Konter, J. Beckmann, and T. M. Loughead (Baskerville: Routledge), 9–22. doi: 10.4324/9781315268248-2

Campbell, D. J., and Pritchard, R. (1976). "Motivation theory in industrial and organizational psychology," in *Handbook of Industrial and Organizational Psychology*, ed. M. D. Dunnette (Chicago, IL: Rand McNally), 63–130.

Carpenter, P. J., and Yates, B. (1997). Relationship between achievement goals and the perceived purposes of soccer for semiprofessional and amateur players. *J. Sport Exer. Psychol.* 19, 302–311. doi: 10.1123/jsep.19.3.302

Carron, A., Colman, M. M., Wheeler, J., and Stevens, D. (2002). Cohesion and performance in sport: a meta-analysis. *J. Sport Exerc. Psychol.* 24, 168–188. doi: 10.1123/jsep.24.2.168

Cerin, E. (2003). Anxiety versus fundamental emotions as predictors of perceived functionality of pre-competitive emotional states, threat and challenge in individual sports. *J. Appl. Sport Psychol.* 15, 223–238. doi: 10.1080/10413200305389

Clancy, R. B., Herring, M. P., MacIntryre, T. E., and Campbell, M. J. (2016). A review of competitive sport motivation research. *Psychol. Sport Exer.* 27, 232–242. doi: 10.1016/j.psychsport.2016.09.003

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale: Erlbaum.

De la Vega, R., Ruiz Barquín, R., Borges, P. J., and Tejero-González, C. M. (2014). Una nueva medida tridimensional del estado de ánimo deportivo: el POMS-VIC. *Cuadernos Psicol. Dep.* 14, 37–46. doi: 10.4321/s1578-84232014000200005

Díaz-Tendero, P., López de la Llave, A., and Pérez-Llantada, M. (2018). "Adaptación del cuestionario TEP para su aplicación "a distancia" con jugadores semiprofesionales de Hockey hierba. Resultados de una experiencia práctica," in *Cu4rto Foro de Investigadores Noveles*, eds M. C. Pérez-Llantada and P. R. Montoro (Madrid: Facultad de Psicología de la UNED), 124–127.

Dosil, J. (2008). *Psicología de la Actividad Física y del Deporte*. Madrid: McGraw-Hill. doi: 10.4272/978-84-9745-351-6

Engelmann, A. (1986). Lep: uma lista, de origen brasileira, para medir a presemça de estados de ânimo no momento em que está sendo respondida. *Ciênc. Cult.* 38, 121–114.

Engelmann, A. (2002). "A new scale for evaluating hedonic preceps," in *Fechner Day. Annual Meeting of The International Society for Psychophysics*, eds J. A. Da Silva, E. H. Matsushima, and N. P. Ribeiro Filho (Cassis: The International Society for Psychophysics), 191–196.

George, D., and Mallery, P. (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference 11.0 Update*. Boston, MA: Pearson Education.

Halldorsson, V., Helgason, A., and Thorlindsson, T. (2012). Attitudes, commitment and motivation amongst Icelandic elite athletes. *Int. J. Sport Psychol.* 43, 241.

Hanin, Y. (1997). Emotions and athletic performance: individual zones of optimal functioning model. *Eur. Yearb. Sport Psychol.* 1, 29–72.

Hanin, Y. (2000b). "Soccer and emotions: enhancing or impairing performance?," in *Soccer and Science in an Interdiciplinary Perspective*, ed. J. Bangsbo (Copenhagen: Institute of Exercise and Sport Sciences Copenhagen University), 69–89.

Hanin, Y. (ed.) (2000a). *Emotions in Sport*. Champaign IL: Human Kinetics. doi: 10.5040/9781492596233

Hardy, L. (1990). "A catastrophe model of performance in sport," in *Stress and Performance in Sport*, eds J. G. Jones and L. Hardy (Chichester: Wiley), 81–106.

Heuze, J., Raimbault, N., and Fontayne, P. (2006). Relationships between cohesion. collective efficacy, and performance in professional basketball teams: an examination of mediating effects. *J. Sports Sci.* 24, 59–68. doi: 10.1080/02640410500127736

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

Jones, G. (1991). Recent developments and current issues in competitive state anxiety research. *Psychologist* 4, 152–155.

Kaiser, H. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychimetrika* 23, 187–200. doi: 10.1007/bf02289233

Kerr, J. H., and Pos, E. H. (1994). Psychological mood in competitive gymnastics: an exploratory field study. *J. Hum. Move. Stud.* 26, 175–185.

Lane, A., and Terry, P. (2005). "Test of a conceptual model of mood performance relationships with a focus on depression: a review and synthesis five years on," in *Promoting Health & Performance for Life: Proceedings of the ISSP 11th World Congress of Sport Psychology [CDROM]*, eds T. Morris, S. Hanrahan, L. Ievleva, G. Kolt, and P. Tremayne (Sydney: Society of Sports Psychology -ISSP).

Larumbe, E. (2006). *Creación de un Instrumento Para la Medida del Estado Psicológico del Corredor de Maratón Popula*. Thesis. Madrid: Universidad de Educación a Distancia UNED.

Larumbe, E., Pérez-Llantada, M., López de la LLave, A., and Buceta, J. (2018). Desarrollo y características psicométricas preliminares del cuestionario PODIUM. *Cuadernos Psicol. Dep.* 15, 41–52. doi: 10.4321/S1578-84232015000300004

Larumbe-Zabala, E., Garciìa-Lluch, J. J., Agea, E., and Peris-Delcampo, D. (2019). Goal-setting strategy and psychological differences in marathon runners compared by gender. *J. Hum. Sport Exer.* 14, 725–735. doi: 10.14198/jhse.2019.144.02

Lasa, A., Holgado, F. P., Carrasco, M. A., del Barrio, and Gandara, M. V. (2008). The structure of Bryant's empathy index for children: a cross-validation study. *Span. J. Psychol.* 11, 670–677. doi: 10.1017/s1138741600004674

Lazarus, R. (2000). How emotions influence performance in competitive sports. *Sport Psychol.* 14, 229–252. doi: 10.1123/tsp.14.3.229

Leo Marcos, F., García Calvo, T., Sánchez Miguel, P., and De la Vega Marcos, R. (2011). Relación entre la cohesión de equipo, la eficacia percibida y el rendimiento en equipos masculinos de jóvenes futbolistas. *Rev. Iberoamer. Psicol. Ejercicio Dep.* 6, 47–62. doi: 10.1174/021347413804756023

Mallett, C. J., and Hanrahan, S. J. (2004). Elite athletes: why does the "fire" burns so brightly? *Psychol. Sport Exer.* 5, 186–200. doi: 10.1016/S1469-0292(02)00043-2

Malmo, R. B. (1959). Activation: a neuropsychological dimension. *Psychol. Rev.* 66, 367–386. doi: 10.1037/h0047858

Martens, R., Burton, D., Vealey, R., Bump, L., and Smith, D. (1990a). "Development and validation of the competitive state anxiety inventroy-2," in *Competitive Anxiety in Sport*, eds R. Martens, R. Vealey, and D. Burton (Champaign: Human Kinetics), 117–178.

Martens, R., Vealey, R., and Burton, D. (1990b). *Competitive Anxiety in Sport*. Champaign: Human Kinetics.

McGrath, J. (1970). *Social and Psychological Factors in Stress*. New York, NY: Holt, Rinehart and Winston.

McNair, D., Lorr, M., and Droppleman, L. (1971). *Manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.

Modroño, C., and Guillén, F. (2006). Motivation and self-concept in windsurfers: a study of professional and recreational participants. *Rev. Psicol. Dep.* 25, 105–112.

Morgan, W. (1980a). Test of champions: the iceberg profile. *Psychol. Today* 14, 92–108.

Morgan, W. (1980b). The trait psychology controversy. *Res. Q. Exer. Sport* 51, 50–76. doi: 10.1080/02701367.1980.10609275

Morgan, W., and Jhonson, R. (1977). Psychologic characterization of the elite wrestler: a mental health model. *Med. Sci. Sports* 9, 55–56. doi: 10.1249/00005768-197721000-00049

Morgan, W., and Jhonson, R. (1978). Personality characteristics of successful and unsuccessful oarsmen. *Int. J. Sport Psychol.* 9, 119–133.

Morgan, W., and Pollock, M. (1977). Psychologic characterization of the elite distance runner. *Ann. N. Y. Acad. Sci.* 301, 382–403. doi: 10.1111/j.1749-6632.1977.tb38215.x

Mouloud, K., and El-Kadder, B. A. (2016). Self-efficacy, achievement motivation and anxiety of elite athletes. *IOSR J. Sport Phys. Educ.* 3, 45–48. doi: 10.9790/6737-03044548

Nagle, F., Morgan, W., Hellickson, R., Serfass, R., and Alexander, J. (1975). Spotting success traits in Olympic contenders. *Phys. Sports Med.* 18, 85–92. doi: 10.1080/00913847.1975.11948293

Nuñez, A., and Garcia-Mas, A. (2017). Relacioìn entre el rendimiento y la ansiedad en el deporte: una revisioìn sistemaìtica. *Rev. Retos* 32, 172–177.

Ponseti, F. J., Sesé, A., and Garcia-Mas, A. (2016). The impact of competitive anxiety and parental influence on the performance of young swimmers. *Rev. Iberoamer. Psicol. Ejercicio Dep.* 11, 229–237.

Schacham, S. (1983). A shortened version of the profile of mood states. *J. Pers. Assess.* 47, 305–306. doi: 10.1207/s15327752jpa4703_14

Silva, R. C. (2017). *Ataque e Defensa: A Influencia da Posiçao Nos Estados de Ânimo Momentâneos em Atletas de Futebol Americano de alto Redimento*. Master's thesis. Sao Paulo: University of Sao Paolo.

Souza, V. C. (2014). *Relaçoes Entre Estados de Ânimo Pré-Competitivos e Desempenho de Atletas de Alto Rendimiento em Jogos de futsal*. Master's thesis. Sao Paulo: University of Sao Paulo.

Spink, K. (1990). Group cohesion and collective efficacy of volleyball teams. *J. Sport Exerc. Psychol.* 12, 301–311. doi: 10.1123/jsep.12.3.301

Terry, P. (1997). "The application of mood profiling with elite sport performers," in *Sport Psychology in Performance*, ed. R. J. Butler (London: Plenum Press), 3–32.

Vargas-Tonsing, T., and Bartholomew, J. (2006). An exploratory study of the effects of pregame speeches on team efficacy beliefs. *J. Appl. Soc. Psychol.* 36, 918–933. doi: 10.1111/j.0021-9029.2006.00049.x

Weinberg, R., and Gould, D. (2007). *Foundations of Sport and Exercise Psychology*. Champaign: Human Kinetics.

Yerkes, R., and Dodson, J. (1908). The relation of strenght of stimulus to rapidity of habit formation. *J. Comp. Neurol. Psychol.* 18, 459–482. doi: 10.1002/cne.920180503

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.

The handling Editor declared a shared affiliation with the authors at time of review.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 November 2020
doi: 10.3389/fpsyg.2020.569438

Emotion Regulation Questionnaire for Cross-Gender Measurement Invariance in Chinese University Students

Ying Zhang[1,2,3] *and Yufang Bian*[1,2,3]*

[1] *Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University, Beijing, China,* [2] *Institute of Mental Health and Education, Beijing Normal University, Beijing, China,* [3] *Child and Family Education Research Center, Beijing Normal University, Beijing, China*

OPEN ACCESS

Edited by:
José Antonio Lozano Lozano,
Autonomous University of Chile, Chile

Reviewed by:
Juan Yang,
Hainan Medical University, China
Rosalba Rosato,
University of Turin, Italy

***Correspondence:**
Yufang Bian
bianyufang66@126.com

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

Received: *04 June 2020*
Accepted: *06 October 2020*
Published: *04 November 2020*

Citation:
Zhang Y and Bian Y (2020) Emotion Regulation Questionnaire for Cross-Gender Measurement Invariance in Chinese University Students. Front. Psychol. 11:569438. doi: 10.3389/fpsyg.2020.569438

Objectives: Emotion regulation has been extensively studied in various areas of psychology. The Emotion Regulation Questionnaire (ERQ) was developed to assess two specific constructs associated with emotion control—cognitive reappraisal and expression suppression (Gross and John, 2003). The instrument displayed sound psychometric properties; however, to date, inquiry regarding the measure's characteristics has been limited. This study aims to measure cross-gender invariance [measurement invariance (MI)] in Chinese undergraduates using the ERQ.

Methods: This study measured the psychometric properties of the ERQ in a sample of 847 Mainland China undergraduates (401 males and 446 females) through confirmatory factor analysis. The tests of MI were used to examine potential structural differences based on gender.

Results: The findings supported the measure's original structure with all demographic groups and demonstrated exceptional fit. Additional normative data for gender and ethnic groups are included as well. The results also supported the use of the instrument in future research.

Conclusion: The two-factor structure in the ERQ establishes a cross-gender equivalence between males and females in Chinese college students. This study supports the use of the instrument in future research.

Keywords: emotion regulation, cognitive reappraisal, expression suppression, across gender, measurement invariance

INTRODUCTION

Emotion regulation implies the process that individuals use to regulate, experience, and express their emotions (Gross, 2002; John and Gross, 2007; Wang et al., 2020). Using emotion regulation strategies, individuals could alter their emotions in physiological activities, subjective experiences, and behavior (Ochsner and Gross, 2008; Miao, 2009; Gratz et al., 2015). Individuals regulate their emotions using the emotion regulation strategy, which enables them to improve, maintain, or reduce one or several emotional reactions (Gross, 1998; Dunsmore et al., 2013). Emotion regulation can influence individuals' physical health (e.g., sleep quality) (Minkel et al., 2012), mental health (e.g., social anxiety and other negative emotions) (Goldin et al., 2012), interpersonal relationships (e.g., partnership and parent–child relationship) (English et al., 2012; Shi et al., 2019). Reportedly, individual emotion regulation could appear and often play a role in daily life and various interpersonal interactions (Gross et al., 2006). Emotion regulation has become a pressing issue in the field of psychology.

Successful emotion regulation strategies are crucial for an individual's emotion (Cai et al., 2012), social support (English et al., 2012; Goldin et al., 2012), and subjective well-being (Parkinson and Totterdell, 1999; Gross and John, 2003; McRae et al., 2012). To clearly and directly assess emotion regulation strategies, Gross (1998) developed the Emotion Regulation Questionnaire (ERQ) based on the process model of emotion regulation [i.e. ERQ, compiled by Gross (1998) at Stanford University, United States, which focuses on the frequency of individual utilization of emotion regulation strategies by measuring two dimensions: "cognitive reappraisal" and "expression suppression," [1] (Chinese version)]. Cognitive reappraisal is an antecedent-focused strategy and often tries to reinterpret events positively (e.g., *When I'm faced with a stressful situation, I make myself think about it in a way that helps me stay calm*) (John and Gross, 2004). Expressive suppression, however, attempts to suppress, hide, or reduce emotional expression (e.g., *I keep my emotions to myself*) (John and Gross, 2004). Gross's ERQ comprises 10 items, including 6 items for measuring the cognitive reappraisal dimension and 4 measuring the expression suppression dimension. In recent years, ERQ has been extensively used in the measurement of special and normal groups and has been translated into different languages and widely used worldwide (Liu et al., 2017; Lotfi et al., 2019; Pastor et al., 2019; Wang et al., 2020). ERQ is acceptable to excellent levels of internal consistency reliability across various types of participants (posttraumatic stress disorder, anxiety disorders, normal adolescents, and young adults) (Gross and John, 2003; Wiltink et al., 2011; Spaapen et al., 2014; Preece et al., 2019).

The effects of cognitive reappraisal and expressive suppression are manifold depending on the cultural background. In the Western cultural background, the impact of cognitive reappraisal is more positive such as better social support and lower level of psychopathology symptoms (Moore et al., 2008; Joormann and Gotlib, 2010; McRae et al., 2012), whereas the impact of expressive suppression is more negative such as higher level of depression and anxiety (Moore et al., 2008; Eftekhari et al., 2009). However, in the Asian cultural background, cognitive reappraisal could be an ineffective strategy for some minority groups experiencing oppression, and expressive suppression appears to be less harmful (Soto et al., 2012; Su et al., 2015; Wang et al., 2020). Indeed, most studies that investigated the ERQ's psychometric properties are under Western cultural background (Australian Bureau of Statistics, 2017), and a few have focused on the Asian cultural background (e.g., Mainland China) (Preece et al., 2019). Wang et al. (2007) explored the ERQ's psychometric properties in Chinese college students, and Wang et al. (2020) tested the ERQ's psychometric properties in Chinese rural-to-urban migrant adolescents and young adults; both studies found that the reliability and validity of ERQ fulfilled the requirements of psychometrics.

The research testing measurement invariance (MI) across different populations using the confirmatory factor analysis (CFA) has highlighted the significance of identifying discrepancies in factor and parameter characteristics and assessing how this could affect and distort between-group comparisons (Meredith, 1993). Wang et al. (2007) and Wang et al. (2020) focused on Oriental culture under the background of people's emotion regulation strategies, and their studies' impact on the measurement tool laid the foundation. Although both studies mentioned above in China reported worthwhile findings, the consideration of MI did not receive attention. Thus, it is crucial to determine whether the underlying traits measured by the measurement (e.g., ERQ in this study) are equivalent across different groups. For example, the ERQ measuring emotion regulation could exhibit variance across gender. Despite this inconsistency, measurement has always been a combination of males and females without distinction, and the latent construct of emotion regulation being measured could be observed in the male group but not in the female group, or *vice versa*. In this instance, variance is expected, and perhaps, the construct cannot be measured in the female or male group. Consequently, the scale could be an excellent measure of the latent construct of emotion regulation in a male population; however, the mean score comparisons between the male and female groups are relatively worthless because of measurement non-equivalence across the items. Such issues are of key significance in cross-gender research and when examining potential intergroup differences (e.g., based on gender, ethnicity, or age) in psychological constructs measured through self-reporting (Little, 1997; Gregorich, 2006). In addition, comparisons of gender differences based on the ERQ or studies of the impact of emotional regulation strategy between different genders should be based on the measurement equivalence of the scale. When the study was based on the scale to conduct further research and found differences between different genders, one should first consider from the angle of exploring ERQ measurement equivalence between different gender groups, that is, the scale to participants of different genders was measured on the equivalence, only to make the equivalence scale further valuable. However, to date, no equivalence study based on this scale has been reported among different genders in Chinese cultural background, and this study is conducted on such considerations.

THIS STUDY

This study uses tests of model invariance to determine whether the scale illustrates consistent measurement characteristics across two specific demographic comparisons—male and female undergraduate participants. The normative data for these gender groups in an undergraduate sample are included to provide further information about how the questionnaire performs across

[1] https://spl.stanford.edu/sites/g/files/sbiybj9361/f/chinese.pdf

varying participant groups. It is hypothesized that this study will support the two-subscale structure illustrated in a previous research, and the measure will demonstrate invariance across gender comparison groups.

MATERIALS AND METHODS

Participants and Procedure

We enrolled junior and senior students from a university in Beijing. A total of 882 participants (47.01% males), aged 19–23 years, were enrolled [mean (M_{age}) = 21.31, standard deviation (SD) = 1.09]. The sample encompassed 93.42% of individuals who reported their ethnicity as Han, and a further 6.58% classified themselves as belonging to an ethnic minority. To control ordering effects, the order of questionnaire administration was counterbalanced in each study. All participants were given information outlining the purpose and possible drawbacks of participation before completing the measures, as well as the opportunity to decline participation if they desired. Participants completed all measures and returned the questionnaires to research assistants before leaving the classroom.

Measures

In this study, the ERQ comprised 10 items. It includes two dimensions—cognitive reappraisal factor (six items; items 1, 3, 5, 7, 8, and 10) and expression suppression factor (four items; items 2, 4, 6, and 9). The ERQ is primarily used to evaluate individual emotion regulation strategies. We used the Likert seven-point scoring method for the items. The higher the score, the higher the frequency of using emotion regulation strategy. The internal consistency (Cronbach's α) in this study was 0.825.

Statistical Analysis

Missing Data

The original sample included 882 Chinese college students; however, as 35 failed to respond to all ERQ items, they were excluded from the analysis. A total of 847 valid questionnaires (401 males and 446 females) were collected (effective rate: 96.03%).

Analytic Stages

Our analyses contained the following two stages: (i) CFA tested the fit of the emotional regulation model; and (ii) MIs of the emotional regulation model were assessed, from the CFA, across gender.

Stage 1: Model Evaluation in CFA

CFA was conducted for the Emotional Regulation model, and the CFA was specified and estimated using Mplus 8.0 software (Muthén and Muthén, 1998-2017). Based on previous studies, we used some fit indices to assess the overall fit of the models; these included chi-square (χ^2), comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). The values >0.90 for the CFI and TLI and <0.08 for the RMSEA and SRMR indicated an adequate fit (Kline, 2010).

Stage 2: Model Specification

Following the generally accepted practice, we assessed the fit of each model by examining multiple fit indices (Kline, 2010). When examining factorial invariance, we followed the established procedures (Meredith, 1993; Gregorich, 2006; Meredith and Teresi, 2006), which were used in the related literature (Engdahl et al., 2011; Wang et al., 2013a). If configural invariance (baseline model, Model A) is supported, further restrictive constraints could be imposed on the model, as was performed in the conventional multiple group CFA invariance test. First, factor loadings were constrained to be equal across gender to test metric or weak invariance (Model B). In addition, a χ^2 difference test was conducted to assess if the baseline model was significantly different from the constrained model. A non-significant χ^2 difference test indicated that factor loadings were invariant across gender, thereby satisfying metric invariance. Furthermore, based on the metric invariance model, intercepts were constrained to be equal across gender to build Model C, a test of scalar or strong invariance. Model D included the restrictions from Model C plus the additional constraint of equal item error variances across the two genders (invariant error variance or strict invariance). Subsequent to Model D, residual error variances were not constrained to be equal across timepoints (Grouzet et al., 2006). Thus, Model E was compared with Model C to preserve nested model testing. Model E comprised the constraints from Model C plus the additional constraint of equal factor variances across the two genders (invariant factor variances). During testing, except for the baseline model (Model A), the first two invariance testing analyses were also called MI, while the next invariance testing analyses were called structural invariance.

Data Analysis

Statistical analyses were performed using SPSS 19.0, JASP-0.11.1.0 ([2]; Marsman and Wagenmakers, 2017; Wagenmakers et al., 2017a; Wagenmakers et al., 2017a,b), and Mplus 8.0 (Muthén and Muthén, 1998-2017). JASP-0.11.1.0 software was primarily used to analyze the kurtosis and skewness of items. Using Mplus 8.0 software, we used the CFA of the ERQ, compared the fitting index, and obtained the best factor model to fit the Chinese college students. In addition, significant skewness and kurtosis values were obtained for each item ($p < 0.01$). We selected the robust maximum-likelihood estimation method for unbiased estimation of non-normal distribution data for data analysis (Satorra and Bentler, 2001). The robust ML estimator with a mean-adjusted χ^2 (maximum likelihood parameter estimates with standard errors and a mean-adjusted χ^2 test statistic) was selected, as these provide parameter estimates that are robust to non-normality (Satorra and Bentler, 2001; Wang et al., 2013a). Furthermore, we use the corrected scaled χ^2 difference test to compare the nested models (Satorra and Bentler, 2001).

We evaluated the fit of each model by examining multiple fit indices (Kline, 2010; Wang et al., 2012). We used the Satorra Bentler chi-square statistic (S-Bχ^2), RMSEA, SRMR, TLI, and CFI. On the basis of extensive simulation studies conducted

[2]https://jasp-stats.org/

by Hu and Bentler (1999), it appears that good-fitting models have CFI and TLI values greater than 0.95, RMSEA values less than 0.06, and less than 0.08 (Wang et al., 2012). The corrected scaled chi-square difference test developed by Satorra and Bentler (2001); Muthén and Muthén (1998-2017) was used to compare nested models. However, tests of the change in CFI (i.e., ΔCFI) are superior to chi-square ($\Delta\chi^2$) difference tests of invariance because they are not affected by the sample size (Cheung and Rensvold, 2002; Meade et al., 2008). Thus, the corrected scaled chi-square difference test and change in CFI were used to compare nested models. When both results contradict each other, however, we primarily depended on results of CFI differences.

According to the suggestion of Cheung and Rensvold (2002), the change in CFI was chosen to evaluate the measurement invariance. When $\Delta CFI < 0.01$, it implies that the invariance hypothesis cannot be rejected, and the model fits well; when $0.01 \leq \Delta CFI \leq 0.02$, it implies that the degree of the model has a moderate deterioration, which cannot reveal that the difference exists and is significant; when $\Delta CFI \geq 0.02$, it signifies a significant difference (Cheung and Rensvold, 2002; Meade et al., 2008; Wang et al., 2013b), and the standard of the nested model is $\Delta CFI < 0.01$, $\Delta TLI < 0.01$ (Wang et al., 2012, Wang et al., 2013b).

Ethics Statement

In this study, the core variables were participants' ERQ scores, and we collected the data in the classroom. Written informed consent was obtained from all principals and participants in this study. The protocol and questionnaires used were approved by the university's Institutional Review Board.

RESULTS

Descriptive Statistics

Table 1 lists the average scores measured by the ERQ and standardized factor loads for each item. Significant multivariate skewness and kurtosis were found ($p < 0.05$, based on univariate and multivariate tests). In the ERQ, the real score was 20–53 (male: 36.83 ± 6.118; female: 32.98 ± 5.732), and the male score was significantly higher than the female score ($t = 3.054$, $p < 0.01$, $d = 0.46$). In the cognitive reappraisal factor score, the male score was 16.02 ± 2.659, while the female score was 14.95 ± 2.802; thus, the male and female scores revealed no statistically significant difference ($t = 1.223$, $p = 0.171$). In the expression suppression factor score, the male score was 22.01 ± 3.754, while the female score was 18.65 ± 4.002; the male score was significantly higher than that of the females ($t = 3.124$, $p < 0.01$, $d = 0.42$). In this study, Cronbach α was 0.825 in the ERQ, and the coefficient α of cognitive reappraisal and expression suppression was 0.831 and 0.778, respectively.

Item analysis was used to discriminate each item (**Table 2**). (i) A critical ratio (decision values of the high- and low-score groups) was used and the correlation of the total items to test the discrimination of each item. We defined the first 27% of the score in the ERQ as the high-score group, while the latter 27% as the low-score group. (ii) Each item score difference in the high- and low-score groups was compared in this study. The results revealed that the ERQ scores in the high- and low-score groups were statistically significant, and the correlation of the total items were 0.38–0.62 ($p < 0.01$).

TABLE 1 | Descriptive statistics results of Emotion Regulation Questionnaire (ERQ).

Item		*M (SD)*	Skewness	Kurtosis	Factor load		*t*	*p*	*Cohen's d*
					CR	ES			
Cognitive reappraisal									
Item 1		3.76 (1.779)	−0.22	0.55	0.591**				
Item3		3.87 (1.754)	−0.21	0.61	0.698**				
Item 5		3.72 (1.782)	−0.25	0.37	0.657**				
Item 7		3.81 (1.791)	−0.25	0.39	0.589**				
Item 8		3.75 (1.802)	−0.19	0.48	0.592**				
Item 10		3.69 (1.793)	−0.13	0.61	0.563**				
Expression suppression									
Item 2		4.21 (1.901)	−0.41	0.89		0.631**			
Item 4		4.16 (1.330)	−0.53	0.87		0.602**			
Item 6		3.91 (1.324)	−0.47	0.89		0.603**			
Item 9		3.87 (1.135)	−0.55	0.91		0.594**			
Scores for different gender									
Total scores	*Males*	36.83 (6.118)					3.054**	<0.01	0.46
	Females	32.98 (5.732)							
CR scores	*Males*	16.02 (2.659)					1.223	0.171	—
	Females	14.95 (2.802)							
ES scores	*Males*	22.01 (3.754)					3.124**	<0.01	0.42
	Females	18.65 (4.002)							

***p < 0.01. CR, cognitive reappraisal; ES, expression suppression.*

TABLE 2 | The *t*-test of high- and low-score group for each item and total item correlation of ERQ.

Item	*t*	*p*	Total item correlation
Cognitive reappraisal			
Item 1	9.46	<0.001	0.530**
Item 3	12.77	<0.001	0.582**
Item 5	10.54	<0.001	0.577**
Item 7	11.39	<0.001	0.589**
Item 8	12.32	<0.001	0.621**
Item 10	11.55	<0.001	0.502**
Expression suppression			
Item 2	10.71	<0.001	0.501**
Item 4	8.92	<0.001	0.384**
Item 6	11.23	<0.001	0.522**
Item 9	10.05	<0.001	0.598**

$^{**}p < 0.01$.

Stage 1: Confirmatory Factor Analysis

The CFA results (**Figure 1**) revealed that S-B χ^2/df = 5.95, p = 0.004, CFI = 0.93, TLI = 0.92, RMSEA = 0.056, and SRMR = 0.038. Specifically, for males, the CFA results revealed that S-B χ^2/df = 3.49, p = 0.002, CFI = 0.94, TLI = 0.93, RMSEA = 0.043, and SRMR = 0.051. For females, the CFA results revealed that S-B χ^2/df = 3.66, p = 0.002, CFI = 0.95, TLI = 0.93, RMSEA = 0.059, and SRMR = 0.044 (**Table 3**).

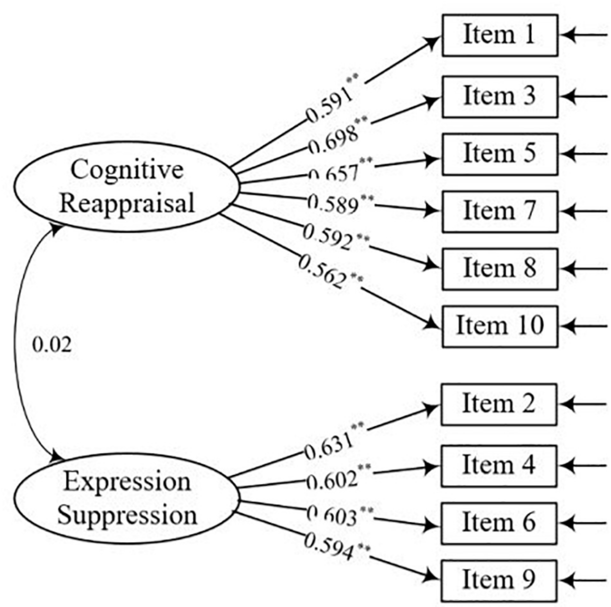

FIGURE 1 | Standardized factor loadings for total sample confirmatory factor analysis (CFA).

TABLE 3 | Two-factor structure model fitting results in ERQ.

	S-Bχ^2/df	CFI	TLI	RMSEA	SRMR
Total	5.95	0.934	0.929	0.056	0.038
Male	3.49	0.941	0.932	0.043	0.051
Female	3.66	0.945	0.934	0.059	0.044

ERQ, emotion regulation questionnaire.S-Bχ^2, Satorra–Bentler scaled χ^2; df, degrees of freedom; TLI, Tucker–Lewis index; CFI, comparative fit index; RMSEA, root-mean-square error of approximation; SRMR, standardized root mean squared residual.

Stage 2: Measurement Invariance Testing Across Gender

The results from the MI across gender revealed that all five steps of MI testing resulted in significant χ^2 (ps < 0.01), excellent (CFIs > 0.95, TLIs > 0.090), and equivalent fit indices (ΔCFIs < 0.01, ΔTLIs < 0.01). Moreover, all goodness-of-fit indices suggested that all models assuming different degrees of invariance were acceptable (**Table 4**).

Configural Invariance (Model A)

In the configural MI testing, the factor load and the intercept of observation variables were performed for free estimation. In this study, each fitting index of Model A fulfilled the measurement standard (CFI $\geq$ 0.90; TLI $\geq$ 0.90), thereby establishing the configural invariance, and Model A fulfilled the requirements as the next MI analysis baseline model (**Table 4**).

Metric Invariance (Model B)

After passing the configural invariance testing, the factor load MI was set according to Model A, and both groups of corresponding factor loads were constrained to be equal to test the weak invariance model. After increasing the factor load equal constrain, if the data fitting situation did not reach the standard in statistics, the constrain was not removed. In this study, comparing the CFIs and TLIs of Model B and Model A, the $|\Delta$CFI$|$ and $|\Delta$TLI$|$ values were 0 and 0.003. As shown in **Table 4**, the model fitted well, and the MI test continued.

Scalar Invariance (Model C)

Based on the construction of Model B, we set the measurement intercepts of two groups equally (Model C). As shown in **Table 4**, we compared the CFIs and TLIs of Model C and Model B, the $|\Delta$CFI$|$ and $|\Delta$TLI$|$ values were 0.003 and 0.001, and the model fitted well, thereby the MI test continued.

Residual Error Invariance (Model D)

Based on Model C, we constrained residual error variances across the groups. Then, we compared CFIs and TLIs values of Model D and Model C, the $|\Delta$CFI$|$ and $|\Delta$TLI$|$ values were 0.004 and 0.002. As shown in **Table 3**, the model fitted well, thereby the MI test continued.

Invariant Factor Variances (Model E)

The final test of this study was to test structural invariance (Model E), which additionally constrained factor variances and covariances (not residual variances), tested against Model C. As shown in **Table 4**, $|\Delta$CFI$|$ and $|\Delta$TLI$|$ values of the two models

TABLE 4 | Goodness-of-fit indices of the compared models.

MI model	S-Bχ^2	*df*	RMSEA [90%CI]	CFI	TLI	Model comparison	ΔCFI	ΔTLI
Model-A	443.589	68	0.051 [0.046, 0.058]	0.964	0.946		–	–
Model-B	457.902	76	0.049 [0.044, 0.056]	0.964	0.949	B vs. A	0	0.003
Model-C	488.271	84	0.050 [0.044, 0.057]	0.961	0.948	C vs. B	−0.003	−0.001
Model-D	513.678	94	0.045 [0.042, 0.055]	0.957	0.951	D vs. C	−0.004	0.002
Model-E	567.237	97	0.043 [0.040, 0.054]	0.953	0.952	E vs. D	−0.004	0.001

Model A indicates no parameters constrained to be equal across groups; model B, factor loadings constrained to be equal; model C, observed variable intercepts and factor loadings constrained to be equal; model D, residual variances, factor loadings, and observed variable intercepts constrained to be equal; model E, factor variances and covariances, factor loadings, and observed variable intercepts constrained to be equal. CI indicates confidence interval. df, degrees of freedom; TLI, Tucker–Lewis index; CFI, comparative fit index; RMSEA, root-mean-square error of approximation; SRMR, standardized root mean squared residual; S-Bχ^2, Satorra–Bentler scaled χ^2.

mentioned above were 0.004 and 0.001, respectively, implying that the factor variance MI was established.

DISCUSSION

This study first tested the two-factor structure of the emotion regulation using the CFA among Mainland China college students. The item analysis revealed that the distinction and discrimination of the items were acceptable, which is consistent with previous studies that used the CFA to compare alternative structures of emotion regulation among Chinese rural-to-urban migrant youth (Wang et al., 2020). The Cronbach's α of ERQ total scores and subscales was acceptable (0.778–0.831), suggesting that the ERQ is a reliable measure of emotion regulation. The CFA results supported the two-factor structure of the ERQ, which demonstrated a clear replication with the results of most previous studies (Wang et al., 2007; Matsumoto et al., 2008). The total internal consistency α coefficient of the ERQ was 0.825, and each dimension was 0.831 (cognitive reappraisal) and 0.778 (expressive suppression), which is acceptable. In addition, α coefficients of the ERQ were similar to that in previous studies in Chinese literature (cognitive reappraisal, $\alpha = 0.85$; expressive suppression, $\alpha = 0.77$) (Wang et al., 2007); however, α coefficients of the ERQ were marginally lower than that of the rural-to-urban migrant adolescents and young adults in China (the total internal consistency α coefficient of the ERQ was 0.82, and each dimension was 0.82 (cognitive reappraisal) and 0.73 (expressive suppression) (Wang et al., 2020); this could be attributable to different characteristics of different groups of people.

This study examined MI across gender and compared the gender difference of emotion regulation strategy based on the ERQ. The findings demonstrated that all models assuming different degrees of invariance were acceptable, suggesting that the ERQ factors have the same meaning across gender, suggesting that comparisons across gender based on the ERQ are meaningful. This study's results of MI across gender corroborated previous research, in which MI was found in a sample of American undergraduates (Melka et al., 2011). Furthermore, the results of this research extend the study area from the perspective of MI in Mainland China with Oriental cultural background.

Comparison of differences in ERQ scores and the two factors between males and females revealed that males' overall emotion regulation is markedly higher than females'. Regarding cognitive reappraisal factors, no significant difference was observed between males and females, whereas, a significant difference was observed between males and females in terms of expression suppression, suggesting that males exhibit more utilization of expression suppression strategies for emotion regulation than females. Notably, previous studies have compared the emotion regulation strategy of people from various backgrounds (Sala et al., 2012). However, as related to gender, if the MI does not hold across groups, differences in observed scores may not be directly comparable. This finding is consistent with previous studies on the differences in emotion regulation between males and females (Hess et al., 2000; Parkins, 2012; Chaplin and Aldao, 2013), and, thus, our results provide additional empirical support from Mainland China for their conclusion.

Our findings provide crucial meaning for practice. First, influenced by Chinese traditional culture, undergraduates in Mainland China are not good at expressing their emotions, which remind college administrators to be concerned about undergraduates, teach them emotion regulation strategies and interpersonal communication strategies, and provide them with opportunities to interact and practice emotion regulation strategies in their relationships, and specific educational schedules should be developed and used for this group. Second, gender differences depicted in ERQ measurement scores reflect the real differences in the cognitive reappraisal and expression suppression between males and females, rather than caused by the variance measured by the ERQ itself (Meredith and Teresi, 2006), thereby providing a comparative psychological basis for related research. Finally, it is significant that future emotion regulation measurement and invariance measurement criteria should consider this character.

This study has some limitations. First, we used a restricted sample of college students from Mainland China; thus, the results might not be entirely generalizable for all Chinese population. Second, the sample was not considered regarding other variables and, thus, was not further explained; however, it could serve as a basis for future research. Finally, we used a more appropriate parameter estimate approach (Flora and Curran, 2004; Melka et al., 2011).

CONCLUSION

This study establishes the ERQ as a structurally consistent and sound measure of cognitive reappraisal and emotional suppression across gender groups. Given the popularity of emotion regulation research in recent years, attempts to elucidate mea sures of associated constructs are vital. This study provides further evidence that the ERQ is a valuable research topic. Nonetheless, continued efforts to use the instrument in future studies are highly recommended.

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 Institutional Review Committee of the Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University. Participants provided their written informed consent to participate in the study.

AUTHOR CONTRIBUTIONS

YZ designed and executed the study, analyzed the data, and wrote the manuscript. YB collaborated with the design of the study. All authors contributed to the article and approved the submitted version.

REFERENCES

Australian Bureau of Statistics (2017). *2016 Census Data Summary: Cultural Diversity in Australia*. Canberra: Australian Bureau of Statistics.

Cai, L., Zhong, M., Zhu, X., Tang, Q., and Wang, Q. (2012). Relationship between cognitive emotion regulation strategies and depressive symptoms in patients with depression. *Chin. J. Clin. Psychol.* 20, 176–179. doi: 10.1016/j.paid.2005.12

Chaplin, T. M., and Aldao, A. (2013). Gender differences in emotion expression in children: a meta-analytic review. *Psychol. Bull.* 139, 735–765. doi: 10.1037/a0030737

Cheung, W. G., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Model. A Multidiscipl. J.* 9, 233–255. doi: 10.1207/S15328007SEM0902_5

Dunsmore, J. C., Booker, J. A., and Ollendick, T. H. (2013). Parental emotion coaching and child emotion regulation as protective factors for children with oppositional defiant disorder. *Soc. Dev.* 22, 444–446. doi: 10.1111/j.1467-9507.2011.00652.x

Eftekhari, A., Zoellner, L. A., and Vigil, S. A. (2009). Patterns of emotion regulation and psychopathology. *Anxiety Stress Cop.* 22, 571–586. doi: 10.1080/10615800802179860

Engdahl, R. M., Elhai, J. D., Richardson, J. D., and Frueh, B. C. (2011). Comparing posttraumatic stress disorder s symptom structure between deployed and non-deployed veterans. *Psychol. Assess.* 23, 1–6. doi: 10.1037/a0020045

English, T., John, O. P., Srivastava, S., and Gross, J. J. (2012). Emotion regulation and peer-rated social functioning: a 4-year longitudinal study. *J. Res. Pers.* 46, 780–784. doi: 10.1016/j.jrp.2012.09.006

Flora, D. B., and Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol. Methods* 9, 466–491. doi: 10.1037/1082-989x.9.4.466

Goldin, P., Ziv, M., and Jazaieri, H. (2012). Cognitive reappraisal self-efficacy mediates the effects of individual cognitive-behavioral therapy for social anxiety disorder. *J. Consult. Clin. Psychol.* 80, 1034–1040. doi: 10.1037/a0028555

Gratz, K. L., Weiss, N. H., and Tull, M. T. (2015). Examining emotion regulation as an outcome, mechanism, or target of psychological treatments. *Curr. Opin. Psychol.* 3, 85–90. doi: 10.1016/j.copsyc.2015.02.010

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Med. Care* 44, 578–594. doi: 10.1097/01.mlr.0000245454.12228.8f

Gross, J. J. (1998). The emerging field of emotion regulation: an integrative review. *Rev. Gen. Psychol.* 2, 271–299. doi: 10.1037/1089-2680.2.3.271

Gross, J. J. (2002). Emotion regulation: affective, cognitive, and social consequences. *Psychophysiology* 39, 281–291. doi: 10.1017/s0048577201393198

Gross, J. J., and John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and wellbeing. *J. Pers. Soc. Psychol.* 85, 348–362. doi: 10.1037/0022-3514.85.2.348

Gross, J. J., Richards, J. M., and John, O. P. (2006). "Emotion regulation in everyday life," in *Emotion Regulation in Couples and Families: Pathways to Dysfunction and Health*, eds D. K. Snyder, J. Simpson, and J. N. Hughes (Washington, DC: American Psychological Association), 13–35. doi: 10.1037/11468-001

Grouzet, F. M. E., Kasser, T., Ahuvia, A., Dols, J. M. F., Kim, Y., Lau, S., et al. (2006). The structure of goal contents across 15 cultures. *J. Pers. Soc. Psychol.* 89, 800-816. doi: 10.1037/0022-3514.89.5.800

Hess, U., Senécal, S., Kirouac, G., Herrera, P., Philippot, P., and Kleck, R. E. (2000). Emotional expressivity in men and women: stereotypes and self-perceptions. *Cogn. Emot.* 14, 609–642. doi: 10.1080/02699930050117648

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

John, O. P., and Gross, J. J. (2004). Healthy and unhealthy emotion regulation: personality processes, individual differences, and life span development. *J. Pers.* 72, 1301–1334. doi: 10.1111/j.1467-6494.2004.00298.x

John, O. P., and Gross, J. J. (2007). "Individual differences in emotion regulation strategies: links to lobal trait, dynamic, and social cognitive constructs," in *Handbook of Emotion Regulation*, ed. J. J. Gross (New York, NY: Guilford Press), 351–372.

Joormann, J., and Gotlib, I. H. (2010). Emotion regulation in depression: relation to cognitive inhibition. *Cogn. Emot.* 24, 281–298. doi: 10.1080/02699930903407948

Kline, R. B. (2010). *Principles and Practice of Structural Equation Modeling*, 3rd Edn, New York, NY: Guilford Press.

Little, T. D. (1997). Mean and covariance structure (MACS) analyses of cross-culture data: practical and theoretical issues. *Multiv. Behav. Res.* 32, 53–76. doi: 10.1207/s15327906mbr3201_3

Liu, W., Chen, L., and Tu, X. (2017). Chinese adaptation of emotion regulation questionnaire for children and adolescents (ERQ-CCA): a psychometric evaluation in Chinese children. *Intern. J. Psychol.* 52, 398–405. doi: 10.1002/ijop.12233

Lotfi, M., Bahrampouri, L., Amini, M., Fatemitabar, R., Birashk, B., Entezari, M., et al. (2019). Persian adaptation of emotion regulation questionnaire for children and adolescents (ERQ-CA). *J. Mazand. Univ. Med. Sci.* 29, 117–128.

Marsman, M., and Wagenmakers, E.-J. (2017). Bayesian benefits with JASP. *Eur. J. Dev. Psychol.* 14, 545–555. doi: 10.1080/17405629.2016.1259614

Matsumoto, D., Yoo, S. H., and Nakagawa, S. (2008). Culture, emotion regulation, and adjustment. *J. Pers. Soc. Psychol.* 94, 925–937. doi: 10.1037/0022-351494.6.925

McRae, K., Jacobs, S. E., Ray, R. D., John, O. P., and Gross, J. J. (2012). Individual differences in reappraisal ability: links to reappraisal frequency, well-being, and cognitive control. *J. Res. Pers.* 46, 2–7. doi: 10.1016/j.jrp.2011.10.003

Meade, A. W., Johnson, E. C., and Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *J. Appl. Psychol.* 93, 568–592. doi: 10.1037/0021-9010.93.3.568

Melka, S. E., Lancaster, S. L., Bryant, A. R., and Rodriguez, B. F. (2011). Confirmatory factor and measurement invariance analyses of the emotion regulation questionnaire. *J. Clin. Psychol.* 67, 1283–1293. doi: 10.1002/jclp.20836

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika* 58, 525–543. doi: 10.1007/BF02294825

Meredith, W., and Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Med. Care* 44, 69–77. doi: 10.1097/01.mlr.0000245438.73837.89

Miao, Y. J. (2009). *Happiness in the Perspective of Psychology-The Theory of Happiness and Its Evaluation*. Tianjin: Tianjin People's Press.

Minkel, J. D., McNealy, K., and Gianaros, P. J. (2012). Sleep quality and neural circuit function supporting emotion regulation. *Biol. Mood Anx. Disord.* 2, 1–9.

Moore, S. A., Zoellner, L. A., and Mollenholt, N. (2008). Are expressive suppression and cognitive reappraisal associated with stress-related symptoms? *Behav. Res. Ther.* 46, 993–1000. doi: 10.1016/j.brat.2008.05.001

Muthén, L. K, Muthén, B. O(1998-2017). *Mplus User S Guide*, 8th Edn, Los Angeles, CA: Muthén & Muthén.

Ochsner, K. N., and Gross, J. J. (2008). Cognitive emotion regulation: Insights from social cognitive and affective neuroscience. *Curr. Direct. Psychol. Sci.* 17, 153–158. doi: 10.1111/j.1467-8721.2008.00566.x

Parkins, R. (2012). Gender and emotional expressiveness: an analysis of prosodic features in emotional expression. *Griffith Work. Pap. Pragmat. Interc. Commun.* 5, 46–54.

Parkinson, B., and Totterdell, P. (1999). Classifying affect-regulation strategies. *Cogn. Emot.* 13, 277–303. doi: 10.1080/026999399379285

Pastor, M. C., López-Penadés, R., Cifre, E., and Moliner-Urdiales, D. (2019). The spanish version of the emotion regulation questionnaire for children and Adolescents (ERQ-CA): a psychometric evaluation in early adolescence. *Span. J. Psychol.* 22:30. doi: 10.1017/sjp.2019.30

Preece, D. A., Becerra, R., Robinson, K., and Gross, J. J. (2019). The emotion regulation questionnaire: psychometric properties in general community samples. *J. Pers. Assess.* 102, 1–9. doi: 10.1080/00223891.2018.1564319

Sala, M. N., Molina, P., Abler, B., Kessler, H., Vanbrabant, L., and van de Schoot, R. (2012). Measurement invariance of the emotion regulation questionnaire (ERQ). a cross-national validity study. *Eur. J. Dev. Psychol.* 9, 751–757. doi: 10.1080/17405629.2012.690604

Satorra, A., and Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 66, 507–514. doi: 10.1007/BF02296192

Shi, L., Guang, P., and Luo, X. (2019). A network analysis of the influence of parental anxiety and its emotional regulation on adolescent anxiety. *Chin. J. Clin. Psychol.* 27, 1237–1241.

Soto, J. A., Armenta, B. E., Perez, C. R., Zamboanga, B. L., UmañaTaylor, A. J., Lee, R. M., et al. (2012). Strength in numbers? Cognitive reappraisal tendencies and psychological functioning among Latinos in the context of oppression. *Cult. Diver. Ethnic Minor. Psychol.* 18, 384–394. doi: 10.1037/a0029781

Spaapen, D., Waters, F., Brummer, L., Stopa, L., and Bucks, R. (2014). The emotion regulation questionnaire: validation of the ERQ-9 in two community samples. *Psychol. Assess.* 26, 46–54. doi: 10.1037/a0034474

Su, J. C., Lee, R. M., Park, I. J. K., Soto, J. A., Chang, J., Zamboanga, B. L., et al. (2015). Differential links between expressive suppression and well-being among Chinese and Mexican American college students. *Asian Am. J. Psychol.* 6, 15–24. doi: 10.1037/a0036116

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., et al. (2017a). Bayesian inference for psychology. Part II: example applications with JASP. *Psychon. Bull. Rev.* 25, 58–76. doi: 10.3758/s13423-017-1323-7

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., et al. (2017b). Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. *Psychon. Bull. Rev.* 25, 35–57. doi: 10.3758/s13423-017-1343-3

Wang, D., Yuan, B., Han, H., and Wang, C. (2020). Validity and reliability of emotion regulation questionnaire (ERQ) in Chinese rural-to-urban migrant adolescents and young adults. *Curr. Psychol.* 4, 1–8. doi: 10.1007/s12144-020-00754-9

Wang, L., Liu, H. C., Li, Z. Q., and Du, W. (2007). Reliability and validity of emotion regulation questionnaire Chinese revised version. *Chin. J. Health Psychol.* 15, 503–505.

Wang, M., Armour, C., Li, X., Dai, X., Zhu, X., and Yao, S. (2013a). The factorial invariance across gender of three well-supported models. *J. Nerv. Ment. Dis.* 201, 145–152. doi: 10.1097/nmd.0b013e31827f627d

Wang, M., Armour, C., Wu, Y., Ren, F., Zhu, X., and Yao, S. (2013b). Factor structure of the CES-D and measurement invariance across gender in mainland chinese adolescents. *J. Clin. Psychol.* 69, 966–979. doi: 10.1002/jclp.21978

Wang, M., Elhai, J. D., Dai, X., and Yao, S. (2012). Longitudinal invariance of posttraumatic stress disorder symptoms in adolescent earthquake survivors. *J. Anx. Disord.* 26, 263–270. doi: 10.1016/j.janxdis.2011.12.009

Wiltink, J., Glaesmer, H., Canterino, M., Wölfling, K., Knebel, A., Kessler, H., et al. (2011). Regulation of emotions in the community: suppression and reappraisal strategies and its psychometric properties. *GMS Psycho. Soc. Med.* 8:78.

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 in Psychology

ORIGINAL RESEARCH
published: 03 December 2020
doi: 10.3389/fpsyg.2020.595651

Psychometric Evaluation of the Czech Version of Group Cohesiveness Scale (GCS) in a Clinical Sample: A Two-Dimensional Model

Adam Klocek†, Tomáš Řiháček*† and Hynek Cígler†*

Department of Psychology, Faculty of Social Studies, Brno, Czechia

OPEN ACCESS

Edited by:
Susana Sanduvete-Chaves, Seville University, Spain

Reviewed by:
Sai-fu Fung, City University of Hong Kong, Hong Kong
Giorgia Molinengo, University of Turin, Italy

***Correspondence:**
Adam Klocek
klocek.adam@mail.muni.cz;
aklocek.ak@gmail.com
Tomáš Řiháček
rihacek@fss.muni.cz

†ORCID:
Adam Klocek
orcid.org/0000-0002-0797-4890
Tomáš Řiháček
orcid.org/0000-0001-5893-9289
Hynek Cígler
orcid.org/0000-0001-9959-6227

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

Received: 17 August 2020
Accepted: 11 November 2020
Published: 03 December 2020

Citation:
Klocek A, Řiháček T and Cígler H (2020) Psychometric Evaluation of the Czech Version of Group Cohesiveness Scale (GCS) in a Clinical Sample: A Two-Dimensional Model. Front. Psychol. 11:595651. doi: 10.3389/fpsyg.2020.595651

The Group Cohesiveness Scale (GCS, 7 items) measures patient-rated group cohesiveness. The English version of the scale has demonstrated good psychometric properties. This study describes the validation of the Czech version of the GCS. A total of 369 patients participated in the study. Unlike the original study, the ordinal confirmatory factor analysis (CFA) supported a two-dimensional solution (RMSEA = 0.075; TLI = 0.986). The analysis demonstrated the existence of two moderately to highly associated (r = 0.79) domains of group cohesiveness—affective and behavioral. The two-dimensional model was invariant across genders, age, education, and time (retest after 6 weeks) up to factor means level. Internal consistency reached satisfactory values for both domains (affective, ω = 0.86; behavioral, ω = 0.81). In terms of convergent validity, only weak association was found between the GCS domains and the group working alliance measured by the Group Outcome Rating Scale (GSRS). This is the first revision of the factor structure of the GCS in the European context. The scale showed that the Czech version of the GCS is a valid and reliable brief tool for measuring both aspects of group cohesiveness.

Keywords: confirmatory factor analysis, group cohesion, Group Cohesiveness Scale, Czech validation study, affective and behavioral group cohesion

INTRODUCTION

Group cohesion is one of the elemental group phenomena that allows other therapeutic processes to occur within the group therapy framework. It is defined as the ability of the members of a group to tolerate negative emotions and self-disclosure (Wongpakaran et al., 2013). Group cohesion partially overlaps with other group phenomena, such as the working alliance and empathy (Johnson et al., 2005). Group cohesion is conceptually akin to the working alliance in individual therapy. Although it is primarily based on the relationships among the group members, it can also be extended to the relationship with the therapists (Budman et al., 1989). Group cohesion is also related to empathy because a cohesive group demands that its members have an understanding of others' feelings and experiences and can effectively express this understanding (Roark and Sharah, 1989).

Until recently, group phenomena and processes were measured by measures such as the Group Climate Questionnaire (MacKenzie, 1983), the Therapeutic Factors Inventory (Lese and MacNair-Semands, 2000), and the Working Alliance Inventory (Horvath and Greenberg, 1989). However, these scales were too lengthy to be used in routine care or rapid hospital environments (compared to research) and were not directly focused on group cohesion. Therefore, the Group Cohesiveness Scale (GCS[1]) was developed (Wongpakaran et al., 2013).

The GCS (Wongpakaran et al., 2013; see **Table 1**) was created from an original pool of 40 items and reduced to seven items representing two domains: cohesion and engagement. The former domain was represented by two items from the Therapeutic Factors Inventory, while the latter was represented by five items from the Group Climate Questionnaire. However, since both domains were similar in content, Wongpakaran et al. (2013) considered them to be representations of the unidimensional group cohesiveness construct.

Alternatively, Wongpakaran et al. (2013) suggested that the GCS items can be differentiated into the affective (items 1, 2, and 3) and behavioral (items 4, 5, 6, and 7) components of group cohesiveness. They argued that these components might be related to each other in a fashion similar to the unidimensional construct of depression, in which the feeling of sadness is functionally different from a behavioral lack of interest, yet both components measure the same latent construct of depression (Wongpakaran et al., 2013).

The distinction between the affective and behavioral components is consistent with the theoretical literature. According to Carron (1982), group cohesion is a "dynamic process that is reflected in the tendency for a group *to stick together* [emphasis added] and remain united in the *pursuit of its goals and objectives.* [emphasis added]" (p. 124). Similarly, Mudrack (1989) divided group cohesion into attraction-to-group (affective component) and commitment to the group task (behavioral component).

Originally, the GCS was standardized in the Thai language (Wongpakaran et al., 2013) in a clinical sample of 96 patients (56% women) with a mean age of 28.22 (SD = 6.84). Patients were hospitalized for up to 2 weeks. A principal component analysis revealed a unidimensional factor structure (57.2% of explained variance). Based on a confirmatory factor analysis (CFA) conducted on the same dataset, the authors claimed that the unidimensional model had moderately acceptable fit despite unsatisfactory RMSEA values ($\chi^2(14) = 32.29$; CFI = 0.94; TLI = 0.90; SRMR = 0.04; RMSEA = 0.12).

Although Wongpakaran et al. (2013) tried to fit a two-dimensional model (i.e., cohesion and engagement), they did not report the results, arguing that the two dimensions were too strongly correlated to be set apart ($r = 0.83$). Instead, they fine-tuned the unidimensional model based on modification indices by allowing residual correlations between pairs of items (items 1 and 2; items 2 and 3), reaching an excellent fit [$\chi^2(12) = 12.41$; CFI = 0.99; TLI = 0.99; SRMR = 0.04; RMSEA = 0.02]. Arguably, by allowing the residual correlations, the authors developed a model that was very similar to (but less parsimonious than) the suggested two-factor model with the affective and behavioral factors. Therefore, we found it desirable to formally test this alternative two-factor model as well. In terms of convergent validity, the GCS was correlated to the Group Benefit Questionnaire ($r = 0.71$, $p < 0.001$) and to the Cohesion to Therapist Scale ($r = 0.77$, $p < 0.001$) in the original study.

The GCS is a relatively new measure that has been employed in a limited number of studies thus far. Psychometric information about the GCS is rather scarce and often unsatisfactory given small sample sizes. Poyner-Del Vento et al. (2018) used the GCS as a measure of group cohesion in a pilot study in a sample of seven female military veterans. They found that removing item 6 ("*The members appear to do things the way they think will be acceptable to the group*") increased the internal consistency of the scale from $\alpha = 0.72$–0.90. Tulin et al. (2018) used the GCS to measure group cohesion in a sample of 109 students with internal consistency of $\alpha = 0.90$. In another sample of 22 students, Ashby et al. (2018) found a mean interitem correlation of $r = 0.43$. This limited evidence does not allow us to thoroughly evaluate the GCS, and the applicability of the measure in Western culture is still missing.

This study aimed to validate the Czech version of the GCS using the ordinal CFA paradigm. Four models were tested, including the unidimensional model (model 1), the unidimensional model with residual covariances between items 1 and 2 and items 2 and 3 allowed (model 2), a two-factor model with the factors of cohesion (items 1 and 2 originally extracted from the Therapeutic Factors Inventory) and engagement (items 3–7 originally extracted from the Group Climate Questionnaire) (model 3), and a two-factor model with affective (items 1–3)

[1]The GCS used in this study is unrelated to the Harvard Community Health Plan Group Cohesiveness Scale (which is also referred to as GCS in the literature; Budman et al., 1993).

TABLE 1 | Group Cohesiveness Scale (Wongpakaran et al., 2013).

Item no.	Item wording (*Czech in italics*)	Cohesiveness (C) or engagement (E) domain	Affective (A) or behavioral (B) domain
1	I feel accepted by the group. (*Cítím se být skupinou přijímaný/á*.)	C	A
2	In my group, we trust each other. (*Ve skupině si vzájemně důvěřujeme*.)	C	A
3	The members like and care about each other. (*Členové skupiny se mají rádi a vzájemně jim na sobě záleží*.)	E	A
4	The members try to understand why they do the things they do; they try to reason it out. (*Členové se snaží porozumět tomu, proč dělají věci, které dělají; snaží se na to přijít*.)	E	B
5	The members feel a sense of participation. (*Členové skupiny cítí, že se podílejí na chodu skupiny*.)	E	B
6	The members appear to do things the way they think will be acceptable to the group. (*Vypadá to, že členové dělají věci způsobem, o němž si myslí, že bude pro skupinu přijatelný*.)	E	B
7	The members reveal sensitive personal information or feelings. (*Členové si sdělují citlivé osobní informace a pocity*.)	E	B

and behavioral factors (items 4–7) (model 4). Furthermore, to assess the convergent validity, we used the Group Session Rating Scale (GSRS, Quirk et al., 2013), a measure of the group working alliance, as a comparison. Although group cohesion and group working alliance are distinct constructs, we expected the GCS scores to be related to the GSRS scores because both instruments measure non-specific group-based relational factors of the therapeutic process.

MATERIALS AND METHODS

Sample and Procedure

The sample included patients from seven clinical sites in the Czech Republic who provided informed consent to participate in research tracking the mechanisms of change during psychotherapy from January 2018 to December 2019. All patients underwent group therapy lasting from 4 to 12 weeks (depending on the site, median of 6 weeks). Data were collected on a paper-and-pencil form on a weekly basis during the whole treatment. Participants completed a battery of questionnaires regarding demographic variables, several outcome variables and several mechanisms of change, including group cohesion and working alliance. The study was approved by the Research Ethics Committee of Masaryk University (Ref. No. EKV-2017-029-R1).

In this study, the dataset used to validate the GCS included data from the second week of therapy (i.e., the first measurement of the group cohesion). Out of 448 patients who provided their baseline data, 380 patients (85%) participated in the second week of treatment. Out of 380 participants, 11 were characterized by missing data regarding the GCS, resulting in a total sample size of $N = 369$ patients. Differences between participants with missing data ($n = 80$) and the final sample ($n = 369$) in the demographics and clinical diagnosis data were investigated using t-tests and χ^2-tests.

Group Therapy

The treatment was integrative with major psychodynamic and minor humanistic and experiential aspects, supplemented with art, physical activity, music, ergo-, drama-, physio-, and bibliotherapy, relaxation and cognitive training, and community meetings[2]. Five sites were characterized by a frequency of five sessions of psychotherapy per week. The remaining two sites had three and four sessions per week, respectively. A session of group therapy lasted 90 min[3].

The sample comprised small closed groups of inpatients within four clinical sites and small open groups of outpatients in a program with a daycare basis within three clinical sites. Twenty-five (16 female) therapists participated in this research ($M_{\text{age}} = 44.13$ years, $SD_{\text{age}} = 10.29$). They were trained in the psychodynamic or psychoanalytic approach ($n = 15$), gestalt ($n = 4$), person-centered approach ($n = 3$), integrative approach ($n = 2$) or Daseinanalysis ($n = 1$). Their experience fluctuated between 1 and 25 years ($M = 12.21$, $SD = 7.30$).

[2]The supplemental therapeutic techniques and session differed by site.

[3]One site was characterized by the session length of 75 min.

Instruments

Group Cohesiveness Scale (GCS)

The seven items of the GCS are scored on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). None of the items is negatively worded. A higher score indicates higher perceived group cohesion. In the original study, the GCS yielded an average score of 4.73 out of 5 ($SD = 0.62$), the internal consistency of the whole scale was $\alpha = 0.87$, and the item-total correlations ranged from 0.497 to 0.752.

The scale was translated into Czech from the English version. Five native Czech speakers (a psychology student, two psychologists, and two laypeople) created five independent Czech translations. A group of three people (the two psychologists and the psychology student) then discussed all the translations and consolidated them into a single version. Third, this version was back-translated into English by a bilingual, native English speaker and compared to the original English version. Fourth, the final Czech version was field-tested with five respondents to check the comprehensibility of the items.

Group Session Rating Scale (GSRS)

The GSRS (Quirk et al., 2013) is a measure of the working alliance in group psychotherapy. It includes four 10-cm-long visual analog scales, each framed by a verbal anchor on both ends. The continuous dimension of each item is framed by bipolar points, and participants rate the group working alliance by making a mark on each scale. The response is measured as the length of the line from the left-hand side to the mark in millimeters. The range of the total score, computed as the sum of all items, can thus reach values between 0 and 400. A higher score indicates a better perceived working alliance. The scale was reported to be unidimensional, and the internal consistency ranged from $\alpha = 0.86$ to 0.90 in the original study.

Data Analysis

Software and General Settings

The statistical procedures were performed using statistical software R, version 4.0.2 (R Core Team, 2020). The significance level was set at $p < 0.05$.

Factorial Validity

The factor structure was estimated through ordinal CFA using the lavaan package (Rosseel, 2012). The ordinal factor analysis is equivalent to the two-parameter logistic graded response model in item response theory. Hence, this approach is not as vulnerable to the violation of assumptions as the standard factor analysis (Raykov and Marcoulides, 2011). Each item has five parameters (one slope and four thresholds between all neighboring response options). All five models were estimated using the stochastic weighted least squares means and variance adjusted estimator method (WLSMV), which seems to perform well with ordered categorical data (Raykov and Marcoulides, 2011). The fit indices employed in this study included χ^2, χ^2/df, root mean square error of approximation (RMSEA), Tucker-Lewis index (TLI), comparative fit index (CFI), and standardized root mean residual (SRMR). According to Hu and Bentler's (1999) and Hooper et al. (2008) evaluation criteria, the χ^2/df should not exceed 3, the

RMSEA should optimally be below 0.05, but values up to 0.10 are still considered to indicate a satisfactory fit. The SRMR should not exceed 0.08. Optimally, the TLI and CFI should be above 0.95; nevertheless, values above 0.90 are still considered to indicate a satisfactory fit.

Within models with more than one dimension (models 3 and 4), factors were allowed to be correlated. Since model 3 contained a factor represented only by two items, these items were constrained to load equally on their factor. Otherwise, models were identified by standardizing the latent variable. The internal consistency was estimated using bootstrapped Cronbach's alpha and McDonald's omega coefficients (McDonald, 1999). In terms of convergent validity, the association between the GCS and the GSRS was tested on the level of latent scores.

Measurement Invariance

The invariance was tested with regard to age, gender, education, and time. Measurement invariance was assessed by testing differences between nested models with continually increasing constraints: configural, metric (factor loadings), scalar (intercepts), strict (residuals), and factor means. Age groups were created by dividing the sample according to a median split. Gender invariance was assessed between male and female participants. Education invariance was assessed between higher (university, high technical school) and lower education (primary and secondary school with or without graduation) levels. Time invariance was assessed between the second and sixth weeks of group therapy (the sixth week was chosen pragmatically because in most sites, the therapy lasted only 6 weeks). We used four different fit indices to test the invariance, namely, $\Delta\chi^2$, ΔCFI, ΔSRMR, and ΔRMSEA. We employed "theta" parametrization and invariance guidelines with regard to ordinal data according to Wu and Estabrook (2016). Two groups are considered to be invariant if the item parameters (i.e., factor loadings, thresholds, intercepts, residuals, and factor means) are similar across groups.

Items 3 and 7 demonstrated missing response frequency at response option 1 (i.e., 1 or "strongly disagree"). The remaining items demonstrated near-to-missing response frequency (0.01) at response option 1. Response option 2 (unnamed) was also very seldom selected by the participants in all items. Therefore, all items were recoded into three categories (i.e., responses from 1 to 3 were recoded as a single category, representing a low level of group cohesion) for the purpose of testing the measurement invariance.

RESULTS

Missing Data

No significant differences between the final sample (N = 369) and the respondents with missing responses or respondents not participating in the study at the second week (n = 80, who were the remaining part of the initial sample of 449 participants) were found for the mean age, gender, education, and psychiatric diagnosis. The pattern of missingness could be considered missing at random. Therefore, only complete cases were included in the analyses.

Descriptive Characteristics

The total sample included 369 patients (73.7% females). Their nationality included Czech (95%), Slovak (2%), and others (3%). The patients' ages ranged from 18 to 71 years (M_{age} = 39.6, SD = 11.1). Psychiatric diagnoses were represented as follows: F4x (n = 261), F3x (n = 69), F6x (n = 53), F5x (n = 8), and F1x (n = 7). Several participants possessed multiple diagnoses (n = 33), mainly a combination of F4x and F6x (n = 13), F3x and F4x (n = 9), and F3x and F6x (n = 7). The remaining demographic variables are reported in **Table 2**.

The mean scores for each GCS item, the GCS total score, and the GSRS total score, as well as the internal consistency of the unidimensional model, are reported in **Table 3**. The average total score was 3.7 (SD = 0.69). Corrected item-total correlations ranged from 0.49 to 0.75.

Factor Structure

First, the assumptions of factor analysis were tested. The data did not show multivariate normality, and the standardized residuals were positively skewed. Homoscedasticity was not observed. After the preliminary data analyses, an ordinal factor analysis was employed to estimate the fit of the factor models using these skewed non-linear data. The RMSEA of the null model was 0.398. This value is above 0.148; thus, the TLI fit index could be interpreted (Kenny et al., 2015).

TABLE 2 | Descriptive characteristics of the sample (N = 369).

Variable	Categories	*n*	Percent
Gender	Female	272	74%
	Male	90	24%
	Missing	7	2%
Household	In partnership	189	51%
	Single	71	19%
	With parents	39	11%
	Other	62	17%
	Missing	8	2%
Marital status	Single	178	48%
	Married	114	31%
	Divorced	67	18%
	Widowed	2	1%
	Missing	8	2%
Education	Primary school	17	5%
	Secondary school	180	49%
	High technical school	22	6%
	University	141	38%
	Missing	9	2%
Occupation	Employee	163	44%
	Unemployed	53	14%
	Invalidity pension	35	10%
	Entrepreneur	23	6%
	Student	20	6%
	Maternity leave	7	2%
	Retirement	4	1%
	Other	15	4%
	Missing	49	13%

TABLE 3 | Descriptive characteristics of scales (*N* = 369).

Item	*M*	*SD*	Range (min-max)	Skewness	Kurtosis	Corrected item-total correlation	Cronbach's alpha if item deleted
GCS 1	3.70	0.91	4 (1 – 5)	−0.01	−0.64	0.63	0.85
GCS 2	3.72	0.94	4 (1 – 5)	−0.04	−0.87	0.75	0.84
GCS 3	3.51	0.92	3 (2 – 5)	0.21	−0.85	0.69	0.85
GCS 4	3.81	0.94	4 (1 – 5)	−0.27	−0.64	0.69	0.85
GCS 5	3.68	0.86	4 (1 – 5)	−0.01	−0.53	0.69	0.85
GCS 6	3.57	0.93	4 (1 – 5)	0.00	−0.44	0.59	0.86
GCS 7	4.00	0.93	3 (2 – 5)	−0.36	−0.97	0.49	0.87
	M	***SD***	**Range**	**Skewness**	**Kurtosis**	**McDonald's omega**	**Cronbach's alpha**
GCS total	25.99	4.82	22 (13 – 35)	0.20	−0.83	0.91	0.87 [0.85–0.89]
GSRS	290.82	75.20	368 (32 – 400)	−0.66	0.09	0.83	0.82 [0.79–0.85]

Second, four different factor solutions were tested for fit and compared (see **Table 4**). We concluded that the best fit was obtained by model 4, a two-factor solution with the affective and behavioral factors (see **Table 5** and **Figure 1**). Model 4 fit the data significantly better than did model 1 [unidimensional; $\Delta\chi^2(1) = 87.66$, $p < 0.0001$] and model 3 [two-factor with the cohesion and engagement factors; $\Delta\chi^2(2) = 104.31$, $p < 0.0001$]. Furthermore, the fit of model 4 did not significantly differ from that of model 2 [unidimensional with residual correlations; $\Delta\chi^2(1) = 2.35$, $p > 0.10$]. However, model 4 can be considered superior in terms of parsimony as well as theoretical justification. While the affective factor represents the same underlying structure as the empirically derived residual correlations in model 2, it explains the item interrelationships more efficiently and is consistent with theoretical expectations (Carron, 1982; Mudrack, 1989).

Measurement Invariance

Measurement invariance was assessed for model 4 with respect to age, gender, and education (see **Table 6**). Several patients were lost due to missing responses on the demographic variables, namely, age (n = 9), gender (n = 7), and education (n = 9). Measurement invariance between the younger (n = 185) and older (n = 175) cohorts was reached on the configural, metric, scalar, factor mean, and residual levels. Even though the χ^2-test was significant on the scalar and residual invariance level, other Δfit indices showed desirable values. Measurement invariance between women (n = 90) and men (n = 272) was reached on the configural, metric, scalar, and factor mean levels. Genders were not invariant only on the level of residual variances. Measurement invariance between lower (n = 197) and higher education levels (n = 163) was reached on the configural level. Even though the χ^2-test was significant on both metric and scalar invariance levels, other Δfit indices showed desirable values, and the fit even increased with more restricted models. We could, therefore, consider the model invariant between education levels on the configural, metric, scalar, factor mean, and residual variance levels. Measurement invariance between the second (n = 369) and sixth weeks (n = 273) was reached on the configural level. Even though the χ^2-test was significant on both the metric and scalar invariance levels, other Δfit indices showed desirable values, and the fit even increased with more restricted models. We could, therefore, consider the model invariant in time on the configural, metric, scalar, and factor mean levels. The final model was non-invariant only on the level of residual variances between the second and sixth weeks of measurement.

Reliability and Convergent Validity

The internal consistency of the final model was ω = 0.86 for the affective and ω = 0.81 for the behavioral domains (see **Table 5**). Additionally, the internal consistency of the general factor in model 1 was ω = 0.91. None of the GCS items would increase the internal consistency when dropped.

Thirteen participants had missing data on the GSRS scale, resulting in 367 patients. With respect to the final two-factor model with affective and behavioral dimensions (model 4), the affective domain was correlated more strongly with the GSRS (r = 0.449, $p < 0.05$) than the behavioral domain was (r = 0.290, $p < 0.05$). Additionally, a small to moderate positive correlation between the latent constructs of the unidimensional GCS (Model 1) and GSRS scales was found (r = 0.394, $p < 0.05$).

DISCUSSION

The present study described the validation of the Czech version of the Group Cohesiveness Scale (GCS). The average item scores and reliability were compatible with those of the original Thai version (Wongpakaran et al., 2013). However, we concluded that, based on a CFA, the most preferable model was a two-factor solution with the correlated affective and behavioral domains (model 4). This solution is more parsimonious than the fine-tuned unidimensional solution (model 2) suggested by Wongpakaran et al. (2013).

The final model demonstrated excellent fit and was invariant across age groups, genders, education levels, and time. The Czech version did not even show any problematic functioning of item 6 as presented in the English translation by Poyner-Del Vento et al. (2018). Theoretically, group cohesion is related to the working alliance (Johnson et al., 2005). However, in our study, we found only small to medium correlations between the GCS subscales and the GSRS. This finding was unexpected,

TABLE 4 | Fit indices of the tested models (N = 369).

Model	χ^2	df	χ^2/df	TLI	CFI	SRMR	RMSEA
Model 1 (unidimensional)	249.67***	14	17.8	0.975	0.983	0.072	0.156[0.133;0.180]
Model 2 (unidimensional, Item 1 ~~ Item 2, Item 2 ~~ Item 3)	78.97***	12	6.6	0.994	0.997	0.041	0.076[0.050;0.105]
Model 3ǂ (two-factor: cohesion and engagement)	238.95***	14	17.1	0.975	0.983	0.071	0.158[0.135;0.182]
Model 4ǂǂ (two-factor: affective and behavioral)	79.71***	13	6.1	0.994	0.986	0.040	0.075[0.049;0.102]

****p < 0.001.*
ǂCorrelation between the cohesion and engagement latent factors was r = 0.88.
ǂǂCorrelation between the affective and behavioral latent factors was r = 0.79.

since the GSRS measures patients' relationships not only with the therapists/group leaders but also with other members of the group; therefore, there is an apparent overlap in what the instrument is expected to measure. Although the affective domain was more promising than the behavioral domain in terms of convergent validity, overall, the convergent validity of the GCS was not particularly supported in this study.

Theoretical Support for Two-Dimensional Group Cohesion

The GCS was conceived as a unidimensional construct by Wongpakaran et al. (2013). However, the unidimensional model (model 1) demonstrated an acceptable fit neither in their study nor in ours. Although the large correlation between the affective and behavioral factors may be interpreted in favor of the unidimensionality of the scale, the two dimensions are still independent to some degree and represent different phenomena conceptually. Theoretical support for the two-factor model with the affective and behavioral domains can already be found in the standardization study by Wongpakaran et al. (2013), even though these authors did not report fit indices for this model. Group cohesiveness has been recognized as a multidimensional construct several times in the past (Hogg, 1993). Mudrack's (1989) definition of group cohesion as a combination of attraction-to-group and commitment to the group task provides a solid rationale for the differentiation of group cohesion into the affective and behavioral domain. The former is associated with the attraction to the group or its members and by collectively sharing positive, as well as negative, emotional experiences (Barsade and Knight, 2015). The latter, on the other hand, is associated with a commitment to the group (Mudrack, 1989) that may be manifested, for instance, by following group rules or giving gifts to other members (Lawler et al., 2000). Another literature supporting the two-dimensional model was Carron et al. (1985) who defined the individual group factor (commitment to other members of group) and task-social factor (interest in the goals of the group). Cota et al. (1995) in their review of group cohesion structure discussed both unidimensionality and multidimensionality resulting in favoring the multidimensional perspective (normative and behavioral components are divided and considered primary components of group cohesion). Kipnes et al. (2002) tested group cohesion dimensionality using two different instruments and claimed that cohesion is a multidimensional construct and offer a hierarchical structure [first order factors will be (1) bond to individual members and (2) level of trust and encouragement of the group as a whole].

TABLE 5 | Standardized regression weights (factor loadings) and errors (N = 369).

	Model 1		Model 4		
	λ	h^2	λ_{F1}	λ_{F2}	h^2
Item 1	0.75	0.56	0.78	–	0.61
Item 2	0.89	0.80	0.94	–	0.87
Item 3	0.82	0.67	0.85	–	0.73
Item 4	0.79	0.63	–	0.84	0.71
Item 5	0.83	0.69	–	0.87	0.75
Item 6	0.69	0.47	–	0.72	0.52
Item 7	0.56	0.31	–	0.58	0.34
McDonald's omega	0.910		0.860	0.811	
Raykov's omega	0.879		0.850	0.797	
Cronbach's alpha	0.896		0.884	0.828	
R^2	61.9%		32.7%	35.6%	

R^2, explained variance of the model; λ, factor loadings; h^2, communality; F1, affective domain; F2, behavioral domain.
Correlation between F1 and F2 in model 4: r = 0.785. McDonald's omega total for model 4 = 0.894.

In summary, given the high internal consistency of the unidimensional solution and the large correlation between the affective and behavioral dimensions, the GCS may be used as an essentially unidimensional measure of group cohesiveness. However, it should be done with caution and with the awareness of the fact that group cohesiveness may be, in fact, composed of different and partially independent phenomena.

Similarities and Differences Between the Thai and Czech Versions

The Czech version of the GCS demonstrated some features similar to those of the Thai version. Both versions were characterized by similar values of item-total correlations and internal consistency. Item loadings in terms of the unidimensional model were very similar for both versions as

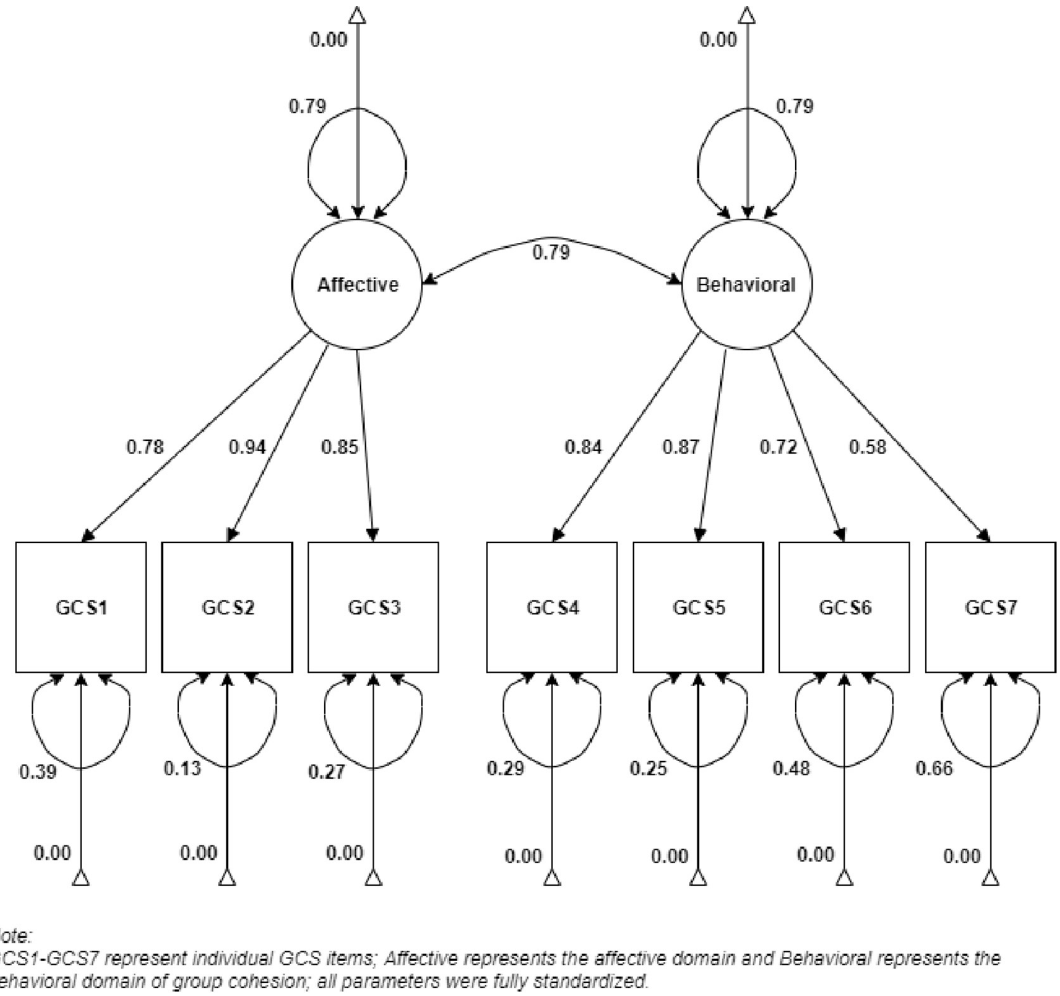

FIGURE 1 | Factor structure of Group Cohesiveness Scale.

well. The GCS scores were relatively skewed in both studies. Patients tended to perceive their groups as rather cohesive in both cultures. Based on these similarities, we can argue that both versions are comparable.

However, certain differences between the Czech and Thai versions can be found. The two-dimensional solution as the best fitting solution is different from the original unidimensional solution. This may be attributed to cultural differences. Furthermore, the mean total score of the unidimensional model was higher in the Thai version (4.7) than in the Czech version (3.7). Therefore, Thai participants might perceive therapeutic groups as generally more cohesive than Czech participants do or might be less willing to report a lack of cohesion.

Limitations

First, the sample was relatively heterogeneous and did not represent both genders equally (70% were female). Although this corresponds to the fact that most psychotherapy clients are women, future studies may investigate male groups to explore possible differences in the factor structure of group cohesion. Second, 67 patients dropped out of the study by the second week (i.e., the time when the first measurement of group cohesion took place). Although there were no significant differences between those who dropped out and those who continued with the treatment, this number of participants could have changed some subtle structures within the data. Third, two models yielded a satisfactory fit. The selection of the final model, even though theoretically anchored, is always relatively arbitrary in such cases. Moreover, none of the models fulfilled the criteria for a good fit regarding the χ^2/df fit index. However, the chi-square test of model fit (and its derivatives) are sample size sensitive and could lead to the rejection of factor model even when residual variances are negligible. Fourth, the final two-factor model was invariant across age cohorts, genders, education levels, and time. Nevertheless, response options 1, 2, and 3 were clustered into a single response option because of missing response patterns in the data. This reduction of thresholds might have distorted our conclusions about the invariance. This response pattern might be explained by the tendency of group members to perceive their group likewise; hence, their responses to the measurement tool

TABLE 6 | Measurement invariance for Model 4 across age, gender, and education level.

	χ^2(df)	TLI	RMSEA	SRMR	$\Delta\chi^2$	Δdf	ΔTLI	ΔRMSEA	ΔSRMR
Age									
Configural	**58.11 (26)**	0.988	0.083	0.042	–	–	–	–	–
Metric *(loadings, free var, free means)*	65.18 (31)	0.989	0.078	**0.042**	5.76	5	0.001	–0.005	0.00
Scalar *(loadings, intercepts, free var, free means)*	79.00 (36)	0.988	0.082	0.043	11.92*	5	–0.001	0.003	0.00
Factor means	61.05 (38)	**0.994**	**0.058**	0.043	0.48	2	0.006	–0.023	0.00
Residuals	101.3 (43)	0.987	0.087	0.052	18.97**	7	–0.002	0.005	0.01
Gender									
Configural	**52.60 (26)**	0.991	0.075	**0.041**	–	–	–	–	–
Metric *(loadings, free var, free means)*	58.77 (31)	0.992	0.071	0.041	5.65	5	0.001	–0.005	0.000
Scalar *(loadings, intercepts, free var, free means)*	65.51 (36)	0.992	0.067	0.041	6.58	5	0.001	–0.003	0.000
Factor means	59.98 (38)	**0.995**	**0.057**	0.041	2.11	2	0.002	–0.011	0.000
Residuals	90.90 (43)	0.990	0.079	0.055	16.74*	7	–0.003	0.011	0.014
Education									
Configural	**64.19 (26)**	0.985	0.091	**0.046**	–	–	–	–	–
Metric *(loadings, free var, free means)*	76.73 (31)	0.985	0.091	0.047	11.54*	5	0.000	0.000	0.001
Scalar *(loadings, intercepts, free var, free means)*	88.38 (36)	0.985	0.090	0.047	10.52*	5	0.000	–0.001	0.001
Factor means	76.33 (38)	**0.990**	**0.075**	0.047	2.35	2	0.005	–0.015	0.000
Residuals	96.26 (43)	0.987	0.083	0.056	11.71	7	0.002	–0.007	0.009
Time (comparing week 2 and week 6)									
Configural	**66.49 (26)**	0.994	0.070	**0.029**	–	–	–	–	–
Metric *(loadings, free var, free means)*	71.35 (31)	0.995	0.064	0.029	3.18	5	0.001	–0.006	0.000
Scalar *(loadings, intercepts, free var, free means)*	85.44 (36)	0.995	0.066	0.029	11.84*	5	0.000	0.002	0.000
Factor means	71.66 (38)	**0.997**	**0.053**	0.029	2.88	2	0.002	–0.013	0.000
Residuals	120.7 (43)	0.993	0.075	0.037	25.92***	7	–0.002	0.010	0.007

**p < 0.05, **p < 0.01, ***p < 0.001.*
Values in bold are the best fitting values when all invariance models within the same groups are compared. The factor mean level and residual level were both compared with the scalar level.

or to particular items could be limited to a very homogenous response style (Evans and Jarvis, 1980).

CONCLUSION

The Czech version of the GCS is a reliable and psychometrically valid tool for the measurement of the affective and behavioral domains of group cohesiveness. Thanks to its brevity, the scale is useful in the rapid hospital or therapeutic environment. As far as we know, this is the first psychometric validation of the GCS in Western culture and the Caucasian population. In this study, we revised the originally proposed unidimensional factor structure (Wongpakaran et al., 2013) and found support for the existence of the affective and behavioral domain of group cohesion.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The datasets analyzed during the current study are available from the corresponding author on reasonable request. Requests to access these datasets should be directed to AK, klocek.adam@mail.muni.cz; https://www.researchgate.net/publication/344571321_GCS_full_dataset.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee of Masaryk University (Ref. No. EKV-2017-029-R1). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AK: conceptualization, theoretical literature search, analysis, writing, and reviewing. TŘ: conceptualization and reviewing. HC: analysis and reviewing. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Czech Ministry of Education, Youth, and Sports, internal grant MUNI/A/0993/2019.

REFERENCES

Ashby, K. M., Collins, D. L., Helms, J. E., and Manlove, J. (2018). Let's talk about race: evaluating a college interracial discussion group on race. *J. Multicult. Counsel. Dev.* 46, 97–114. doi: 10.1002/jmcd.12095

Barsade, S. G., and Knight, A. P. (2015). Group affect. *Annu. Rev. Organ. Psychol. Organ. Behav.* 2, 21–46. doi: 10.1146/annurev-orgpsych-032414-111316

Budman, S. H., Soldz, S., Demby, A., Davis, M., and Merry, J. (1993). What is cohesiveness? An empirical examination. *Small Group Res.* 24, 199–216. doi: 10.1177/1046496493242003

Budman, S. H., Soldz, S., Demby, A., Feldstein, M., Springer, T., and Davis, M. S. (1989). Cohesion, alliance and outcome in group psychotherapy. *Psychiatry* 52, 339–350. doi: 10.1080/00332747.1989.11024456

Carron, A. V. (1982). Cohesiveness in sport groups: interpretations and considerations. *J. Sport Psychol.* 4, 123–138.

Carron, A. V., Widmeyer, W. N., and Brawley, L. R. (1985). The development of an instrument to assess cohesion in sport teams: the group environment questionnaire. *J. Sport Exerc. Psychol.* 7, 244–266. doi: 10.1123/jsp.7.3.244

Cota, A. A., Evans, C. R., Dion, K. L., Kilik, L., and Longman, R. S. (1995). The structure of group cohesion. *Pers. Soc. Psychol. Bull.* 21, 572–580. doi: 10.1177/0146167295216003

Evans, N. J., and Jarvis, P. A. (1980). Group cohesion: a review and reevaluation. *Small Group Behav.* 11, 359–370. doi: 10.1177/104649648001100401

Hogg, M. A. (1993). Group cohesiveness: a critical review and some new directions. *Eur. Rev. Soc. Psychol.* 4, 85–111. doi: 10.1080/14792779343000031

Hooper, D., Coughlan, J., and Mullen, M. R. (2008). Structural equation modelling: guidelines for determining model fit. *Electron. J. Bus. Res. Methods* 6, 53–60.

Horvath, A. O., and Greenberg, L. S. (1989). Development and validation of the working alliance inventory. *J. Counsel. Psychol.* 36:223. doi: 10.1037/0022-0167.36.2.223

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. Multidiscip. J.* 6, 1–55. doi: 10.1080/10705519909540118

Johnson, J. E., Burlingame, G. M., Olsen, J. A., Davies, D. R., and Gleave, R. L. (2005). Group climate, cohesion, alliance, and empathy in group psychotherapy: multilevel structural equation models. *J. Couns. Psychol.* 52, 310–321. doi: 10.1037/0022-0167.52.3.310

McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Lawrence Erlbaum Associate.

Kenny, D. A., Kaniskan, B., and McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociol. Methods Res.* 44, 486–507. doi: 10.1177/0049124114543236

Kipnes, D. R., Piper, W. E., and Joyce, A. S. (2002). Cohesion and outcome in short-term psychodynamic groups for complicated grief. *Int. J. Group Psychother.* 52, 483–509. doi: 10.1521/ijgp.52.4.483.45525

Lawler, E. J., Thye, S. R., and Yoon, J. (2000). Emotion and group cohesion in productive exchange. *Am. J. Sociol.* 106, 616–657. doi: 10.1086/318965

Lese, K. P., and MacNair-Semands, R. R. (2000). The therapeutic factors inventory: development of a scale. *Group* 24, 303–317. doi: 10.1023/A:1026616626780

MacKenzie, K. R. (1983). "The clinical application of group measure," in *Advances in Group Psychotherapy: Integrating Research and Practice*, eds R. R. Dies, and K. R. MacKenzie (New York: International Universities Press), 159–170.

Mudrack, P. E. (1989). Defining group cohesiveness: a legacy of confusion? *Small Group Behav.* 20, 37–49. doi: 10.1177/104649648902000103

Poyner-Del Vento, P., Goy, E., Baddeley, J., and Libet, J. (2018). The caregivers' attachment and relationship education class: a new and promising group therapy for caregivers of individuals with Parkinson's disease. *J. Couple Relationsh. Ther.* 17, 97–113. doi: 10.1080/15332691.2017.1341356

Quirk, K., Miller, S., Duncan, B., and Owen, J. (2013). Group session rating scale: preliminary psychometrics in substance abuse group interventions. *Counsel. Psychother. Res.* 13, 194–200. doi: 10.1080/14733145.2012.744425

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raykov, T., and Marcoulides, G. A. (2011). *Introduction to Psychometric Theory*. London: Routledge.

Roark, A. E., and Sharah, H. S. (1989). Factors related to group cohesiveness. *Small Group Behav.* 20, 62–69. doi: 10.1177/104649648902000105

Rosseel, Y. (2012). Lavaan: an R package for structural equation modeling and more. Version 0.5–12 (BETA). *J. Stat. Softw.* 48, 1–36. doi: 10.1002/9781119579038.ch1

Tulin, M., Pollet, T. V., and Lehmann-Willenbrock, N. (2018). Perceived group cohesion versus actual social structure: a study using social network analysis of egocentric Facebook networks. *Soc. Sci. Res.* 74, 161–175. doi: 10.1016/j.ssresearch.2018.04.004

Wongpakaran, T., Wongpakaran, N., Intachote-Sakamoto, R., and Boripuntakul, T. (2013). The Group Cohesiveness Scale (GCS) for psychiatric inpatients. *Perspect. Psychiatr. Care* 49, 58–64. doi: 10.1111/j.1744-6163.2012.00342.x

Wu, H., and Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika* 81, 1014–1045. doi: 10.1007/s11336-016-9506-0

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 in Psychology

ORIGINAL RESEARCH
published: 11 January 2021
doi: 10.3389/fpsyg.2020.607671

Toward a Contextually Valid Assessment of Partner Violence: Development and Psycho-Sociometric Evaluation of the Gendered Violence in Partnerships Scale (GVPS)

Katharina Goessmann[1], Hawkar Ibrahim[1,2,3]*, Laura B. Saupe[1,4] and Frank Neuner[1,3]*

[1] Department of Clinical Psychology and Psychotherapy, Bielefeld University, Bielefeld, Germany, [2] Department of Clinical Psychology, Koya University, Koya, Iraq, [3] Vivo International, Konstanz, Germany, [4] Department of Clinical and Biological Psychology, Catholic University of Eichstätt-Ingolstadt, Eichstätt, Germany

OPEN ACCESS

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

Reviewed by:
Dominic Willmott, Manchester Metropolitan University, United Kingdom
Montse Subirana-Malaret, University of Barcelona, Spain
Olga Cunha, Universidade Lusófona do Porto, Portugal

***Correspondence:**
Katharina Goessmann
kgoessmann@uni-bielefeld.de; katharina.goessmann@uni-bielefeld.de
Hawkar Ibrahim
hawkar@uni-bielefeld.de

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

Received: 17 September 2020
Accepted: 10 December 2020
Published: 11 January 2021

Citation:
Goessmann K, Ibrahim H, Saupe LB and Neuner F (2021) Toward a Contextually Valid Assessment of Partner Violence: Development and Psycho-Sociometric Evaluation of the Gendered Violence in Partnerships Scale (GVPS). Front. Psychol. 11:607671. doi: 10.3389/fpsyg.2020.607671

This article presents a new measure for intimate partner violence (IPV), the Gendered Violence in Partnerships Scale (GVPS). The scale was developed in the Middle East with the aim to contribute to the global perspective on IPV by providing a contextual assessment tool for partner violence against women in violent-torn settings embedded in a patriarchal social structure. In an effort to generate a scale including IPV items relevant to the women of the population, a pragmatic step-wise procedure, with focus group discussions and expert panels, was performed. The study's analyses resulted in an 18-item checklist featuring four subscales of the GVPS that are based on a new typology of male-to-female partner violence presenting an alternative to the commonly used classification by type of abuse (i.e., physical, psychological, sexual acts). Therein, dominating behaviors, existential threats, impulsive aggression, and aggravated physical assault were identified as reflective of the lived realities of women in the war-torn Middle East, which was confirmed in factor analysis. The scale's psychometric properties were assessed with data from 1,009 displaced women in Iraq, and associations with measures of psychopathology were determined. Implications for IPV assessment and prevention possibilities in humanitarian contexts and beyond are discussed.

Keywords: violence against women, partner violence, scale development, violence, assessment, Middle East

INTRODUCTION

Intimate partner violence (IPV) against women is an ongoing global human rights issue that brings about a wide range of devastating effects on the health and wellbeing of individuals as well as societies at large (Heise and García-Moreno, 2002; Bonomi et al., 2006; Ellsberg and Emmelin, 2014). IPV is a multifaceted phenomenon that can manifest in a myriad of often co-occurring forms, including physical, verbal, and sexual behaviors. It occurs across all social, religious, and cultural contexts (Krane, 1996; Ellsberg et al., 2015), with 30% of all women worldwide reporting having experienced physical or sexual forms of IPV during their lifetime (Devries et al., 2013).

Studying the global prevalence, impact, and conditions of IPV is a difficult task that involves several challenges and requires thoughtful ethical considerations (Bender, 2017). Although IPV continues to affect many women across the world, it is typically considered a private issue and often remains hidden from direct observation. Many affected women fear negative consequences when reporting experiences of partner violence (Krane, 1996; Pournaghash-Tehrani, 2011). Stigmatization and victim-blaming due to socially embedded gender inequality and inadequate support systems seem also to hinder the reporting of IPV incidents (Overstreet et al., 2019). Given such challenges, underreporting is likely, and getting help is difficult for many affected women worldwide.

Still, existing prevalence research indicates alarming rates of IPV particularly in low and middle-income countries, with prevalence rates between 35 and 66% reported in countries in South Asia, Andean South America, Oceania, most parts of Africa, and the Middle East (Devries et al., 2013). One possible explanation for these high numbers may be related to intertwined impacts of factors identified as increasing the risk of violence against women, such as poverty and gender inequalities (Ellsberg et al., 2015; Heise and Kotsadam, 2015; González and Rodríguez-Planas, 2020). IPV seems to be influenced by social and cultural factors, since conditions like high gender inequality and economic dependence are associated with increased levels of oppression and violence against women (Jewkes, 2002; Ebbeler et al., 2017). Besides, in some geographical regions, the high prevalence rates may be further explained by armed conflicts and subsequent social instabilities, which have been associated with violence against women on several, including interpersonal, levels (Catani, 2010; Stark and Ager, 2011). Research indicates that rates of domestic violence against women increase when men seek to reassert power and reestablish their dominant gender roles when such roles are challenged by war or post-war living conditions (DeLargy, 2013; Guruge et al., 2017; MacKenzie and Foster, 2017).

A contextually valid assessment of IPV in diverse contexts is challenged by the fact that existing instruments have predominantly been developed and validated within relatively stable European or United States-American populations. Those instruments are often exhaustively used without or with limited cultural adaptations, which bears the risk of hiding potential context-specific phenomena and relationships (Haddad et al., 2011; Amawi et al., 2014; Wangel and Ouis, 2019). Furthermore, with most of the existing prevalence research still conducted in relatively stable Western countries, population-based data of IPV from other geographical regions, especially from more fragile (e.g., conflict-torn) societies, remain scarce (Falb et al., 2015; Heise and Kotsadam, 2015). The paucity of IPV assessment and instrument development in non-Western contexts and an inadequate variety of items call for a local development and extensive empirical validation of instruments. The comprehensive understanding of IPV globally requires the rigorous investigation of violence in a variety of contexts, including in unstable and violence-affected populations where IPV rates are reportedly high (Stark and Ager, 2011), and where the complexity of the occurrence of IPV may be influenced by several individual and structural factors (Pournaghash-Tehrani, 2011; Jayasundara et al., 2014; Wachter et al., 2018; Goessmann et al., 2019). Local pragmatic approaches are required in order to perform assessments of IPV which adequately reflect the experiences of women within their social environments. As recommended by researchers and practitioners in the field, the involvement of local communities in the definition and development process is crucial to this in order to reduce power imbalances between researchers and participants in women and violence research (Webb, 1993; Hossain and McAlpine, 2017; Fineran and Kohli, 2020). Thus, the present development study followed a pragmatic approach in which the inclusion of IPV items relevant to the lived experiences of women was paramount.

While the acts of violence perpetrated against women in heterosexual partnerships may be as diverse as the partnerships themselves, their underlying dynamics are often quite similar with aggressions mostly being used to exert physical, emotional, psychological or economic control over the partner (DeKeseredy, 2011; Devries et al., 2013). Those dynamics are usually connected to larger societal factors, such as the still widespread inequality between men and women under patriarchal order, of which IPV can be both a reflection and a constituent (Heise, 1998; Fulu and Miedema, 2015). However, the gender-related aspects of violence against women in partnerships have largely been ignored in existing IPV measurement research (Reed, 2008; Hamby, 2014; Ali et al., 2016; Bender, 2017). Violence against women, including partner violence against women, is per definition any act directed against women because they are women, or that affects women disproportionally (OHCHR, 1992). Consequently, many types of IPV are inherently gendered, such as sexual acts (e.g., forced penetrative intercourse, forced impregnation) or controlling and coercing behaviors, which make a woman dependent on their male partner and reduce their autonomy.

However, existing assessment tools of IPV usually don't reflect underlying gendered dynamics. The majority of IPV instruments apply a descriptive, tripartite categorization based on the mere appearance of the violent act, classifying partner violence as either physical, psychological/emotional, or sexual (Gómez-Fernández et al., 2019). While this typology attempts to assess all manifestations of violence, it has drawn criticism in recent years for its contribution to overlooking the gendered nature of IPV against women (Reed, 2008; Reed et al., 2010; DeKeseredy, 2011; Hamby, 2014; Ali et al., 2016; Bender, 2017). In an effort to complement the descriptive categorization of physical, psychological or sexual IPV, a growing body of theory has proposed the use of alternative categories and the inclusion of a greater variety of violent acts (Johnson and Ferraro, 2000; Johnson and Leone, 2005; Kelly and Johnson, 2008; Ali et al., 2016; Velonis, 2016; Mennicke, 2019). Researchers have suggested distinguishing acts of IPV, for example, according to violence severity and intensity, situational influences, perpetrator's motivations, societal patterns of gender-related dominance/control, and the impacts and personal meanings of the abuse for both the perpetrator and the victim, to allow valid IPV assessment that takes the context of the violence into account (Johnson and Ferraro, 2000; Bogat et al., 2005;

Ali et al., 2016; Bender, 2017). Accordingly, a number of other categories and patterns of violent acts against women in partnerships have been identified. While an extensive review of the growing literature in this regard is not feasible within the frame of this study, some theoretical developments should be mentioned. For example, research has distinguished acts that are motivated by the aim to control or dominate the women (Strauchler et al., 2004; Ali et al., 2016). Such acts have been reported to be prevalent in women's lived partnership experiences especially, but not only, in settings with pronounced patriarchal society structures subordinating women (Felson and Messner, 2000). Examples of controlling IPV may include following the partner around, determining their clothing, limiting their social interactions, as well as reproductive coercion. Other IPV events, such as manipulation and economic threats like being denied financial means, may jeopardize the partner's sense of personal safety and potential for self-sufficiency (Adams et al., 2008; Voth Schrag, 2015). Such acts are often not explicitly considered in IPV research, and thus remain a largely invisible facet of partner violence (Postmus et al., 2020).

Yet other acts of partner violence may be of rather impulsive types. For example in already ongoing conflict situations, verbal aggression (e.g., yelling, calling names) is likely to be followed by, or simultaneously occurring with, physical violence such as hitting or throwing things (Winstok and Smadar-Dror, 2018). Therein, aggression levels, conflict management styles, or substance abuse (e.g., alcohol) may have important impacts on the level of physical and verbal violence used impulsively within heterosexual partnerships (Derefinko et al., 2011; Graham et al., 2011; Cascardi et al., 2018). Regarding physically violent behaviors, research has distinguished a category of highly intense physical attacks, such as attacks with weapons or fire, which can be extremely harmful and even fatal (WHO, 2005; Stark, 2010). For such acts, a pronounced gender pattern has been identified which suggests that women are much more frequently victimized by severe physical violence than are men (Hamby, 2005; Ansara and Hindin, 2010).

The present study reports the development process of a new IPV event checklist from its initial efforts to empirical testing among displaced Syrian and Iraqi women in northern Iraq. We purposefully chose the study's location for several reasons. In Iraq, a country with comparatively high structural gender inequalities (World Bank Group, 2019), legislation granting equal rights to women and men is reportedly deficient and not implemented consistently; thus, society-wide human rights violations against women are prevalent (Davis, 2016). That includes the Kurdistan region of Iraq (KRI), where a recent study showed that women in Erbil had little knowledge of existing law enforcement structures and were reluctant to seek justice in cases of domestic violence (Malik et al., 2017). Furthermore, acceptance of physical violence against women seems to be widespread, as 63% of women participating in a survey study conducted in Iraq indicated that they approved of the use of beatings in partnerships (Linos et al., 2012). Despite its comparatively high IPV levels, the Middle East is among the regions for which very few validated IPV instruments exist (Boy and Kulczycki, 2008; Azadarmaki et al., 2016), one of the few exceptions is the Arab version of the Composite Abuse Scale (Alhabib et al., 2013). Hence, additional IPV instruments are necessary, which consider the full variety of experiences of women living in the context. The decision to use data from forcibly displaced women for the development of this new IPV scale based on the specific characteristics of these women. As mentioned earlier, armed conflicts seem to increase rates of interpersonal violence, including violence against women (Catani, 2010; Stark and Ager, 2011; Wachter et al., 2018). In various post-war and displacement settings, particularly high IPV rates are reported (e.g., Ward, 2002; Annan and Brier, 2010; Clark et al., 2010; Jewkes et al., 2017), including in Iraq (Goessmann et al., 2019). Refugee camps in the war-torn KRI thus provide a suitable environment in which to gather data on IPV exposure and to validate a new instrument for its assessment.

MATERIALS AND METHODS

The present study is part of a collaborative research project conducted by the University of Bielefeld, Germany, and Koya University, Iraq, which aims to investigate the experiences and psychological states of refugees and forcibly displaced people living in camps in the KRI. The project and its procedures featuring Syrian and Iraqi individuals and married couples have been approved by the ethical committees of the two universities involved.

The study was conducted in three phases. The first phase encompassed the initial development of the scale. Based on the results from focus group discussions with violence-affected women in northern Iraq identifying acts and patterns of IPV relevant to their living contexts, a panel of clinical experts arranged the resulting IPV items into four categories. In the second phase, data on IPV exposure and psychopathology were collected among a sample of 1,009 Iraqi and Syrian displaced women. The third phase consisted of the statistical analysis to assess the psychometric properties of the scale using confirmatory factor analysis (CFA), measuring the prevalence of violence exposure and mental health impairment among the participants, and determining the scale's convergent validity.

Phase 1: Development of the IPV Instrument

Item Generation

The first step of the development of the instrument was the generation of suitable items. Two focus group discussions with displaced Iraqi and Syrian women were conducted to discuss IPV acts and themes with the aim to incorporate types of violence into the proposed measure that are relevant for populations of women living in socially strained societies with high levels of gender inequality. This approach sought to increase the research's local relevance following recommendations for gender-based violence research methods in humanitarian settings proposed by Hossain and McAlpine (2017). Each focus group consisted of 12 women residing in a refugee camp in the KRI who had been invited to participate through oral invitations by camp community mobilizers. The group discussions were held by a local female social worker specialized in working with women affected by violence. Using example items from pre-existing IPV scales such as the WHO Violence Against Women Instrument (Nybergh et al., 2013) and the Composite Abuse Scale (Hegarty et al., 1999, 2005), the participating women were asked about acts of partner violence that play a role in their own lives or within their communities.

After in-depth consultation with local experts in violence research, all IPV acts identified by the focus group participants were assessed for face validity by three members of the research team, as well as for their alignment with recommendations and guidelines for domestic violence research (WHO, 2001; Ellsberg and Heise, 2005; Hossain and McAlpine, 2017) to determine their inclusion in the questionnaire. That resulted in a list of 23 items covering acts of physical, emotional, verbal, sexual, controlling, and economic abuse. Since the focus group discussions had been conducted in Kurdish and Arabic languages, the generated item list was translated to English for further analyses. All translations including those described below were performed by multilingual clinical experts following translation guidelines for transcultural research (Human Services Research Institute, 2005).

Item Categorization

The next step of the process was to prepare the item list for psychometrically evaluation among Iraqi and Syrian displaced couples. A panel of six local and international clinical experts in violence research organized the identified 23 items based on patterns identified by the women in the focus group discussions and based on theoretical considerations of the content, meaning, and motivational characteristics of the acts within the given context of gendered societal norms of the Middle East. That resulted in a typology classifying violent acts against women in partnerships into four categories that were labeled as *Dominating behaviors*, *Existential threats*, *Impulsive aggressions (physical and verbal)*, and *Aggravated physical assault*.

Assigned to the category of *Dominating behaviors* were those acts of IPV which had been described by women in the focus groups as reflecting the husband's intention to control, such as violating their freedom through deprivation of rights, interdictions, and coercive sexual acts. Seven items were identified to be fitting to these criteria, namely: (1) Being followed or watched, (2) being prevented from visits to family or friends, (3) controlled clothing decisions, (4) being prevented from working/studying, (5) forced sexual intercourse, (6) disregard during sexual intercourse, and (7) forceful impregnation.

Assigned to the *Existential threats* category were those IPV behaviors which, while they are also closely related to the subordination of women and their forced obedience to a male partner, were described as potentially posing severe risk of losing status and of social disadvantage. The items included in this subscale are all more or less economic and finance-related, such as being denied access to financial means or being forced to sell one's possessions. Six items were included in this category: (8) Threat to get another wife/partner beside the spouse, (9) threat of being divorced, (10) threat to be thrown out of the house, (11) being forced to ask family or friends for money, (12) being forced to sell own personal possessions, and (13) to be denied financial means even if they are available.

Items reflecting acts described as mainly impulsive and to be occurring during situational partner conflicts, such as yelling or throwing things, were assigned to the *Impulsive aggressions* category. The six items assigned to this category were (14) name-calling, (15) use of disrespectful language, (16) pushing, hitting, kicking, beating, punching, slapping, (17) pulling the hair, (18) twisting arms, and (19) throwing things at the partner.

Finally, the category of *Aggravated physical assault* comprises IPV acts of intense physical violence with potential health- and life-threatening consequences (e.g., burning, attacks with weapons, etc.). This category had four items assigned to it: (20) Strangulation/attempting to strangle the partner, (21) burning or scalding, (22) attempt to kill the partner with a weapon, and (23) attacking the partner with a weapon/gun or knife.

The instrument was conceptualized as a checklist, as the aim of the study was to develop a short and pragmatic IPV instrument that is applicable in a variety of social contexts including complex humanitarian settings. However, in order to allow comprehensive assessments of both types and frequencies of IPV among this study's participants, the preliminary 23-item questionnaire featured an answer format using a five-point-scale (scoring 0–4), indicating the frequency of the occurrence of each act with regard to the past year (*never*, *once*, *once per month*, *once per week*, or *daily*).

Phase 2: Data Collection

Participants' Characteristics

The sample recruited for the statistical analyses of this study consisted of 1,009 married Syrian (48.4%) and Iraqi (51.5%) women. Participants' ages ranged from 15 to 75 years (M = 33.58, SD = 11.42) and the mean age for getting married was 19.43 (SD = 4.41, range 7–41). Almost all of the participating women were currently married (97.2%), while the other participants were either widowed (1.9%), divorced (0.5%), or separated (0.4%), though all participants had been with their partner within the past year at the time of assessment. The average duration of formal education was 4.26 years (SD = 4.32, range 0–18). Very few (6.4%) participants had an income of their own: 10,327.65 Iraqi Dinar (less than 8 EUR or 9 USD; SD = 65,479.04, range 0–900,000 Iraqi Dinar) per month on average. The women had 3.8 children on average (SD = 2.72, range 0–15); less than a tenth (8.9%) had no children at all.

Procedures and Instruments

Data collection was conducted in camps for displaced people located in Duhok and Sulaymaniyah, KRI. Due to lack of reliable census data for the camps, sampling was performed using a pragmatic approach. The camps were subdivided into sections according to approximately equal household numbers. Households in each section were randomly selected for participation by spinning a pen from the section center on a camp map. Interview staff then visited the selected households to determine eligibility of women for the study. Approval of the study procedures was provided by the camp administrations and the ethical committees of Bielefeld University, Germany, and Koya University, KRI. Structured interviews were conducted with participants in either Arabic (41.7%) or Kurdish (58.3%).

Interviews lasted between 60 and 90 min and took place in the homes of the participants without any other person present to insure privacy and confidentiality. All interviewers were fluent in Kurdish and Arabic, held University degrees in either psychology or social work, and had been trained in data collection procedures prior to conducting assessments. Due to cultural considerations, participants' consent was obtained in verbal rather than written form after informing them about the study's procedures and their rights as participants (Ibrahim and Hassan, 2017). For underage participants, their parents' consent was additionally obtained. A comprehensive risk management procedure was established to protect participants and staff. Women who reported being affected by severe violence or mental health issues were offered counseling by psychologically trained staff and were referred to further health care providers if needed. Telephone numbers of emergency and violence prevention hotlines and contacts to local support organizations in and outside of camps were also handed out to participating women. Details on the comprehensive measures taken to protect respondents and staff during and around data collection, including the focus groups, to ensure ethically sound research procedures have been described in previous publications generated from this research project (Ibrahim et al., 2018a,b, 2019; Goessmann et al., 2019).

As the present study focuses on violence against women in partnerships, it includes only data from women. In addition to collecting information on participants' experiences of IPV, sociodemographic information and mental health issues, in terms of depression and posttraumatic stress disorder (PTSD), were also assessed using validated instruments. PTSD symptoms were measured using the Arab and Kurdish versions of the PTSD Checklist for DSM-5 (PCL-5) which had been validated in the KRI (Ibrahim et al., 2018a). Depression symptoms were measured with the 15-item depression subscale of the Hopkins Symptom Checklist (HSCL-D), a valid cross-cultural instrument that has been utilized previously in displaced Arabic populations (Tinghög and Carstensen, 2010; Al-Turkait et al., 2011). Internal consistency of both the PCL-5 and HSCL-D in this sample was good, as indicated by Cronbach's alpha values of $\alpha = 0.91$ (PCL-5) and $\alpha = 0.86$ (HSCL-D), respectively.

Phase 3: Statistical Analysis

Means, standard deviations, ranges, and frequencies were calculated to describe the sample characteristics as well as violence and psychopathology prevalence. To examine the proposed item structure of the Gendered Violence in Partnerships Scale (GVPS), CFA was performed. Model fit of the CFA was tested along the criteria for model fit indices suggested by Schermelleh-Engel et al. (2003). Since the Chi-square test is extremely sensitive to large sample sizes (Bentler and Bonett, 1980), we instead relied on other fit indices including Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Incremental Fit Index (IFI), Root-Mean-Square Error of Approximation (RMSEA), and Standardized Root-Mean-Square Residual (SRMR). Model fit was considered to be acceptable if the model featured TLI, IFI, and CFI values of ≥ 0.90, an RMSEA value of ≤ 0.06 with 90% confidence interval values <0.05 (lower value) and <0.08 (upper value), and an SRMR value below 0.08. In the first step of the CFA, the 23 continuous items of the preliminary questionnaire were entered according to the four preassigned subscales (**Table 1**). After an initial examination of the results, the item loadings and the scale structure were discussed by a panel of transcultural clinical experts. The discussion resulted in recoding and rearranging of the items based on similarity and co-occurrence of items and on contextually informed considerations to reflect the participating women's lived realities. Then, the CFA was run again. The final scale resulting from the second step of the CFA was then used to calculate descriptive statistics of the participants' IPV experiences (see **Figure 1**), as well as indicators of internal consistency and convergent validity based on associations with mental health measures (see **Table 3**). Data analyses were carried out using IBM SPSS and Amos version 25. The internal consistency of the resulting IPV scale and the subscales was tested using Cronbach's alpha reliability. Convergent validity was measured on the basis of correlation analyses of the IPV sum score and the IPV subscale scores with mental health indicators (depression and PTSD scores).

RESULTS

Confirmatory Factor Analysis of the Gendered Violence in Partnerships Scale

The first step of the CFA included all 23 items initially derived from the focus group discussions and resulted in a model with inadequate fit (CFI = 0.85, RMSEA = 0.089 [90%-CI = 0.085 −0.092, PCLOSE = 0.00], SRMR = 0.06). Some items showed very low factor loadings (i.e., below 0.40; see **Table 1**). After discussing the results with a group of experts, a re-arrangement of the items was made with the main aim not to lose any informative value of the initial item list. Since the items with low factors loadings represented rather rare events but were nonetheless reflective of relevant experiences of the women participating in the focus groups (e.g., "Has your partner tried to kill you with a gun?"), their information was retained.

In step 2 of the factor analysis, instead of excluding items with lower factor loadings, thematically related items were combined and rephrased to create four new items (see **Table 2**). This led to the reduction of the total number of items from 23 to 18. In the subscale of *Dominating behaviors*, two items addressing sexual violence (having one's own sexual needs ignored by their partner; getting impregnated against their will) were combined to create one new item of sexual subjugation. The two items regarding the forced acquisition of money (being forced to ask family or friends for money; being forced to sell one's personal possessions) were combined into one new item under the subscale of *Existential threats*. The three items representing physical attacks without objects (being pushed/kicked/slapped; having their hair pulled; having their arms twisted) were combined into one item in the subscale of *Impulsive aggressions*. Lastly, in the *Aggravated physical assault* subscale, the two items addressing physical violence with weapons (attempted murder with a weapon; attack with a gun or knife) were combined into one item. The number of items in each of the four subscales (*Dominating behaviors*,

TABLE 1 | Factor loadings of step 1 of the factor analysis including 23 items.

Item	Factor 1	Factor 2	Factor 3	Factor 4
Has your partner impregnated you against your will?[1]	0.350			
Has your partner disregarded you during sex and only focused on their own pleasure?[1]	0.335			
Has your partner forced you to have sex when you did not want to?	0.622			
Has your partner prevented you from working/studying?	0.447			
Has your partner controlled what you wear?	0.531			
Has your partner prevented you from visiting your family or friends?	0.655			
Has your partner followed you or watched you?	0.634			
Has your partner left you alone in the house without money even though they had money?		0.596		
Has your partner forced you to sell your personal possessions (e.g., house or jewelry)?[2]		0.707		
Has your partner forced you to ask your family or friends for money?[2]		0.532		
Has your partner threatened to throw you out of the house?		0.830		
Has your partner threatened you with divorce?		0.783		
Has your partner threatened to get another wife/partner?		0.605		
Has your partner thrown things at you?			0.770	
Has your partner twisted your arms?[3]			0.826	
Has your partner pulled your hair?[3]			0.840	
Has your partner pushed, hit, kicked, beaten, punched, or slapped you?[3]			0.824	
Has your partner used disrespectful language toward you?			0.750	
Has your partner called you names?			0.738	
Has your partner attacked you with a weapon (such as a gun or a knife)?[4]				0.361
Has your partner tried to kill you with a weapon?[4]				0.488
Has your partner burned or scalded you?				0.742
Has your partner tried to strangle you?				0.927

Model fit: χ^2[224, N = 1009] = 1996.66 (p < 0.001), CFI = 0.85, RMSEA = 0.089 [90%-CI = 0.085 −0.092, PCLOSE = 0.00], SRMR = 0.06.
[1] Two items which were combined into one item on sexual subjection in step 2 of the CFA.
[2] Two items that were combined into one item on forced money acquisition in step 2 of the CFA.
[3] Three items that were combined into one item on physical attacks in step 2 of the CFA.
[4] Two items that were combined into one item on physical attacks with weapons in step 2 of the CFA.

Existential threats, *Impulsive aggressions*, and *Aggravated physical assault*) were reduced to six, five, four, and three items, respectively (see **Table 2**). Since the study aimed for the creation of a pragmatic IPV assessment instrument to be used in unstable contexts such as displacement camps where the feasibility of extensive surveys is limited, the scale's response format was changed to a binary format (no/yes; scored 0–1). In order to provide complete versions of the instrument in the languages in which the original items were generated, the four combined items were back-translated to Arabic, Kurdish Kurmanji, and Kurdish Sorani. Translations were performed by clinical experts with experience in instrument translations, and translation accuracy was verified by independent language experts. The adaptations of step 2 resulted in a checklist of 18 items with acceptable model fit (CFI = 0.93, TLI = 0.91, IFI = 0.93, RMSEA = 0.055 [90%-CI = 0.05 −0.06, PCLOSE = 0.06], SRMR = 0.04) and moderate to high factor loadings between 0.40 and 0.80 (see **Table 2**). The resulting GVPS was then used to determine participants' overall IPV prevalence and subscale scores (see below and in **Figure 1**).

IPV Exposure and Psychopathology

Experience of IPV was high among the interviewed women, with 442 (43.8%) reporting that they had experienced at least one violent act perpetrated by their partner within the past year. The most common type of IPV reported was *Dominating behaviors* (reported by 32%), followed by *Impulsive aggressions* (25.1%), *Existential threats* (24.3%), and *Aggravated physical assault* (3.7%). Specific acts of IPV that were reported by more than 10% of the participating women included control of clothing, denial of sexual and reproductive rights, threats to get another wife, threats to be thrown out, being called names, disrespectful language use, and physical attacks (hitting, kicking, twisting arms, pulling hair). Frequencies of all individual acts and subtypes of IPV reported by the participants can be found in **Figure 1**. Psychopathology was high, with 72% of the participants (M = 35.32, SD = 17.74) endorsing PTSD symptom levels above the adapted PCL-5 cut-off value of 23 (Ibrahim et al., 2018a). An even larger proportion of participants (81.9%; M = 2.23, SD = 0.76) endorsed clinically relevant levels of depressive symptoms as measured by the HSCL-D (score > 1.55).

Reliability and Validity

The full GVPS, as well as its four subscales, showed moderate to good internal consistency indicated by Cronbach's alpha reliability values between 0.65 and 0.88. For the *Dominating behaviors* subscale reliability was α = 0.65, for the *Existential threats* subscale α = 0.72, for the *Impulsive aggressions* subscale α = 0.78, for the *Aggravated physical assaults* subscale α = 0.70, and for the full scale it was α = 0.88. The four subscales correlated significantly with each other and with the sum score. Correlation

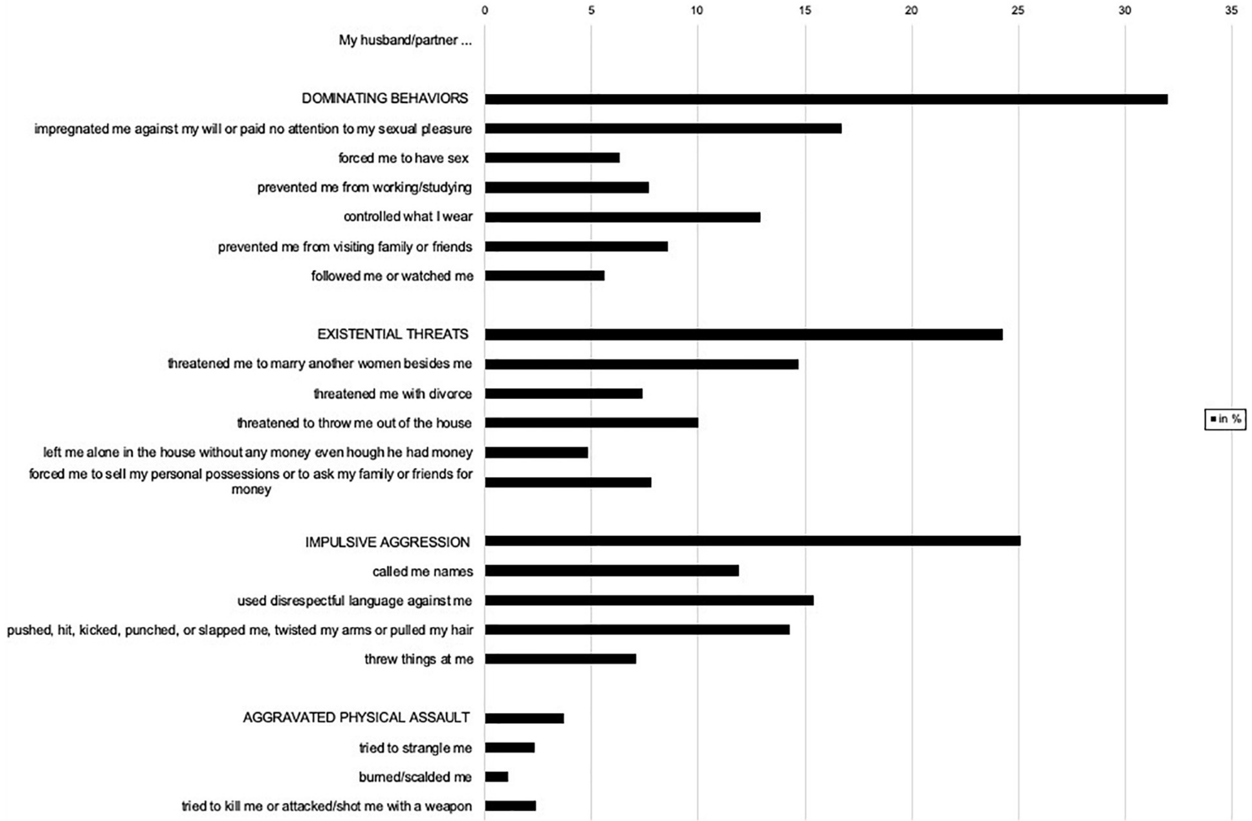

FIGURE 1 | Past-year male-perpetrated partner violence reported by women (N = 1009).

coefficients ranged from 0.48 to 0.89. The subscale *Aggravated physical assaults* showed the lowest correlations with the other subscales as well as with the GVPS sum score ($r = 0.65$, $p < 0.01$), while the correlations of the three other subscales with the GVPS sum score were all well above 0.80 (see **Table 3**).

The total score as well as the subscale scores of the GVPS showed good convergent validity with measurements of women's mental health status. All measures of PTSD and depression symptoms were significantly correlated with the GVPS score and the four subscale scores (see **Table 3**). Correlations with depression were similarly high for experiences of *Existential threats* ($r = 0.21$, $p < 0.01$), *Impulsive aggressions* ($r = 0.19$, $p < 0.01$) and *Dominating behaviors* ($r = 0.18$, $p < 0.01$), and lowest for experiences of *Aggregated physical assaults* ($r = 0.13$, $p < 0.01$). The correlations of *Existential threats* and *Impulsive aggressions* with depression were both significantly higher than the correlation of *Aggregated physical assaults* with depression, $Z = -2.67$, $p < 0.01$ and $Z = -1.90$, $p < 0.05$. All other subscale correlations with depression did not differ significantly from each other. The pattern was similar for PTSD symptoms, with *Existential threats*, *Impulsive aggressions* and *Dominating behaviors* all showing significant correlations above 0.26 with the PCL-5 sum score. The correlation of PTSD with *Aggregated physical assaults* ($r = 0.17$, $p < 0.01$) was significantly smaller than the correlation of PTSD with *Existential threats* ($Z = -3.28$, $p < 0.001$), *Impulsive aggressions* ($Z = -2.89$, $p < 0.01$), and *Dominating behaviors* ($Z = -2.97$, $p < 0.001$).

DISCUSSION

The present study described the development process and psychometric evaluation of the GVPS (see **Table 4**), a new checklist for the assessment of IPV, that was evaluated in a displacement setting in the Middle East. The study fills a gap in the literature of adequate IPV assessment for women in violent-torn environments by providing a pragmatic, contextually valid event checklist. The primary aim of the study's three-phase development procedure was to ensure the process to be locally informed in order to create a pragmatic instrument that reflected the living situations of the women involved. To this end, focus groups of Syrian and Iraqi displaced women discussed and identified acts and patterns of IPV prevalent in their community. The emerging IPV items were then checked for face validity by local and international experts and arranged into four thematic categories, and the resulting item list was psychometrically analyzed using the data from 1,009 Syrian and Iraqi displaced women.

The results of the study provide evidence for the validity and reliability of the GVPS. A two-step factor analysis confirmed

TABLE 2 | Factor loadings of the final factor solution of the GVPS including 18 binary-coded items.

Item (new item number in 18-item checklist)	Factor 1	Factor 2	Factor 3	Factor 4
Has your partner impregnated you against your will or has your partner neglected you during sex and only focused on their own pleasure? (sexual subjugation)	0.433			
Has your partner forced you to have sex when you did not want to?	0.583			
Has your partner prevented you from working/studying?	0.406			
Has your partner controlled what you wear?	0.475			
Has your partner prevented you from visiting your family or friends?	0.558			
Has your partner followed or watched you?	0.525			
Has your partner left you alone in the house without any money even though they had money?		0.509		
Has your partner forced you to sell your personal possessions (e.g., house or jewelry) or forced you to ask your family or friends for money?		0.569		
Has your partner threatened to throw you out of the house?		0.704		
Has your partner threatened you with divorce?		0.664		
Has your partner threatened to get another wife/partner?		0.553		
Has your partner thrown things at you?			0.688	
Has your partner pushed, hit, kicked, beaten, punched, or slapped you, twisted your arms or pulled your hair?			0.722	
Has your partner used disrespectful language toward you?			0.695	
Has your partner called you names?			0.664	
Has your partner tried to kill you or attacked you with a weapon, gun or knife?				0.633
Has your partner burned or scalded you?				0.598
Has your partner tried to strangle you?				0.800

Model fit: χ^2[129, N = 1009] = 518.07 (p < 0.001), CFI = 0.93, TLI = 0.91, IFI = 0.93, RMSEA = 0.055 [90%-CI = 0.05 –0.06, PCLOSE = 0.06], SRMR = 0.04.
New items obtained through combinations of previous items are in bold.

TABLE 3 | Bivariate correlations between IPV scores and mental health scores.

	1	2	3	4	5	6	7	8	9	10	11
1 IPV CL sum score	—	0.86**	0.65**	0.89**	0.85**	0.22**	0.29**	0.23**	0.20**	0.24**	0.28**
2 Subscale Impulsive aggressions (physical/verbal)		—	0.48**	0.70**	0.58**	0.19**	0.26**	0.22**	0.16**	0.21**	0.25**
3 Subscale Aggravated physical assault			—	0.53**	0.51**	0.13**	0.17**	0.11**	0.16**	0.17**	0.15**
4 Subscale Existential threats				—	0.64**	0.21**	0.27**	0.23**	0.19**	0.22**	0.27**
5 Subscale Dominating behaviors					—	0.18**	0.26**	0.18**	0.15**	0.18**	0.22**
6 HSCL-D sum score						—	0.68**	0.54**	0.30**	0.64**	0.65**
7 PCL-5 sum score							—	0.85**	0.57**	0.90**	0.88**
8 PCL-5 Intrusions								—	0.39**	0.67**	0.65**
9 PCL-5 Avoidance									—	0.41**	0.41**
10 PCL-5 Cognitions and mood										—	0.71**
11 PCL-5 Arousal											—

Pearson's correlations, two-tailed. HSCL-D, Hopkins Symptom Checklist for Depression; PCL-5, PTSD checklist for DSM-5.
$^{**}p$ < 0.01.

the scale's psychometric properties and its proposed factor structure of *Dominating behaviors*, *Existential threats*, *Impulsive aggressions*, and *Aggregated physical assault*. While in the first step of the factor analysis, the initial model with 23 items showed inadequate model fit and some critically low factor loadings, the model fit of the 18-item version was acceptable, with moderate to high factor loadings on all four subscales. Its model fit indices are in line with other measurement scales evaluated using CFA (e.g., Boduszek et al., 2018; Hooker et al., 2019). Somewhat lower but still acceptable factor loadings between 0.40 and 0.50 were observed for three items in the *Dominating behaviors* subscale, which also had the lowest Cronbach's alpha coefficient (α = 0.65). This finding might best be explained by the inclusion of both sexually and non-sexually dominating acts in this subscale. The reason for the integration of those different acts into one subscale was based on the supposition that their common elements were their oppressive nature and the manner in which they put women in a position of subordination under a male partner's control and domination (Kelly and Johnson, 2008). Research has indicated that sexual coercion, as well as psychological control

TABLE 4 | Gendered violence in partnerships scale (GVPS) (English version).

	No (0)	Yes (1)
Dominating behaviors		
Has your partner followed you or watched you?		
Has your partner controlled what you wear?		
Has your partner prevented you from visiting your family or friends?		
Has your partner prevented you from working/studying?		
Has your partner forced you to have sex when you did not want to?		
Has your partner impregnated you against your will or has your partner neglected you during sex and only focused on his own pleasure? (sexual subjugation)		
Existential threats		
Has your partner threatened you with divorce?		
Has your partner threatened to throw you out of the house?		
Has your partner forced you to sell your personal possessions (e.g., house or jewelry) or forced you to ask your family or friends for money?		
Has your partner left you alone in the house without any money even though they had money?		
Has your partner threatened to get another wife/partner?		
Impulsive aggressions (verbal and physical)		
Has your partner called you names?		
Has your partner pushed, hit, kicked, beaten, punched, or slapped you, twisted your arms or pulled your hair?		
Has your partner thrown things at you?		
Has your partner used disrespectful language toward you?		
Aggravated physical assault		
Has your partner tried to strangle you?		
Has your partner tried to kill you or attacked you with a weapon, gun or knife?		
Has your partner burned or scalded you?		

Versions of the GVPS in Arabic, Kurdish Kurmanji, and Kurdish Sorani are available upon request.

and intimidation, increase negative health outcomes for women, especially if they co-occur, which highlights the role of gender and power relations for the impacts of IPV (Pico-Alfonso et al., 2006; Caldwell et al., 2012). The item with the lowest scoring on the *Dominating behaviors* subscale was the prevention from working or studying. In previous IPV instruments, this behavior has been assigned to acts of economic oppression (Adams et al., 2008), which are represented in our *Existential Threats* subscale. However, discussions with local experts revealed that, in the given social context, being denied access to education or work is considered an act of control rather than an existential threat, since men's intention to regulate women's every behavior and whereabouts is the driving underlying motivation for it. The *Dominating behaviors* subscale thus makes theoretical sense in the context, and its internal consistency of $\alpha = 0.65$ can be considered acceptable for a subscale with six items covering two different aspects (i.e., psychological and sexual acts) of controlling and dominating behaviors (Streiner, 2003).

The four established subscales correlated significantly with each other as well as with the GVPS sum score. Only the correlations of the *Aggravated physical assault* subscale were somewhat lower than those between the other three subscales. A possible explanation for this finding might be the relative rareness of the events covered by the three items of the *Aggravated physical assault* subscale (i.e., strangulation; burning/scalding; attacks with weapons), which limits its representativeness for the GVPS as a whole and enhances skewness of the distributions. Despite their comparatively rare report, collecting information on acts of extreme physical violence is crucial to understand the full extent of the variety of women's IPV experiences. It is important to keep in mind that more severely abused women tend to be less likely to report their abuse or participate in surveys, and reaching them might require particular efforts (Waltermaurer et al., 2003). That may also explain the lower reported frequencies of the *Aggravated physical assault* items, a finding for which possible reporting biases should also be considered responsible due to potential fear and shame. By contrast, *Dominating behaviors* were the most prevalent forms of violence reported in this sample, indicating that acts of manipulation and control, as well as sexual coercion, are common experiences in the daily lives of many of the participating women. The frequent report of events such as having their sexual and reproductive rights denied or clothing regulations draws a dark picture of the subordination of women and highlights the patriarchal contexts in which IPV often occurs. The belief that holds women to be inferior to men is still prevalent across the globe. Women's rights are disrespected in many ways, and women are often expected to subordinate themselves, which in turn can facilitate their victimization of physical violence (Namy et al., 2017). The high correlations of both the *Existential threats* and the *Dominating behaviors* subscales with the *Impulsive aggressions* subscale, which showed particularly high prevalence for the items on physical violence and disrespectful language use, also indicated the connection of subordination and physical violence victimization. Overall, the prevalence of 44% found in this sample for the whole GVPS exceeds the average prevalence level of 35% previously reported for Middle Eastern countries (Devries et al., 2013) and shows partner violence, in its numerous forms, to be a significant issue among displaced Syrian and Iraqi couples. This is in line with previous research indicating burdened and violent partnership and family relations in conflict-affected contexts (Catani, 2010; Stark and Ager, 2011). However, in light of frequent underreporting of IPV, especially in cases where the relationship with the abuser is ongoing and if women themselves tend to justify spousal violence (Al-Modallal, 2015), it has to be kept in mind that this number might still be an underestimation of the actual severity of abuse experienced by the interviewed women.

The study's results further provide initial indications of good convergent validity of the GVPS and its subscales. Significant correlations of IPV with measures of depression and PTSD symptomatology were found, which is in line with previous research highlighting the negative impacts of IPV on women's mental health (Ellsberg and Emmelin, 2014). Some of the highest correlations of psychopathology measures with subtypes of IPV were those which are a product of male dominance over women (i.e., dominating behaviors and existential threats). This sheds light on the often neglected living situation of Iraqi and Syrian women in the KRI, which seems to be characterized by an intertwining of health impairment and ongoing violence embedded in patriarchal societal structures (Johnson and Leone, 2005). Elucidating the dynamics of the home context of Iraqi and Syrian women seems critical, as the effects of psychologically manipulative acts on (mental) health often go unnoticed, particularly if they occur in combination with other forms of IPV (Arriaga and Schkeryantz, 2015). Existential threats in particular bear the risk of holding the affected women in a continuous state of helplessness, as possibilities to seek support are limited by the abuse itself, for example when contact with friends and relatives is forbidden.

The present study has some important implications for theoretical IPV research and practice as well as for intervention efforts to improve the living conditions of violent-affected women. Following the call by scholars and practitioners in the field to develop instruments differentiating between thematic types of IPV and to empirically validate them in different settings and populations (Kelly and Johnson, 2008; Ali et al., 2016), this study is a first step toward these goals providing an instrument with new meaningful IPV categories that go beyond a categorization of IPV into physical, psychological and sexual abuse types. Assessing IPV experiences clustered in patterns of dominating behaviors, existential threats, impulsive aggressions, and aggravated physical assaults might be useful to professionals to detect underlying dynamics and thus tailor specified interventions. Furthermore, discussing those patterns with violence-affected women might help them gain awareness for indications and associations of IPV in their partnerships. The results on frequency levels and mental health associations found in this study are indications of the value of the proposed GVPS subscales for comprehensive assessments of the prevalence and manifestations of IPV in a setting of predominant gender hierarchies. The high level of IPV exposure, as well as its correlations with depression and PTSD psychopathology found in this study, indicate ongoing insecurity and hardship for women living in (post-)war contexts, an issue that calls for focused attention and action within humanitarian care efforts. Although we developed the scale in a specific context, and the test of the GVPS in other social and cultural settings is still pending, the GVPS items cover a variety of violent acts potentially relevant to the lived realities of many women worldwide and thus offer possibilities to investigate conditions and circumstances of IPV. Future usage and application of the GVPS in different settings should demonstrate that the new categorization provided by this instrument is helpful in the study of the causes and consequences of IPV and may thus help gather further details on the circumstances of violence against women in partnerships to plan and conduct interventions appropriately.

Strengths and Limitations

To our knowledge, this is the first study to scientifically develop and validate an IPV instrument among a population of displaced women in the Middle East using a pragmatic approach with focus groups and expert group discussions, including members of the local communities. As opposed to previous developments or adaptations of IPV scales, our procedure followed a bottom-up approach using the experiences and perspectives of the target population as a starting point for the scale development. This approach enabled the creation of an instrument that takes the lived realities of violence-affected women into account by actively engaging them in the process. The collaboration of international and local clinical experts in the scale development further enhanced the adoption of multiple perspectives in the process and dismantled popular prerogatives of interpretation (Webb, 1993; Hossain and McAlpine, 2017).

Further, the project's realization by an international team and within Arab and Kurdish speaking areas made the simultaneous development of the scale in four languages possible; thus, the instrument is now available in English, Arabic, Kurdish Kurmanji, and Kurdish Sorani. Another advantage of the study is the size of the sample used for data collection. The sample is a good representation of women living in a setting of gender inequality and daily struggle due to ongoing social and political instabilities. However, it has to be noted that some specific characteristics of the participating displaced women might impact their IPV levels. Thus, their experiences do not necessarily reflect the lived realities of other women in Syria, Iraq, or elsewhere. For example, some of the items, such as the threat of getting another wife, are highly context-dependent. The *ad hoc* development of the scale's subscales reflects the study's pragmatic approach to create a contextually valid IPV instrument; however, it might limit the scale's generalizability across social and cultural contexts. The present study demonstrated the scale's suitability and utility in a post-war environment in the Middle East. However, the item categorization and subscales of the GVPS was based on IPV patterns identified as prevalent in the local population and might not be transferable to other contexts. Thus, the applicability, factor structure, and validity of the GVPS need to be tested by future studies in other contexts in order to prevent the risk of premature and inappropriate cross-cultural generalizations (Clark and Walker, 2011), and to make broader analyses of types and circumstances of IPV possible.

Furthermore, the external validity of the scale was tested only by using associations to women's mental health outcomes. Other validity measures, such as future predictive validity could not be assessed here due to the study's design. Future studies should investigate broader associations of IPV, including perpetrators' characteristics, as well as the question of recidivism of different IPV types to identify risk factors and promote prevention. The GVPS provides a promising tool for such analyses in longitudinal designs.

CONCLUSION

The present study introduces the GVPS, a new event checklist to assess experiences of IPV against women that was developed among women from displaced communities in northern Iraq. The development process followed a pragmatic approach aiming to increase local validity of the resulting scale by directly involving local communities and clinical experts who discussed themes and events of IPV against women in the social and cultural context. Statistical analysis found indicators of good psychometric properties of the scale and confirmatory factor analyses of the item structure confirmed the typology of four thematic subscales of IPV to be reflective of the involved women's living situations (*Dominating behaviors*, *Existential threats*, *Impulsive aggressions*, and *Aggregated physical assault*). Furthermore, the study's findings on IPV prevalence and associations with psychopathology significantly extend existing knowledge about IPV and its impacts in settings with high levels of social and political challenges. This newly developed IPV assessment tool might help to understand theoretically the nature of violence and abuse against women in highly patriarchal societies by integrating notions of power relations in gender-based violence. Furthermore, it has the potential to enable health professionals to reliably and validly estimate the suffering that stems from IPV in Iraq, other Arab countries, and beyond in order to promote the development of adequate interventions combatting the global issue of gender-based violence.

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 committees of Bielefeld University, Bielefeld, Germany and Koya University, Koya, Iraq. Participants' informed consent was obtained in oral form. After receiving detailed information on participation and the study's objectives, each participant's informed consent and agreement to participate was documented carefully by the interviewer on a standardized consent information form.

AUTHOR CONTRIBUTIONS

FN and HI obtained the funding of the research. KG was the main contributor of the specific design of this study. KG and HI managed and supervised to collect the data. HI, LS, and KG performed the data analysis. FN, LS, and HI contributed to data interpretation and the structure of the manuscript. KG drafted the manuscript and all authors contributed to the final version.

FUNDING

Funding of this study was provided by the Volkswagen Foundation (grant number 91474-1). A part of this study was supported by a scholarship from the German Academic Exchange Service (DAAD) to HI (grant number 57129429, Personal Ref. No. 91562643). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We acknowledge support for the publication costs by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University. We are grateful to all women who participated in this study. We thank the mental health professionals from local and international NGOs, and the camp hospitals who provided health support for participants in need. Special thanks go to our local research team for their dedicated effort in this research project. We also thank the camp managers, administrations, and security staff for the facilitation of the study's procedures. Thank you to Justin Preston for language editing.

REFERENCES

Adams, A. E., Sullivan, C. M., Bybee, D., and Greeson, M. R. (2008). Development of the scale of economic abuse. *Viol. Against Women* 14, 563–588. doi: 10.1177/1077801208315529

Alhabib, S., Feder, G., and Horwood, J. (2013). English to arabic translation of the composite abuse scale (CAS): a multi-method approach. *PLoS One* 8:75244. doi: 10.1371/journal.pone.0075244

Ali, P. A., Dhingra, K., and McGarry, J. (2016). A literature review of intimate partner violence and its classifications. *Aggress. Viol. Behav.* 31, 16–25. doi: 10.1016/j.avb.2016.06.008

Al-Modallal, H. (2015). Disclosure of spousal violence: viewpoints from the Jordanian society. *J. Int. Viol.* 32, 164–182. doi: 10.1177/0886260515585538

Al-Turkait, F. A., Ohaeri, J. U., El-Abbasi, A. H. M., and Naguy, A. (2011). Relationship between symptoms of anxiety and depression in a sample of Arab college students using the Hopkins Symptom Checklist 25. *Psychopathology* 44, 230–241. doi: 10.1159/000322797

Amawi, N., Mollica, R. F., Lavelle, J., Osman, O., and Nasir, L. (2014). Overview of research on the mental health impact of violence in the Middle East in light of the Arab Spring. *J. Nerv. Mental Dis.* 202, 625–629. doi: 10.1097/NMD.0000000000000174

Annan, J., and Brier, M. (2010). The risk of return: intimate partner violence in Northern Uganda's armed conflict. *Soc. Sci. Med.* 70, 152–159. doi: 10.1016/j.socscimed.2009.09.027

Ansara, D. L., and Hindin, M. J. (2010). Exploring gender differences in the patterns of intimate partner violence in Canada: a latent class approach. *J. Epidemiol. Commun. Health* 64, 849–854. doi: 10.1136/jech.2009.095208

Arriaga, X. B., and Schkeryantz, E. L. (2015). Intimate relationships and personal distress: the invisible harm of psychological aggression. *Pers. Soc. Psychol. Bull.* 41, 1332–1344. doi: 10.1177/0146167215594123

Azadarmaki, T., Kassani, A., Menati, R., Hassanzadeh, J., and Menati, W. (2016). Psychometric properties of a screening instrument for domestic violence in a sample of Iranian women. *Nurs. Midwif. Stud.* 5:e27763. doi: 10.17795/nmsjournal27763

Bender, A. K. (2017). Ethics, methods, and measures in intimate partner violence research: the current state of the field. *Viol. Against Women* 23, 1382–1413. doi: 10.1177/1077801216658977

Bentler, P. M., and Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* 88, 588–606. doi: 10.1037/0033-2909.88.3.588

Boduszek, D., Debowska, A., Sherretts, N., and Willmott, D. (2018). Psychopathic Personality Traits Scale (PPTS): construct validity of the instrument in a sample of U.S. prisoners. *Front. Psychol.* 9:1596. doi: 10.3389/fpsyg.2018.01596

Bogat, G. A., Levendosky, A. A., and Von Eye, A. (2005). The future of research on intimate partner violence: person-oriented and variable-oriented perspectives. *Am. J. Commun. Psychol.* 36, 49–70. doi: 10.1007/s10464-005-6232-7

Bonomi, A. E., Thompson, R. S., Anderson, M. L., Reid, R. J., Carrell, D., Dimer, J. A., et al. (2006). Intimate partner violence and women's physical, mental, and social functioning. *Am. J. Prevent. Med.* 30, 458–466. doi: 10.1016/j.amepre.2006.01.015

Boy, A., and Kulczycki, A. (2008). What we know about intimate partner violence in the Middle East and North Africa. *Viol. Against Women* 14, 53–70. doi: 10.1177/1077801207311860

Caldwell, J. E., Swan, S. C., and Woodbrown, V. D. (2012). Gender differences in intimate partner violence outcomes. *Psychol. Viol.* 2, 42–57. doi: 10.1037/a0026296

Cascardi, M., Chesin, M., and Kammen, M. (2018). Personality correlates of intimate partner violence subtypes: a latent class analysis. *Aggress. Behav.* 44, 348–361. doi: 10.1002/ab.21756

Catani, C. (2010). War at home: a review of the relationship between war trauma and family violence. *Verhaltenstherapie* 20, 19–27. doi: 10.1159/000261994

Clark, C. J., Everson-Rose, S. A., Suglia, S. F., Btoush, R., Alonso, A., and Haj-Yahia, M. M. (2010). Association between exposure to political violence and intimate-partner violence in the occupied Palestinian territory: a cross-sectional study. *Lancet* 375, 310–316. doi: 10.1016/S0140-6736(09)61827-4

Clark, J. J., and Walker, R. (2011). Research ethics in victimization studies: widening the lens. *Viol. Against Women* 17, 1489–1508. doi: 10.1177/1077801211436167

Davis, L. (2016). Iraqi women confronting ISIL: protecting women's rights in the context of conflict. *Southw. J. Int. Law* 22, 27–78.

DeKeseredy, W. S. (2011). Feminist contributions to understanding woman abuse: myths, controversies, and realities. *Aggress. Violent Behav.* 16, 297–302. doi: 10.1016/j.avb.2011.04.002

DeLargy, P. (2013). "Sexual violence and women's health in war," in *Women and Wars*, ed. C. Cohn (Malden, MA: Polity Press), 54–79.

Derefinko, K., Dewall, C. N., Metze, A. V., Walsh, E. C., and Lynam, D. R. (2011). Do different facets of impulsivity predict different types of aggression? *Aggress. Behav.* 37, 223–233. doi: 10.1002/ab.20387

Devries, K. M., Mak, J. Y., Petzold, M., Child, J. C., Falder, G., Lim, S., et al. (2013). The global prevalence of intimate partner violence against women. *Science* 340, 1527–1528. doi: 10.1126/science.1240937

Ebbeler, C., Grau, I., and Banse, R. (2017). Cultural and individual factors determine physical aggression between married partners: evidence from 34 countries. *J. Cross Cult. Psychol.* 48, 1098–1118. doi: 10.1177/0022022117719497

Ellsberg, M., Arango, D. J., Morton, M., Gennari, F., Kiplesund, S., Contreras, M., et al. (2015). Prevention of violence against women and girls: what does the evidence say? *Lancet* 385, 1555–1566. doi: 10.1016/S0140-6736(14)61703-7

Ellsberg, M., and Emmelin, M. (2014). Intimate partner violence and mental health. *Glob. Health Action* 7:25658. doi: 10.3402/gha.v7.25658

Ellsberg, M., and Heise, L. L. (2005). *Researching Violence Against Women: A Practical Guide for Researchers and Activists*. Washington, DC: World Health Organization. doi: 10.4135/9781446269930

Falb, K. L., Annan, J., and Gupta, J. (2015). Achieving gender equality to reduce intimate partner violence against women. *Lancet Glob. Health* 3, e302–e303. doi: 10.1016/S2214-109X(15)00006-6

Felson, R. B., and Messner, S. F. (2000). The control motive in intimate partner violence. *Soc. Psychol. Q.* 63, 86–94. doi: 10.2307/26955883

Fineran, S., and Kohli, H. K. (2020). Muslim refugee women's perspectives on intimate partner violence. *J. Fam. Soc. Work* 23, 199–213. doi: 10.1080/10522158.2020.1742839

Fulu, E., and Miedema, S. (2015). Violence against women: globalizing the integrated ecological model. *Viol. Against Women* 21, 1431–1455. doi: 10.1177/1077801215596244

Goessmann, K., Ibrahim, H., Saupe, L. B., Ismail, A. A., and Neuner, F. (2019). The contribution of mental health and gender attitudes to intimate partner violence in the context of war and displacement: evidence from a multi-informant couple survey in Iraq. *Soc. Sci. Med.* 237:112457. doi: 10.1016/j.socscimed.2019.112457

Gómez-Fernández, M. A., Goberna-Tricas, J., and Payá-Sánchez, M. (2019). Characteristics and clinical applicability of the validated scales and tools for screening, evaluating and measuring the risk of intimate partner violence. Systematic literature review (2003–2017). *Aggress. Violent Behav.* 44, 57–66. doi: 10.1016/j.avb.2018.11.006

González, L., and Rodríguez-Planas, N. (2020). Gender norms and intimate partner violence. *J. Econ. Behav. Organ.* 178, 223–248. doi: 10.1016/j.jebo.2020.07.024

Graham, K., Bernards, S., Wilsnack, S. C., and Gmel, G. (2011). Alcohol may not cause partner violence but it seems to make it worse: a cross national comparison of the relationship between alcohol and severity of partner violence. *J. Int. Viol.* 26, 1503–1523. doi: 10.1177/0886260510370596

Guruge, S., Ford-Gilboe, M., Varcoe, C., Jayasuriya-Illesinghe, V., Ganesan, M., Sivayogan, S., et al. (2017). Intimate partner violence in the post-war context: women's experiences and community leaders' perceptions in the Eastern Province of Sri Lanka. *PLoS One* 12:e0174801. doi: 10.1371/journal.pone.0174801

Haddad, L. G., Shotar, A., Younger, J. B., Alzyoud, S., and Bouhaidar, C. M. (2011). Screening for domestic violence in Jordan: validation of an Arabic version of a domestic violence against women questionnaire. *Int. J. Womens Health* 3, 79–86. doi: 10.2147/IJWH.S17135

Hamby, S. L. (2005). Measuring gender differences in partner violence: implications from research on other forms of violent and socially undesirable behavior. *Sex Roles* 52, 725–742. doi: 10.1007/s11199-005-4195-7

Hamby, S. L. (2014). Intimate partner and sexual violence research: scientific progress, scientific challenges, and gender. *Trau. Viol. Abuse* 15, 149–158. doi: 10.1177/1524838014520723

Hegarty, K., Bush, R., and Sheehan, M. (2005). The composite abuse scale: further development and assessment of reliability and validity of a multidimensional partner abuse measure in clinical settings. *Viol. Vict.* 20, 529–548. doi: 10.1891/vivi.2005.20.5.529

Hegarty, K., Sheehan, M., and Schonfeld, C. (1999). A multidimensional definition of partner abuse: development and preliminary validation of the Composite Abuse Scale. *J. Fam. Viol.* 14, 399–415. doi: 10.1023/A:1022834215681

Heise, L. L. (1998). Violence against women: an integrated, ecological framework. *Viol. Against Women* 4, 262–290. doi: 10.1177/1077801298004003002

Heise, L. L., and García-Moreno, C. (2002). "Violence by intimate partners," in *World Report on Violence and Health*, eds E. G. Krug, L. L. Dahlberg, J. A. Mercy, A. B. Zwi, and R. Lozano (Geneva: World Health Organization), 87–121.

Heise, L. L., and Kotsadam, A. (2015). Cross-national and multilevel correlates of partner violence: an analysis of data from population-based surveys. *Lancet Glob. Health* 3, e332–e340. doi: 10.1016/S2214-109X(15)00013-3

Hooker, J. L., Dow, D., Morgan, L., Schatschneider, C., and Wetherby, A. M. (2019). Psychometric analysis of the repetitive behavior scale-revised using confirmatory factor analysis in children with autism. *Aut. Res.* 12, 1399–1410. doi: 10.1002/aur.2159

Hossain, M., and McAlpine, A. (2017). *Gender Based Violence Research Methodologies in Humanitarian Settings: An Evidence Review and Recommendations*. Cardiff: Elrha. doi: 10.15713/ins.mmj.3

Human Services Research Institute (2005). *Toolkit on Translating and Adapting Instruments*. Available online at: http://www.hsri.org/files/uploads/publications/PN54_Translating_and_Adapting.pdf (accessed November 20, 2019).

Ibrahim, H., Catani, C., Ismail, A. A., and Neuner, F. (2019). Dimensional structure and cultural invariance of DSM V post-traumatic stress disorder among Iraqi and Syrian displaced people. *Front. Psychol.* 10:1505. doi: 10.3389/fpsyg.2019.01505

Ibrahim, H., Ertl, V., Catani, C., Ismail, A. A., and Neuner, F. (2018a). The validity of posttraumatic stress disorder checklist for DSM-5 (PCL-5) as screening instrument with kurdish and arab displaced populations living in the kurdistan region of Iraq. *BMC Psychiatry* 18:259. doi: 10.1186/s12888-018-1839-z

Ibrahim, H., Ertl, V., Catani, C., Ismail, A. A., and Neuner, F. (2018b). Trauma and perceived social rejection among Yazidi women and girls who survived enslavement and genocide. *BMC Med.* 16:154. doi: 10.1186/s12916-018-1140-5

Ibrahim, H., and Hassan, C. Q. (2017). Post-traumatic stress disorder symptoms resulting from torture and other traumatic events among Syrian Kurdish refugees in Kurdistan Region. *Iraq. Front. Psychol.* 8:241. doi: 10.3389/fpsyg.2017.00241

Jayasundara, D., Nedegaard, R., Sharma, B., and Flanagan, K. (2014). Intimate partner violence in Muslim communities. *Arts Soc. Sci. J.* 1:003. doi: 10.4172/2151-6200.S1-003

Jewkes, R. (2002). Intimate partner violence: causes and prevention. *Lancet* 359, 1423–1429. doi: 10.1016/S0140-6736(02)08357-5

Jewkes, R., Jama-Shai, N., and Sikweyiya, Y. (2017). Enduring impact of conflict on mental health and gender-based violence perpetration in Bougainville, Papua New Guinea: a cross-sectional study. *PLoS One* 12:e0186062. doi: 10.1371/journal.pone.0186062

Johnson, M. P., and Ferraro, K. J. (2000). Research on domestic violence in the 1990s: making distinctions. *J. Marr. Fam.* 62, 948–963. doi: 10.1111/j.1741-3737.2000.00948.x

Johnson, M. P., and Leone, J. M. (2005). The differential effects of intimate terrorism and situational couple violence: findings from the national violence against women survey. *J. Fam. Issues* 26, 322–349. doi: 10.1177/0192513X04270345

Kelly, J. B., and Johnson, M. P. (2008). Differentiation among types of intimate partner violence: research update and implications for interventions. *Fam. Court Rev.* 46, 476–499. doi: 10.1111/j.1744-1617.2008.00215.x

Krane, J. (1996). Violence against women in intimate relations: insights from cross-cultural analyses. *Transcult. Psychiatric Res. Rev.* 33, 435–465. doi: 10.1177/136346159603300404

Linos, N., Khawaja, M., and Kaplan, R. L. (2012). Women's acceptance of spousal abuse in Iraq: prevalence rates and the role of female empowerment characteristics. *J. Fam. Viol.* 27, 625–633. doi: 10.1007/s10896-012-9462-0

MacKenzie, M., and Foster, A. (2017). Masculinity nostalgia: how war and occupation inspire a yearning for gender order. *Secur. Dialog.* 48, 206–223. doi: 10.1177/0967010617696238

Malik, I. A., Shabila, N. P., and Al-Hadithi, T. S. (2017). Women's knowledge of the domestic violence legislation in Erbil, Iraq and their response to spousal violence. *J. Fam. Viol.* 32, 47–53. doi: 10.1007/s10896-016-9829-8

Mennicke, A. (2019). Expanding and validating a typology of intimate partner violence: intersections of violence and control within relationships. *Viol. Against Women* 25, 379–400. doi: 10.1177/1077801218780362

Namy, S., Carlson, C., O'Hara, K., Nakuti, J., Bukuluki, P., Lwanyaaga, J., et al. (2017). Towards a feminist understanding of intersecting violence against women and children in the family. *Soc. Sci. Med.* 184, 40–48. doi: 10.1016/j.socscimed.2017.04.042

Nybergh, L., Taft, C., and Krantz, G. (2013). Psychometric properties of the WHO Violence Against Women instrument in a female population-based sample in Sweden: a cross-sectional survey. *BMJ Open* 3:e002053. doi: 10.1136/bmjopen-2012-002053

OHCHR (1992). *General Recommendation No. 19. Committee on the Elimination of Discrimination Against Women, Eleventh session*. New York, NY: United Nations.

Overstreet, N. M., Willie, T. C., and Sullivan, T. P. (2019). Stigmatizing reactions versus general negative reactions to partner violence disclosure as predictors of avoidance coping and depression. *J. Int. Viol.* 34, 1734–1752. doi: 10.1177/0886260516653753

Pico-Alfonso, M. A., Garcia-Linares, M. I., Celda-Navarro, N., Blasco-Ros, C., Echeburúa, E., and Martinez, M. (2006). The impact of physical, psychological, and sexual intimate male partner violence on women's mental health: depressive symptoms, posttraumatic stress disorder, state anxiety, and suicide. *J. Womens Health* 15, 599–611. doi: 10.1089/jwh.2006.15.599

Postmus, J. L., Hoge, G. L., Breckenridge, J., Sharp-Jeffs, N., and Chung, D. (2020). Economic abuse as an invisible form of domestic violence: a multicountry review. *Trau. Viol. Abuse* 21, 261–283. doi: 10.1177/1524838018764160

Pournaghash-Tehrani, S. (2011). Domestic violence in Iran: a literature review. *Aggress. Viol. Behav.* 16, 1–5. doi: 10.1016/j.avb.2010.12.001

Reed, E. (2008). Intimate partner violence: a gender-based issue? *Am. J. Public Health* 98, 197–198. doi: 10.2105/AJPH.2007.125765

Reed, E., Raj, A., Miller, E., and Silverman, J. G. (2010). Losing the "gender" in gender-based violence: the missteps of research on dating and intimate partner violence. *Viol. Against Women* 16, 348–354. doi: 10.1177/1077801209361127

Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods Psychol. Res.* 8, 23–74.

Stark, E. (2010). Do violent acts equal abuse? Resolving the gender parity/asymmetry dilemma. *Sex Roles* 62, 201–211. doi: 10.1007/s11199-009-9717-2

Stark, L., and Ager, A. (2011). A systematic review of prevalence studies of gender-based violence in complex emergencies. *Trau. Viol. Abuse* 12, 127–134. doi: 10.1177/1524838011404252

Strauchler, O., McCloskey, K., Malloy, K., Sitaker, M., Grigsby, N., and Gillig, P. (2004). Humiliation, manipulation, and control: evidence of centrality in domestic violence against an adult partner. *J. Fam. Viol.* 19, 339–354. doi: 10.1007/s10896-004-0679-4

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J. Pers. Assess.* 80, 99–103. doi: 10.1207/S15327752JPA8001

Tinghög, P., and Carstensen, J. (2010). Cross-cultural equivalence of HSCL-25 and WHO (ten) wellbeing index: findings from a population-based survey of immigrants and non-immigrants in Sweden. *Commun. Mental Health J.* 46, 65–76. doi: 10.1007/s10597-009-9227-2

Velonis, A. J. (2016). "He never did anything you typically think of as abuse": experiences with violence in controlling and non-controlling relationships in a non-agency sample of women. *Viol. Against Women* 22, 1031–1054. doi: 10.1177/1077801215618805

Voth Schrag, R. J. (2015). Economic abuse and later material hardship: is depression a mediator? *J. Women Soc. Work* 30, 341–351. doi: 10.1177/0886109914541118

Wachter, K., Horn, R., Friis, E., Falb, K. L., Ward, L., Apio, C., et al. (2018). Drivers of intimate partner violence against women in three refugee camps. *Viol. Against Women* 24, 286–306. doi: 10.1177/1077801216689163

Waltermaurer, E. M., Ortega, C. A., and McNutt, L. A. (2003). Issues in estimating the prevalence of intimate partner violence: assessing the impact of abuse status on participation bias. *J. Int. Viol.* 18, 959–974. doi: 10.1177/0886260503255283

Wangel, A.-M., and Ouis, P. (2019). Reflections on the translation into Arabic and validation process of the NorAQ abuse questionnaire. *J. Int. Viol.* 34, 585–598. doi: 10.1177/0886260516644595

Ward, J. (2002). *If not Now, When? Addressing Gender-Based Violence in Refugee, Internally Displaced, and Post-Conflict Settings: A Global Overview*. New York, NY: Reproductive Health for Refugees Consortium.

Webb, C. (1993). Feminist research: definitions, methodology, methods and evaluation. *J. Adv. Nurs.* 18, 416–423. doi: 10.1046/j.1365-2648.1993.18030416.x

WHO (2001). *Putting Women First: Ethical and Safety Recommendations for Research on Domestic Violence Against Women*. Available online at: https://www.who.int/gender/violence/womenfirtseng.pdf (accessed March 3, 2019).

WHO (2005). *WHO Multi-Country Study on Women's Health and Domestic Violence Against Women: Initital Results on Prevalence, Health Outcomes and Women's Responses*, eds C. García-Moreno, H. A. F. M. Jansen, M. Ellsberg, L. L. Heise, and C. H. Watts (Geneva: World Health Organization). doi: 10.1016/S0140-6736(06)69523-8

Winstok, Z., and Smadar-Dror, R. (2018). Gender, escalatory tendencies, and verbal aggression in intimate relationships. *J. Inter. Viol.* 1–20. doi: 10.1177/0886260518805764 [Epub ahead of print].

World Bank Group (2019). *Women, Business and the Law 2019: A Decade of Reform*. Washington, DC: World Bank. doi: 10.1596/31327

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 in Psychology

ORIGINAL RESEARCH
published: 10 February 2021
doi: 10.3389/fpsyg.2021.624938

Check for updates

Not all Perfectionists Are as They Are Assessed: An Investigation of the Psychometric Properties of the Perfectionism Inventory in the Teaching Profession

Elena Mirela Samfira[1] *and Laurenţiu P. Maricuţoiu*[2,3]*

[1] *Teacher Training Department, Banat University of Agricultural Sciences and Veterinary Medicine from Timisoara, Timisoara, Romania,* [2] *Department of Psychology, West University of Timisoara, Timisoara, Romania,* [3] *Clinica Universitara de Terapii si Consiliere PsihoPedagogica, West University of Timisoara, Timisoara, Romania*

OPEN ACCESS

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

Reviewed by:
Jian Li, Beijing Normal University, China
Elisabeth Prestele, University of Koblenz and Landau, Germany
José Manuel García-Fernández, University of Alicante, Spain
Siti Mistima Maat, National University of Malaysia, Malaysia

****Correspondence:***
Laurenţiu P. Maricuţoiu
laurentiu.maricutoiu@e-uvt.ro

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

Received: *01 November 2020*
Accepted: *21 January 2021*
Published: *10 February 2021*

Citation:
Samfira EM and Maricuţoiu LP (2021) Not all Perfectionists Are as They Are Assessed: An Investigation of the Psychometric Properties of the Perfectionism Inventory in the Teaching Profession. Front. Psychol. 12:624938. doi: 10.3389/fpsyg.2021.624938

Perfectionism has been studied for almost 30 years. In the present study, we investigated the internal validity of The Perfectionism Inventory (PI—Hill et al., 2004) in an occupation that encourages perfectionistic tendencies in own behavior or in students' behavior. We collected data from a large sample of schoolteachers ($N = 633$, 81.18% female, 63.02% from urban areas, 46.66% from secondary schools, mean age = 42.11 years) recruited using a snowball sampling approach, and we analyzed the factor structure of the PI using confirmatory factor analyses. We found that the 8-factor structure of PI provided a reasonable fit root mean square error of approximation [RMSEA = 0.055, 90% CI = (0.053–0.057); SRMR = 0.071]. However, additional analyses revealed problematic divergent validity only in the case of the scales associated with self-evaluative perfectionism, not in the case of the scales associated with conscientious perfectionism. We found that teachers displayed distinguishably different forms of perfectionism only when it referred to own person, not when it referred to perfectionism imposed to others. Based on these findings, we suggested that the PI could provide a useful framework for investigating the role of conscientious-related forms of perfectionism in the development of teacher beliefs regarding their school behavior.

Keywords: perfectionism, The Perfectionism Inventory, schoolteacher, confirmatory factor analysis, internal validity

INTRODUCTION

Perfectionism is a complex, multidimensional personality trait (Hill et al., 2016; Stoeber, 2017) which is strongly related to various affective disorders such as anxiety, depression (Egan et al., 2011), suicide tendencies (Smith et al., 2018), and insomnia (Schmidt et al., 2018). When they define perfectionism, scientists refer to the idea of having high standards of performance (Hewitt et al., 2017), and to the idea of having overly critical evaluations of own behavior (Frost et al., 1990; Hewitt and Flett, 1991).

In educational settings, perfectionism is an important research topic because it is related to achievement and because it is highly relevant for understanding goal attainment (Flett and Hewitt, 2016). The educational environment encourages high standards of academic achievement (Flett et al., 2009; Schruder et al., 2014), therefore it can encourage perfectionistic tendencies in students and in teachers (Gilman and Ashby, 2006). As part of the educational environment, teaching is a complex task that requires teachers to set achievement goals to students, and to monitor how students fulfill these goals (Shim et al., 2020). Despite its relevance for the outcomes of the educational environment (for a detailed discussion, see Starley, 2019), research on teacher perfectionism is limited (Starley, 2019; Shim et al., 2020). As recent research suggested that perfectionistic tendencies are associated with a wide range of affective disorders (Egan et al., 2011; Maricuţoiu et al., 2019), it is important to understand how perfectionistic tendencies are manifested by teachers in the educational environment.

In this paper, our aim was to investigate the internal validity of a comprehensive measure of perfectionism [i.e., the Perfectionism Inventory (PI)—Hill et al., 2004], on a large sample of schoolteachers. There are two main arguments for conducting this research. First, perfectionism is an important teacher variable that is associated with teaching efficacy and teacher burnout (Ghorbanzadeh and Rezaie, 2016), while teacher pressure to perform was found to be related to clinical symptoms in students (Lozano et al., 2019). This means that accurate assessment of teacher perfectionism can be important for understanding both teacher-related and student-related variables. Second, the teacher perfectionistic tendencies can be enhanced by the nature of their job. As teachers are required by their students and by their peers to behave without making mistakes (Pelletier et al., 2002), it is possible that they manifest different forms of perfectionism simultaneously. Therefore, the differential diagnosis of various forms of perfectionism might be difficult in the teachers' case.

Initial research studies identified the two forms of perfectionism: adaptive (or positive) and maladaptive (or negative) perfectionism (Terry-Short et al., 1995; Flett and Hewitt, 2006; Ulu and Tezer, 2010). Adaptive perfectionism is generally understood as perfectionistic strivings (i.e., putting effort into achieving high-quality outcomes and high performance standards), while maladaptive perfectionism is generally seen as having perfectionistic concerns (i.e., overcritical self-views, uncertainty and doubts regarding own capacities or regarding the outcomes of own actions). Although these forms of perfectionism seem to be functionally opposite, they are generally seen as independent forms of perfectionism that could be observed simultaneously in one's behavior (Stoeber et al., 2020). Beyond the functional vs. dysfunctional aspects attributed to perfectionism, the concept evolved toward a multidimensional approach in the 1990s. This means that researchers identified various forms of manifestation for perfectionistic strivings and for perfectionistic concerns, which were later seen as super-ordinate (or second-order) dimensions of perfectionism. Initially, two multidimensional perspectives of perfectionism dominated the literature (and the perspective developed by Frost et al., 1990; i.e., the perspective suggested by Hewitt and Flett, 1991). The perspective developed by Hewitt and Flett (1991) described perfectionism as a three-dimensional construct: self-oriented perfectionism, other-oriented perfectionism, and socially-prescribed perfectionism. *Self-oriented perfectionism* (*SOP*) reflects the tendency of an individual to set exacting standards for oneself and stringently evaluating and censuring own behavior. *Other-oriented perfectionism* (*OOP*) reflects the tendency of an individual to have exaggerated expectations about capabilities of others and to be overcritical with them. *Socially-prescribed perfectionism* (*SPP*) reflects the perceived need of an individual to attain high standards and expectations imposed by significant others, who exert pressure on them to be perfect (Hewitt and Flett, 1991). On the other hand, the perspective proposed by Frost et al. (1990) had six dimensions: concern over mistakes, personal standards, parental expectations, parental criticism, doubting of actions, and organization. *Concern over mistakes* was conceptualized as negative responses to mistakes, a tendency to interpret mistakes as failures, and a tendency to believe that an individual will lose the respects of others after failures. *Personal standards* was conceptualized as the settings of very high standards of performance and a tendency for self-evaluation based on performance. *Parental expectations* and *Parental criticism* reflects the tendency to believe that parents set excessive goals and are overly critical. *Doubting of actions* was conceptualized as a tendency to feel projects/results are not accomplished to satisfaction. *Organization* was conceptualized to stress the importance of neatness, organization, and order (Frost et al., 1990).

Individuals with perfectionistic strivings (i.e., adaptive perfectionists) tend to recognize their limitations and find appropriate coping strategies (Flett et al., 2009), see the difficulties they face in performing tasks as real challenges, and themselves as competent persons (Frost et al., 1990). Adaptive perfectionism is closely related to experiencing strong feelings of pride associated with low feelings of shame and guilt (Stoeber et al., 2008), and it was seen as a healthy form of perfectionism (Flett and Hewitt, 2006). By contrast, people with perfectionistic concerns (i.e., *maladaptive perfectionists*) put effort to be perfect, but see themselves as being too far from perfection (Slaney et al., 2002). Maladaptive perfectionists are more likely to think in a dichotomous manner, often being overwhelmed by fear of failure and not disappointing others (Gilman and Ashby, 2006). However, recent evidence (e.g., Maricuţoiu et al., 2019) suggested that extreme levels of adaptive perfectionism are also associated with clinical syndromes of depression and anxiety.

More recently, these perspectives were combined in a comprehensive questionnaire by Hill et al. (2004). *The Perfectionism Inventory* (PI—Hill et al., 2004) combined all dimensions theorized in the 1990s in a single questionnaire with eight scales. The main advantage of using the PI over the utilization of the existing scales was that it reduced the redundancy resulted from the overlapping concepts of these scales, while providing a comprehensive assessment of perfectionism (Hill et al., 2004). In the PI, the authors grouped perfectionism dimensions in two main categories: Conscientious Perfectionism (included the factors *Organization, High Standards for Others, Striving for Excellence,* and *Planfulness*), and Self-evaluative Perfectionism (included the factors *Concern over Mistakes*, *Need for Approval*, *Parental Pressure*, and *Rumination*). The existence of second-order factors was confirmed through a confirmatory analysis of the eight scale scores (Hill et al., 2004; Cruce et al., 2012). Hill et al. (2004) argued that the use of the eight facets of perfectionism (i.e., rather than the use of the two second-order factors) could provide a more psychologically meaningful image of perfectionism.

Given the multidimensional nature of perfectionism, researchers delimited between core facets of perfectionistic concerns and strivings, and variables that are peripheral to perfectionism (Stoeber and Otto, 2006; Stricker et al., 2019). The peripheral variables include antecedents of perfectionism development (i.e., parentally prescribed perfectionism), perfectionism oriented to others, and correlates of perfectionism (e.g., planfulness, rumination, or need for approval). Therefore, the PI (Hill et al., 2004) is a diagnostic tool that includes both core and peripheral perfectionism variables. Research studies that used the PI in work contexts reported that its scales are positively correlated with perceived stress and burnout (Craioveanu, 2014), or with active coping (Crăciun and Dudău, 2014). Both Conscientious perfectionism and Self-evaluative perfectionism were positively related to stress and burnout, and had similar correlation values with these scales (Craioveanu, 2014). However, high Conscientious perfectionism was more strongly associated with high levels of active coping, as compared with Self-evaluative perfectionism (Crăciun and Dudău, 2014). On the other hand Self-evaluative perfectionism displayed stronger negative relationships with both forms of social support coping, as compared with Conscientious perfectionism (Crăciun and Dudău, 2014). Finally, the overall score of the PI was strongly associated (i.e., correlations above.40) with most symptoms assessed by the Symptom Checklist (SCL-90-R; Derogatis, 1983), except for phobia and obsession (Craioveanu, 2014).

Although the idea of combining scales from different perspectives into a single inventory was commendable, a major limitation of the Hill et al. (2004) work was that they did not present evidence for the psychometric properties of the entire set of items. In their initial work, the authors of the PI conducted separate principal components analyses for each factor. However, a single factor analysis (confirmatory or exploratory) of the entire set of items is still missing from the literature. Therefore, because we still have little evidence to assess the internal validity of the PI scales, we aimed to fill this gap by conducting a thorough investigation of the PI psychometric properties.

Perfectionism was also studied in the educational context, where high standards are promoted. Fletcher et al. (2014) stated that perfectionism exists and develops within contexts that involve relationships with parents, teachers, colleagues, coaches, and other categories. School teachers are particularly prone to developing occupational stress (Stoeber and Rennert, 2008; Sadoughi, 2017; e.g., Salmela-Aro et al., 2019), and perfectionism plays an important role in this process (Flett et al., 1995; Friedman, 2000). The educational environment is a context in which high standards are encouraged (Flett et al., 2009) and performance is expected (Schruder et al., 2014). These aspects can enhance the students' and the teachers' perfectionist tendencies (Gilman and Ashby, 2006).

Lortie (1975) argued that teachers suffer from a culture of high standards, they frequently realize that they cannot live up to the standards imposed by themselves or by others. Schoolteachers perceive a real social pressure to be perfect—from students, peers, and parents (Pelletier et al., 2002), and the fear of imperfection determines teachers to be more authoritarian (Dinkmeyer et al., 1980). More recently, Shim et al. (2020) reported that teachers that are concerned regarding their mistakes are less likely to promote the intrinsic value of learning to their students. In a similar vein, high levels of perfectionism concerns are associated with teaching efficacy and teacher burnout (Ghorbanzadeh and Rezaie, 2016). To prevent such perfectionistic behaviors and their consequences, Jones (2016) suggested that highly experienced teachers could show pre-service students how to give up their need for perfect order in their classrooms. In a similar vein, Starley (2019) emphasized the role of the educational psychologist in developing coping strategies for teachers with maladaptive perfectionist behaviors.

The research studies presented above used different perfectionism measures, based on more or less different theoretical perspectives. In the present contribution, we present evidence regarding the psychometric properties of the PI (Hill et al., 2004), a questionnaire that combined the most influential theoretical perspectives on perfectionism. By analyzing the entire item pool of the PI, we provide evidence regarding its internal validity. Furthermore, we focused on schoolteachers because the educational environment encourages the achievement of high standards (Flett et al., 2009), where there is a strong expectancy for high performance (Schruder et al., 2014). Being a teacher involves job-specific responsibilities that are similar to various facets of perfectionism. These responsibilities include encouraging students to achieve higher standards (i.e., having high standards for others), organizing and planning each lesson in detail, having a high concern over mistakes (i.e., close self-monitoring in order to avoid teaching mistakes). Therefore, the perfectionistic tendencies described above (i.e., having high standards for students, organizing and planning each lesson, monitoring the mistakes made by students) "come with the job" in the case of teachers, and this could have a negative impact on the psychometric properties of the PI (Hill et al., 2004). Because these forms of perfectionism are job-related actions, teachers' responses to items corresponding to these scales will not reflect own personal options, but rather the degree to which the respondent is performant as a teacher. This could lead to large correlations between these scales, resulted from the fact that all these behaviors are required by the respondents' job.

MATERIALS AND METHODS

Participants

We recruited 633 participants from public schools in Western regions of Romania, using a snowball sampling approach detailed in the Procedure sub-section. Most participants were female (81.18%), taught in primary schools (35.20%) and in secondary schools (45.66%), and were mostly from schools located in urban areas (55.92%). Their mean age was 42.11 years (SD = 9.80), and their mean tenure was 17.61 years (SD = 10.06). More details regarding the study sample are presented in **Table 1**.

TABLE 1 | Descriptive statistics of the sample included in the study.

		Age (years)		Teaching experience (years)	
	N	**Mean**	**SD**	**Mean**	**SD**
Total sample	633	42.11	9.80	17.61	10.06
Male	106	41.75	11.18	13.98	9.58
Female	505	42.26	8.98	18.30	9.95
School level					
Primary school teachers	219	41.84	9.43	19.04	10.76
Secondary school teachers	284	41.97	9.57	16.28	9.63
High-school teachers	112	44.06	8.58	18.17	9.12
Type of locality					
Urban	354	42.79	9.23	18.13	9.95
Rural	230	41.69	9.69	17.02	10.19

N = 633. Any differences between the cumulated number of teachers for each category and the declared sample size are due to existing non-responses in that particular category.

Measure

Perfectionism was assessed using the PI (Hill et al., 2004). The PI (Hill et al., 2004) has 59 items corresponding to 8 forms of perfectionism: *Organization* (sample item: "I like to always be organized and disciplined,"), *High Standards for Others* (sample item: "I usually let people know when their work isn't up to my standards"), *Striving for Excellence* (sample item: "My work needs to be perfect, in order for me to be satisfied"), *Planfulness* (sample item: "I think through my options carefully before making a decision"), *Concern over Mistakes* (sample item: "If I make mistakes, people might think less of me"), *Need for Approval* (sample item: "I'm concerned with whether or not other people approve of my actions"), *Parental Pressure* (sample item: "I always felt that my parent(s) wanted me to be perfect"), and *Rumination* (sample item: "When I make an error, I generally can't stop thinking about it"). Respondents must rate their agreement with each item using a 5-point Likert scale (from *1—strongly disagree* to *5—strongly agree*). The PI (Hill et al., 2004) was translated from English into Romanian by a university English teacher. Later, for the correspondence of the meaning, the Romanian version was back-translated into English by another university English teacher. Finally, translators and researchers analyzed the translation process to ensure that the true meaning of the concepts was preserved after the translation process. The reliability indices (i.e., Cronbach's alpha) of the PI scales was good: *Organization* ($\alpha = 0.862$), *High Standards for Others* ($\alpha = 0.734$), *Striving for Excellence* ($\alpha = 0.811$), *Planfulness* ($\alpha = 0.740$), *Concern over Mistakes* ($\alpha = 0.830$), *Need for Approval* ($\alpha = 0.850$), *Parental Pressure* ($\alpha = 0.906$), and *Rumination* ($\alpha = 0.848$).

Procedure

The sample of teachers was selected using two ways: (i) with the support of the school management or (ii) with the help of a teacher that recruited our participants among his/her colleagues. The teacher, with the agreement of the school principal, asked colleagues if they would like to participate in the study. All teachers who accepted, first completed an informed consent form, according to the Ethic standards in research with human subjects. Both the consent form and the PI were administered in a paper-and-pencil format. The participants were not remunerated for participating in this study.

Data Analyses

We conducted confirmatory factor analyses (CFA) using the *lavaan* package (Rosseel, 2012) in R. Firstly, we tested the eight-factor structure suggested by Hill et al. (2004). Secondly, we tested the two-factor solution theorized by Hill et al. (2004), which contains a factor named *Conscientious perfectionism* and a factor named *Self-evaluative perfectionism*. Initial investigations regarding the distribution of the responses to PI items indicated that the responses were not normally distributed (i.e., most Shapiro-Wilk tests were statistically significant). Therefore, we estimated our models using the *maximum likelihood* method, with robust standard errors (MLR). The MLR estimation implemented in *lavaan* allows for fitting models with non-normal distribution using Yuan-Bentler corrections for non-normal and missing data (Rosseel, 2012). Following the recommendations provided by Kenny et al. (2015), we computed the root mean square error of approximation (RMSEA) for the baseline model to check whether incremental fit indices (e.g., the comparative fit index, the incremental fit index, the Tucker-Lewis index) are informative in the case of our model. The RMSEA for the baseline model was 0.128, which is smaller than the threshold value of 0.158 suggested by Kenny et al. (2015) for considering the incremental indices. Therefore, we assessed model fit using the root mean square error of approximation (RMSEA, the acceptable fit is indicated by values below 0.08—Browne and Cudeck, 1993), and the standardized root mean square residual (SRMR—acceptable fit is indicated by values below 0.08—Hu and Bentler, 1999).

In addition to the confirmatory analyses, we also investigated the convergent and discriminant validity of the factor solutions. For the convergent validity, we used the criteria proposed by Anderson and Gerbing (1988): the factor loadings should be larger than 0.40, and the average variance extracted (AVE) for each factor should be above 0.50. To assess the divergent validity of each latent variable, we compared the squared root of its AVE

TABLE 2 | Fit indices of the two alternative models.

Model	Robust discrepancy	Robust RMSEA		Robust SRMR
		Value	90% CI	
8-factor model	$\chi^2(1567) = 4182.99$, $p < 0.001$	0.055	[0.053–0.057]	0.071
2-factor model	$\chi^2(1594) = 7458.66$, $p < 0.001$	0.082	[0.080–0.083]	0.124
Additional model	$\chi^2(1482) = 4606.48$, $p < 0.001$	0.058	[0.056–0.060]	0.075

TABLE 3 | Correlations matrix between the eight latent factors.

	Org	StrExc	Plan	HSO	CoM	Nap	ParPr	Rum
Org	*0.67*							
StrExc	0.40	*0.68*						
Plan	0.64	0.45	*0.55*					
HSO	0.16	0.58	0.26	*0.54*				
CoM	0.09	0.60	0.31	0.73	*0.63*			
Nap	0.07	0.56	0.26	0.78	0.95	*0.65*		
ParPr	0.06	0.51	0.17	0.44	0.55	0.48	*0.76*	
Rum	0.10	0.61	0.30	0.74	0.92	0.95	0.55	*0.67*

N = 633. Squared AVE values are presented on the diagonal, in italics. Org, organization; StrExc, striving for excellence; Plan, planfulness; HSO, high standards for others; CoM, concern over mistakes; Nap, need for approval; ParPr, perceived parental pressure; Rum, rumination.

with the correlation values between that latent variable and the other latent factors. The divergent validity is not supported if the correlation values are higher than the square root of the AVE (Chin, 1998).

RESULTS

Confirmatory Factor Analyses

The fit indices of our CFAs (presented in **Table 2**) suggested that the eight-factors model had fit indices below the 0.08 threshold value [RMSEA = 0.055, 90% CI = (0.053–0.057); SRMR = 0.071]. On the other hand, the results of the two-factors model suggested that this perspective does not provide adequate fit [RMSEA = 0.082, 90% CI = (0.080–0.083); SRMR = 0.124]. Although it had acceptable fit indices, the eight-factors model (presented in **Table 4**) had some issues regarding the convergent and divergent validity of its factors. Firstly, although most factor loadings were larger than 0.40 (i.e., only the loadings of item 13 and item 3 did not reach this threshold), the average variance extracted by the eight-factors solution reached the 0.50 value only in the case of *Perceived Parental Pressure* factor (AVE = 0.59). The other AVE values suggested that the latent factors explained between 29% (the case of *High Standards for Others*) and 46% (the case of *Striving for Excellence*) of the variance of their items. This means that the eight-factors solution does not meet the criteria for convergent validity, as defined by Anderson and Gerbing (1988).

Secondly, the divergent validity of the eight factors was generally poor. The correlation matrix between the eight scales is presented in **Table 3**, and the squared value of the AVE index is included in the diagonal. The results presented in **Table 3** suggested that divergent validity is problematic in the case of about half of the scales (i.e., *High Standards for Others*, *Concern over Mistakes*, *Need for Approval*, and *Rumination*). In the case of these scales, the squared value of the AVE is smaller than the correlation value between that scale and other factors included in the questionnaire. Simply put, these scales share more variance with other scales, than with own items. Furthermore, inter-factor correlation values are up to 0.95, which raised serious concerns regarding the divergent validity of these factors.

Additional Analyses

Given the poor divergent validity of the scales, we concluded that the scales do not assess different psychological variables. Therefore, we conducted an additional analysis to investigate whether the items have specific variance on the latent variables defined by the eight-factors model, or on the latent variables defined by the two-factors model. In this analysis (i.e., a bifactor analysis), the variance of each item is distributed between the solutions (i.e., the eight-factors and the two-factors) that are tested simultaneously in an orthogonal model (see **Figure 1** for a representation of the eight-factors, two-factors, and additional model). Consequently, the variance of each item is divided between a latent variable from the eight-factors solution and a latent variable from the two-factors solution. This analytical approach is superior to the traditional higher-order confirmatory factor analyses because it is more appropriate when it comes to dealing with multidimensionality issues (i.e., it leaves the possibility of having dimensions of the phenomenon that are independent of the general factor—Dunn and McCray, 2020), and it provides better fit of the dataset (Cucina and Byle, 2017).

To investigate how each latent variable accounts for the total variance of the items included in the analysis, we calculated the explained common variance (ECV). The ECV is computed as the

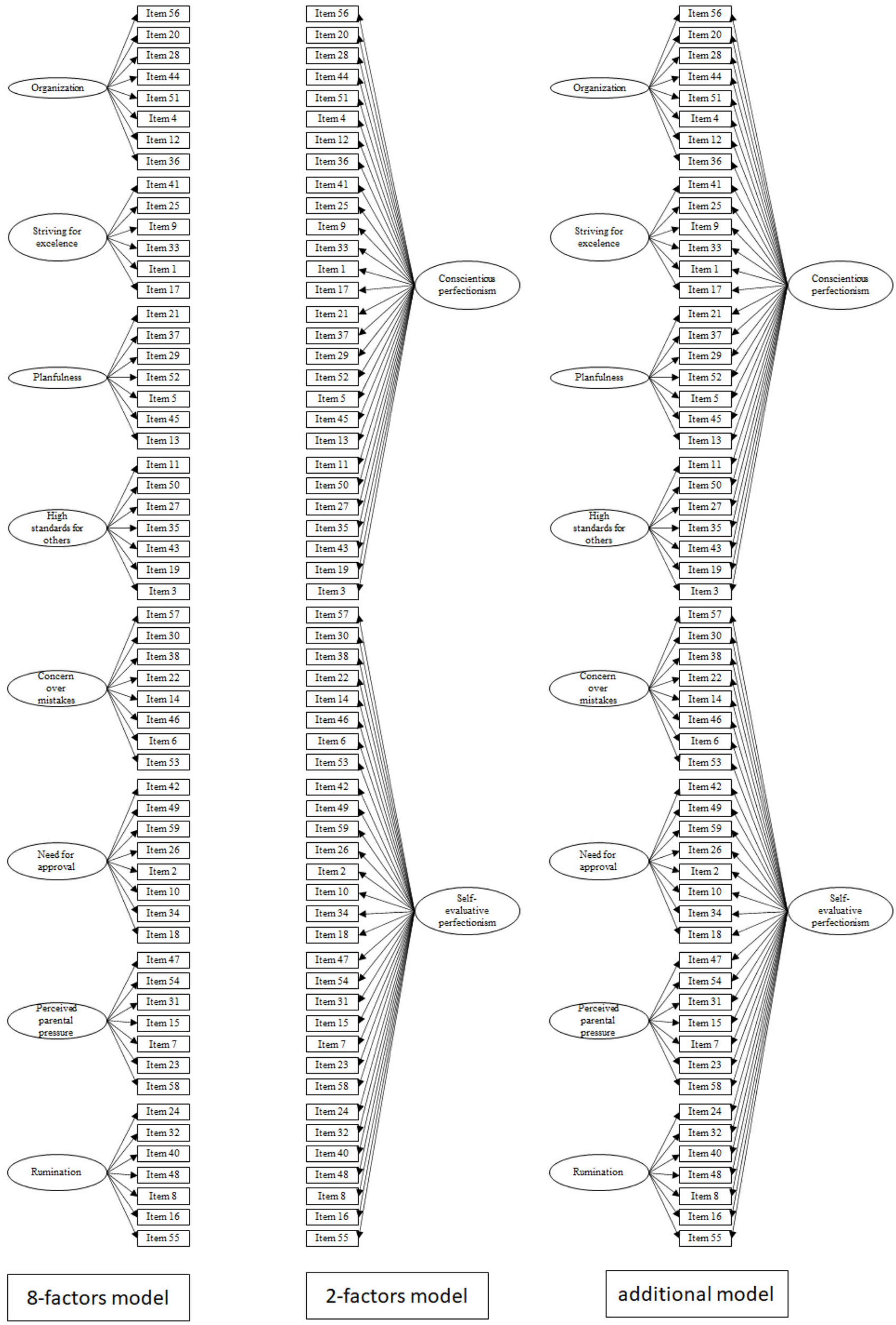

FIGURE 1 | The three models tested.

sum of the squared loadings of those latent variables, divided by the sum of all squared loadings in the model. Therefore, the ECV can be interpreted as a percentage of variance accounted by a latent variables, out of the entire variance captured by that model.

Although this type of analysis (i.e., a bifactor analysis) usually contrasts a multi-factor solution with a one-factor solution, the existence of a one-factor solution is unlikely because many correlations presented in **Table 2** also had values close to 0. Furthermore, a two-factor solution was also tested by previous studies (Hill et al., 2004; Cruce et al., 2012).

To ease their interpretation, the item loadings resulted from the additional analysis are also presented in **Table 3**, together with the loadings from the models that specified only the eight-factors model and only the two-factor model). A visual investigation of factor loadings presented in **Table 4** revealed that item loadings had close to null values in the case of *Concern over Mistakes*, *Need for Approval*, and *Rumination*. This suggests that the variance of these items is not specific to latent variables from the 8-factors model, but to the *Self-evaluation* latent variable. Regarding the remaining five latent variables, most of their items had loadings above 0.40, which suggests that the latent variables are distinct enough to account for item variance. The ECV index suggested that the *Conscientiousness* and *Self-evaluation* accounted for 61% of all explained variance, while the other eight factors only accounted for 39%. On the one hand, this result is a strong argument for reconsidering the eight-factors solution. On the other hand, the *Conscientiousness* and *Self-evaluation* are not similar regarding their capacity to explain item variance. *Self-evaluation* accounts for about 40% of all item variance, while most of its subcomponents explain less than 5% of item variance (i.e., *Concern over mistakes* = 2%; *Need for approval* < 1%; *Rumination* = 2%), while *Perceived parental pressure* is the only subcomponent that still has specific variance (i.e., ECV = 0.12). The *Conscientiousness* latent variable accounts for 21% of the explained variance, while its sub-components explain 21% increment of the explained variance (*Organization* = 6%, *Striving for Excellence* = 7%, *Planfulness* = 4%, *High Standards for Others* = 7%). This means that the *Conscientiousness* sub-components can be differentiated and should not be integrated into a single, second-order factor.

DISCUSSION

In the present research study, we investigated the internal validity of the PI (Hill et al., 2004) in an occupation that encourages perfectionistic tendencies in own behavior or in students' behavior (Shim et al., 2020). Our focus on teacher perfectionism was motivated by the fact that previous studies reported that it is a powerful predictor for teacher efficiency and teacher burnout (Craioveanu, 2014; Ghorbanzadeh and Rezaie, 2016), and can have an impact on students' variables (Lozano et al., 2019). We collected data from a large sample of schoolteachers, and we analyzed the factor structure of the PI (Hill et al., 2004) using confirmatory factor analyses.

Our CFA results suggested that the initial, eight-factor structure of TPI provided a reasonable fit on our sample of teachers. This result was encouraging because Hill et al. (2004) did not conduct a factor analysis (confirmatory or exploratory) on the entire set of items. However, additional analyses revealed that most of the latent factors explained suboptimal percentages of item variance (i.e., values below 50%), which suggested that most of the item variance remained unexplained by the eight-factor solution. Furthermore, we found evidence for problematic divergent validity in the case of about half of the scales. Although strong between-scale correlations were also present in the original study (Hill et al., 2004), the median correlation value was larger in our study (i.e., $r = 0.49$), as compared with the original study ($r = 0.37$). Based on these findings, we concluded that the eight-factor solution had serious psychometrics limitations regarding convergent and divergent validity, and we conducted additional investigations.

The original model, Hill et al. (2004) theoretized that specific factors are not independent from the general factors. However, the bifactor analysis addresses some practical issues regarding the divergent validity of the specific factors. These issues were not initially anticipated by the theoretical framework developed by Hill et al. (2004), and neither by the empirical evidence that they presented (i.e., their factor analyses based on scale scores). The bifactor analyses indicated that 61% of all explained variance can be attributed to the factors suggested by Hill et al. (2004): *Conscientiousness* and *Self-evaluative perfectionism*. However, the two factors had rather different roles. On the one hand, *Self-evaluative perfectionism* accounted for most of the explained variance (40% of the total variance), while its sub-dimensions (i.e., *Need for Approval, Rumination, Concern over Mistakes*) had very weak relations with own items. This suggests that these sub-dimensions do not have specific variance and their scores do not capture different forms of perfectionism. Previous studies reported that socially-prescribed perfectionism (e.g., high need for approval or high concern over mistakes) is related to experiencing self-conscious emotions such as shame, guilt and embarrassment (Tangney, 2002). Because these forms of perfectionism were not differentiated on our teacher sample, the PI (Hill et al., 2004) has limited capabilities regarding the differential diagnostic of the perfectionist tendencies that could explain psychological strain. However, because confirmatory analyses on the entire set of PI items are scarce, it is premature to conclude that the components of *Self-evaluative perfectionism* are generally indistinguishable one from another. For example, results suggested that the *Perceived Parental Pressure* captures specific variance that is distinguishable from its super-ordinate factor (i.e., *Self-evaluative perfectionism*). Therefore, it seems that this scale has good discriminant validity and could be seen as a form of perfectionism that is separated from the super-ordinate factors. This result can be explained by the fact that *Perceived Parental Pressure* can be interpreted as an antecedent to perfectionism (Stricker et al., 2019). To conclude, future studies should provide additional evidence regarding the specificity of the scales that compose these two forms of perfectionism. On the other hand, the *Conscientious perfectionism* supra-factor had a different role. In this case, the explained variance was equally distributed between *Conscientious perfectionism* (that accounted for 21% of the total explained variance) and its sub-scales (i.e.,

TABLE 4 | Standardized loadings of the CFA analyses.

	8 factors model								2 factors model		Bifactor model									
	Org	**StrExc**	**Plan**	**HSO**	**CoM**	**Nap**	**ParPr**	**Rum**	**Cs**	**SEv**	**Org**	**StrExc**	**Plan**	**HSO**	**CoM**	**Nap**	**ParPr**	**Rum**	**Cs**	**SEv**
it56	0.77								0.64		0.47								0.59	
it20	0.75								0.59		0.65								0.50	
it28	0.73								0.64		0.50								0.56	
it44	0.67								0.62		0.24								0.64	
it51	0.67								0.56		0.37								0.53	
it4	0.61								0.49		0.56								0.38	
it12	0.58								0.65		−0.01								0.74	
it36	0.58								0.56		0.17								0.57	
it41		0.78							0.51			0.71							0.37	
it25		0.73							0.41			0.62							0.39	
it9		0.72							0.41			0.56							0.44	
it33		0.65							0.50			0.56							0.30	
it1		0.56							0.59			0.45							0.35	
it17		0.42							0.50			0.09							0.60	
it21			0.66						0.51				0.50						0.46	
it37			0.65						0.49				0.44						0.46	
it29			0.61						0.47				0.50						0.40	
it52			0.54						0.43				0.35						0.42	
it5			0.53						0.46				0.31						0.43	
it45			0.49						0.40				0.32						0.36	
it13			0.39						0.43				0.00						0.47	
it11				0.64					0.27					0.62					0.21	
it50				0.62					0.18					0.55					0.08	
it27				0.59					0.20					0.63					0.11	
it35				0.57					0.39					0.48					0.37	
it43				0.51					0.29					0.48					0.24	
it19				0.45					0.26					0.44					0.21	
it3				0.32					0.27					0.28					0.22	
it57					0.72					0.68					0.22					0.68
it30					0.71					0.68					0.06					0.69
it38					0.66					0.64					0.28					0.63
it22					0.64					0.62					0.01					0.64
it14					0.63					0.62					−0.12					0.64
it46					0.62					0.60					0.44					0.58
it6					0.53					0.50					0.05					0.52
it53					0.40					0.37					0.42					0.35
it42						0.76				0.71						0.02				0.72
it49						0.72				0.69						0.01				0.71
it59						0.72				0.69						0.02				0.69
it26						0.64				0.64						0.02				0.64
it2						0.61				0.57						0.01				0.60
it10						0.60				0.59						0.00				0.60
it34						0.58				0.55						0.00				0.57
it18						0.55				0.52						0.01				0.55
it47							0.87			0.58							0.71			0.48
it54							0.86			0.59							0.70			0.49
it31							0.79			0.48							0.71			0.37
it15							0.74			0.42							0.70			0.31

(Continued)

TABLE 4 | Continued

	8 factors model								2 factors model		Bifactor model									
	Org	**StrExc**	**Plan**	**HSO**	**CoM**	**Nap**	**ParPr**	**Rum**	**Cs**	**SEv**	**Org**	**StrExc**	**Plan**	**HSO**	**CoM**	***N*ap**	**ParPr**	**Rum**	**Cs**	**SEv**
it7							0.72			0.53							0.55			0.47
it23							0.72			0.56							0.54			0.49
it58							0.65			0.39							0.60			0.28
it24								0.72		0.70								0.16		0.66
it32								0.71		0.67								0.51		0.66
it40								0.71		0.67								0.40		0.69
it48								0.70		0.68								0.07		0.50
it8								0.66		0.66								0.00		0.64
it16								0.66		0.63								0.09		0.70
it55								0.52		0.52								0.06		0.66
AVE	0.45	0.46	0.31	0.29	0.40	0.42	0.59	0.45	0.21	0.36	–	–	–	–	–	–	–	–	–	–
ECV	0.15	0.11	0.09	0.08	0.13	0.14	0.17	0.13	0.37	0.63	0.06	0.07	0.04	0.07	0.02	<0.01	0.12	0.02	0.21	0.40

N = 633. Org, organization; StrExc, striving for excellence; Plan, planfulness; HSO, high standards for others; CoM, concern over mistakes; Nap, need for approval; ParPr, perceived parental pressure; Rum, rumination; Cs, conscientious perfectionism; SEv, slf-evaluative perfectionism.

Organization, Striving for Excellence, Planfulness, High Standards for Others—that together accounted for 24% of the explained variance). This result suggests that the four scales assess different forms of perfectionism, each with its unique variance.

From a teacher assessment perspective, our results suggested that *Self-evaluative perfectionism* could be used as a single composite score, while component (or scale) scores should be used in the case of *Conscientious perfectionism*. This is important because the two forms of perfectionism also have different functionalities. On the one hand, the *Self-evaluative perfectionism* is associated with low levels of trait emotional stability (i.e., trait neuroticism—Cruce et al., 2012), while *Conscientious perfectionism* is associated with trait conscientiousness (Cruce et al., 2012). Previous research studies suggested that teacher neuroticism is associated with low students' self-efficacy, while teachers' conscientiousness was a predictor for the students' reports of support from the teacher (Kim et al., 2018). Based on these findings, future studies should investigate whether different forms of teacher perfectionism (i.e., self-evaluative or conscientious perfectionism) are associated with students' variables. Furthermore, the *Conscientious perfectionism* scales could be linked with individual differences in structuring and conducting teaching activities. For example, Decker and Rimm-Kaufman (2008) reported that trait conscientiousness was significantly associated with the schoolteachers' focus on the teaching process. According to their results, highly conscientious schoolteachers believe that classroom activities should have a set of explicit rules that need to be reinforced constantly, that they should organize and discuss the schedule of the day with their students, and that the teachers' primary goal is to establish and maintain classroom control (Decker and Rimm-Kaufman, 2008). Based on the relations presented above, it is possible that different forms of conscientiousness perfectionism could be related to different teacher beliefs regarding the instructional process. In this vein, future studies could investigate the relations between the PI scales (Hill et al., 2004) and various models that describe teachers' beliefs regarding the instructional process (Decker and Rimm-Kaufman, 2008), or their approaches to teaching (Trigwell and Prosser, 2004). For example, the assessment of teachers' beliefs regarding the instructional process include items that refer to scheduling the school day, establishing a morning routine in the classroom, or reinforcing the rules for students' classroom behavior (Decker and Rimm-Kaufman, 2008). Future studies could investigate whether the endorsement of the teachers' beliefs mentioned earlier is associated with teachers' forms of conscientious perfectionism.

LIMITATIONS

The present research study has some limitations that should be acknowledged. Firstly, our sample was unbalanced in terms of participants' gender (i.e., about 80% of the participants were female) and the school level (i.e., only 18% of the participants were high school teachers). On the one hand, previous research on Romanian samples did not yield gender differences regarding the levels of perfectionism (Macsinga and Dobriţa, 2010). On the other hand, the gender differences regarding some components related to Self-evaluative perfectionism (e.g., rumination—Johnson and Whisman, 2013) are very well documented in the literature. However, given the gender imbalance present in the schoolteacher population, a gender-balanced sample was difficult to attain. Regarding the school level, it is possible that high school teachers approach teaching in a different manner, as compared with primary or secondary school teachers. Therefore, their responses to some items (e.g., High standards for others) could have been different. Finally, future research studies should extend this investigation by including other multidimensional perfectionism scales (e.g., Frost et al., 1990; Hewitt and Flett, 1991) and external criteria relevant for the educational environment (e.g., approaches to teaching—Trigwell and Prosser, 2004).

CONCLUSIONS

In the present paper, we investigated the factor structure of a comprehensive inventory of perfectionism scales (i.e., the PI—Hill et al., 2004) on a large teacher sample. We found that teachers provided differentiated responses to the items of conscientious perfectionism scales, not to the items of the self-evaluative perfectionism scales. This suggests that the PI (Hill et al., 2004) could be useful to investigate how perfectionism is related to various teaching behaviors linked to conscientiousness, but the PI could be a limited measure in explaining teacher strain and teacher unwell-being.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ES selected the topic, organized the collection of the data, and contributed in the writing of the manuscript. LM contributed to the design of the study, performed the statistical analyses, and contributed in the writing of the manuscript. Both authors contributed to the manuscript revision and approved the submitted version.

REFERENCES

Anderson, J. C., and Gerbing, D. W. (1988). Structural equation modelling in practice: a review and recommended two-step approach. *Psychol. Bull.* 103, 411–423. doi: 10.1037/0033-2909.103.3.411

Browne, M. W., and Cudeck, R. (1993). "Alternative ways of assessing model fit," in *Testing Structural Equation Model*s, eds K. A. Bollen and J. S. Long (Newbury Park, CA: Sage), 136–162.

Chin, W. W. (1998). "The partial least squares approach to structural equation modelling," in *Modern Methods for Business Research,* ed G. A. Marcoulides (New York, NY: Psychology Press), 295–336.

Crăciun, B., and Dudău, D. P. (2014). The relations between perfectionism, perfectionistic self-presentation and coping strategies in medical employees. *Proc.Soc. Behav. Sci.* 127, 509–513. doi: 10.1016/j.sbspro.2014.03.300

Craioveanu, P. M. (2014). Correlations between perfectionism, stress, psychopathological symptoms and burnout in the medical field. *Proc.Soc. Behav. Sci.* 127, 529–533. doi: 10.1016/j.sbspro.2014.03.304

Cruce, S. E., Pashak, T. J., Handal, P. J., Munz, D. C., and Gfeller, J. D. (2012). Conscientious perfectionism, self-evaluative perfectionism, and the five-factor model of personality traits. *Pers. Indiv. Differ.* 53, 268–273. doi: 10.1016/j.paid.2012.03.013

Cucina, J., and Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test bateries. *J. Intell.* 5:27. doi: 10.3390/jintelligence5030027

Decker, L., and Rimm-Kaufman, S. (2008). Personality characteristics and teacher beliefs among pre-service teachers. *Teach. Educ. Q.* 35, 2, 45–64. Available online at: www.jstor.org/stable/23479223 (accessed November 16, 2019).

Derogatis, L. R. (1983). *SCL-90-R: Administration, Scoring, and Procedural Manual-II.* Baltimore, MD: Clinical Psychometric Research.

Dinkmeyer, D., McKay, G., and Dinkmeyer, D. (1980). *Systematic Training for Effective Teaching*. Circle Pins, MN: American Guidance Services.

Dunn, K. J., and McCray, G. (2020). The place of the bifactor model in confirmatory factor analysis investigations into construct dimensionality in language testing. *Front. Psychol.* 11:1357. doi: 10.3389/fpsyg.2020.01357

Egan, S. J., Wade, T. D., and Shafran, R. (2011). Perfectionism as a transdiagnostic process: a clinical review. *Clin. Psychol. Rev.* 31, 203–212. doi: 10.1016/j.cpr.2010.04.009

Fletcher, K. L., Speirs Neumeister, K. L., and Flett, G. L. (2014). Introduction to the special issue: perfectionism in the school context. *Psychol. Sch.* 51, 897–898. doi: 10.1002/pits.21790

Flett, G. L., Blankstein, K. R., and Hewitt, P. L. (2009). Perfectionism, performance, and state positive affect and negative affect after a classroom test. *Can. J. Sch. Psychol.* 24, 4–18. doi: 10.1177/0829573509332457

Flett, G. L., and Hewitt, P. L. (2006). Positive vs. negative perfectionism in psychopathology: a comment on Slade and Owens's dual process model. *Behav. Modif.* 30, 472–495. doi: 10.1177/0145445506288026

Flett, G. L., and Hewitt, P. L. (2016). Still measuring perfectionism after all these years: reflections and an introduction to the special issue on advances in the assessment of perfectionism. *J. Psychoeduc. Assess.* 34, 615–619. doi: 10.1177/0734282916651540

Flett, G. L., Hewitt, P. L., and Hallett, C. J. (1995). Perfectionism and job stress in teachers. *Can. J. Sch. Psychol.* 11, 32–42. doi: 10.1177/082957359501100105

Friedman, I. A. (2000). Burnout in teachers: shattered dreams of impeccable professional performance. *J. Clin. Psychol.* 56, 595–606. doi: 10.1002/(sici)1097-4679(200005)56:5<595::aid-jclp2>3.0.co;2-q

Frost, R. O., Marten, P., Lahart, C., and Rosenblate, R. (1990). The dimensions of perfectionism. *Cogn. Ther. Res.* 14, 449–468. doi: 10.1007/BF01172967

Ghorbanzadeh, A., and Rezaie, G. (2016). The relationship between English language teacher perfectionism, efficacy, and burnout. *Int. J. For. Lang. Teach. Res.*. 4, 97–106. Available online at: http://jfl.iaun.ac.ir/article_563447.html (accessed January 14, 2020).

Gilman, R., and Ashby, J. S. (2006). "Perfectionism," in *Children's Needs III: Development, Prevention, and Intervention*, *3rd Edn*, eds G. G. Bear and K. M. Minke (Bethesda, MD: National Association of School Psychologists), 303–312.

Hewitt, P. L., and Flett, G. L. (1991). Perfectionism in the self and social contexts: conceptualization, assessment, and association with psychopathology. *J. Pers. Soc. Psychol.* 60, 456–470. doi: 10.1037/0022-3514.60.3.456

Hewitt, P. L., Flett, G. L., and Mikail, S. F. (2017). *Perfectionism: A Relational Approach to Conceptualization, Assessment, and Treatment*. New York, NY: The Guilford Press.

Hill, A. P., Appleton, P. R., and Mallinson, S. H. (2016). Development and initial validation of the Performance Perfectionism Scale for Sport (PPS-S). *J. Psychoeduc. Assess.* 34, 653–669. doi: 10.1177/0734282916651354

Hill, R. W., Huelsman, T. J., Furr, R. M., Kibler, J., Vicente, B. B., and Kennedy, C. (2004). A new measure of perfectionism: the perfectionism inventory. *J. Pers. Assess.* 82, 80–91, doi: 10.1207/s15327752jpa8201_13

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

Johnson, D. P., and Whisman, M. A. (2013). Gender differences in rumination: a meta-analysis. *Pers. Indiv. Differ.* 55, 367–374. doi: 10.1016/j.paid.2013.03.019

Jones, B. K. (2016). Enduring in an "impossible" occupation: perfectionism and commitment to teaching. *J. Teach. Educ.* 67, 437–446. doi: 10.1177/0022487116668021

Kenny, D. A., Kaniskan, B., and McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Soc. Method Res.* 44, 486–507. doi: 10.1177/0049124114543236

Kim, L. E., Dar-Nimrod, I., and MacCann, C. (2018). Teacher personality and teacher effectiveness in secondary school: personality predicts teacher support and student self-efficacy but not academic achievement. *J. Educ. Psychol.* 110, 309–323. doi: 10.1037/edu0000217

Lortie, D. C. (1975). *Schoolteacher, 2nd Edn*. Chicago, IL: The University of Chicago Press.

Lozano, L. M., Valor-Segura, I., García-Cueto, E., Pedrosa, I., Llanos, A., and Lozano, L. (2019). Relationship between child perfectionism and psychological disorders. *Front. Psychol*. 10:1855. doi: 10.3389/fpsyg.2019.01855

Macsinga, I., and Dobriţa, O. (2010). More educated, less irrational: gender and educational differences in perfectionism and irrationality. *Rom. J. Appl. Psychol.* 12, 2, 79–85. Available online at: http://www.rjap.psihologietm.ro/Download/rjap122_5.pdf (accessed January 14, 2020).

Maricuţoiu, L. P., Măgurean, S., and Tulbure, B. T. (2019). Perfectionism in a transdiagnostic context: an investigation of the criterion validity of the almost perfect scale-revised. *Eur. J. Psychol. Assess.* 36, 573–583. doi: 10.1027/1015-5759/a000541

Pelletier, L. G., Séguin-Lévesque, C., and Legault, L. (2002). Pressure from above and pressure from below as determinants of teachers' motivation and teaching behaviors. *J. Educ. Psychol.* 94, 186–196. doi: 10.1037/0022-0663.94.1.186

Rosseel, Y. (2012). Lavaan: an R package for structural equation modeling and more. Version 0.5-12 (BETA). *J. Stat. Softw*. 48, 1–36. doi: 10.18637/jss.v048.i02

Sadoughi, M. (2017). The relationship between personality traits, perfectionism and job burnout: the case of Iranian high-school teachers. *Int. J. Acad. Res. Prog. Educ. Dev*. 6, 64–77. doi: 10.6007/IJARPED/v6-i1/2576

Salmela-Aro, K., Hietajärvi, L., and Lonka, K. (2019). Work burnout and engagement profiles among teachers. *Front. Psychol.* 10:2254. doi: 10.3389/fpsyg.2019.02254

Schmidt, R. E., Courvoisier, D. S., Cullati, S., Kraehenmann, R., and Linden, M. V. D. (2018). Too imperfect to fall asleep: perfectionism, pre-sleep counterfactual processing, and insomnia. *Front. Psychol.* 9:1288. doi: 10.3389/fpsyg.2018.01288

Schruder, C. R., Curwen, T., and Sharpe, G. W. (2014). Perfectionistic students: contributing factors, impacts, and teacher strategies. *J. Edu. Soc. Behav. Sci.* 4, 139–155. doi: 10.9734/BJESBS/2014/5532

Shim, S., Cho, Y., and Knapke, M. (2020). Perils of perfectionistic concerns among teachers. *Psychol. Sch.* 57, 1116–1131. doi: 10.1002/pits.22384

Slaney, R. B., Rice, K. G., and Ashby, J. S. (2002). "A programmatic approach to measuring perfectionism: the almost perfect scales," in *Perfectionism: Theory, Research, and Treatment,* eds G. L. Flett and P. L. Hewitt (Washington, DC: American Psychological Association), 63–88. doi: 10.1037/10458-003

Smith, M. M., Sherry, S. B., Chen, S., Saklofske, D. H., Mushquash, C., Flett, G. L., et al. (2018). The perniciousness of perfectionism: a meta-analytic review of the perfectionism–suicide relationship. *J. Pers*. 86, 522–542. doi: 10.1111/jopy.12333

Starley, D. (2019). Perfectionism: a challenging but worthwhile research area for educational psychology. *Educ. Psychol. Pract*. 35, 121–146. doi: 10.1080/02667363.2018.1539949

Stoeber, J. (2017). *The Psychology of Perfectionism: Theory, Research, Applications*. London: Routledge. doi: 10.4324/9781315536255

Stoeber, J., Kempe, T., and Keogh, E. J. (2008). Facets of self-oriented and socially prescribed perfectionism and feelings of pride, shame, and guilt following success and failure. *Pers. Indiv. Differ.* 44, 1506–1516. doi: 10.1016/j.paid.2008.01.007

Stoeber, J., Madigan, D. J., and Gonidis, L. (2020). Perfectionism is adaptive and maladaptive, but what's the combined effect? *Personal. Individ. Diff.* 161:109846, doi: 10.1016/j.paid.2020.109846

Stoeber, J., and Otto, K. (2006). Positive conceptions of perfectionism: approaches, evidence, challenges. *Pers. Soc. Psychol. Rev.* 10, 295–319. doi: 10.1207/s15327957pspr1004_2

Stoeber, J., and Rennert, D. (2008). Perfectionism in school teachers: relations with stress appraisals, coping styles, and burnout. *Anxiety Stress Copin.* 21, 37–53. doi: 10.1080/10615800701742461

Stricker, J., Buecker, S., Schneider, M., and Preckel, F. (2019). Multidimensional perfectionism and the big five personality traits: a meta-analysis. *Eur. J. Personal.* 33, 176–196. doi: 10.1002/per.2186

Tangney, J. P. (2002). "Perfectionism and the self-conscious emotions: shame, guilt, embarrassment, and pride," in *Perfectionism: Theory, Research, and Treatment,* eds. G. L. Flett and P. L. Hewitt (Washington, DC: American Psychological Association), 199–215. doi: 10.1037/10458-008

Terry-Short, L. A., Owens, R. G., Slade, P. D., and Dewey, M. E. (1995). Positive and negative perfectionism. *Pers. Indiv. Differ*. 18, 663–668. doi: 10.1016/0191-8869(94)00192-U

Trigwell, K., and Prosser, M. (2004). Development and use of the approaches to teaching inventory. *Educ. Psychol. Rev.* 16, 409–424. doi: 10.1007/s10648-004-0007-9

Ulu, I. P., and Tezer, E. (2010). Adaptive and maladaptive perfectionism, adult attachment, and big five personality traits. *J. Psychol*. 144, 327–340. doi: 10.1080/00223981003784032

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 in Psychology

ORIGINAL RESEARCH
published: 03 June 2021
doi: 10.3389/fpsyg.2021.564011

Psychometric Properties of the Independent and Interdependent Self-Construal Questionnaire: Evidence From the Czech Republic

David Lacko[1,2]*, *Jiří Čeněk*[3,4] *and Tomáš Urbánek*[2]

[1] Department of Psychology, Faculty of Arts, Masaryk University, Brno, Czechia, [2] Institute of Psychology, Czech Academy of Sciences, Brno, Czechia, [3] Department of Social Studies, Faculty of Regional Development and International Studies, Mendel University in Brno, Brno, Czechia, [4] Department of Information and Library Studies, Faculty of Arts, Masaryk University, Brno, Czechia

OPEN ACCESS

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

Reviewed by:
Helena Carvalho,
University Institute of Lisbon (ISCTE),
Portugal
Fco. Pablo Holgado-Tello,
National University of Distance
Education (UNED), Spain

***Correspondence:**
David Lacko
david.lacko@mail.muni.cz

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

Received: 20 May 2020
Accepted: 07 May 2021
Published: 03 June 2021

Citation:
Lacko D, Čeněk J and Urbánek T
(2021) Psychometric Properties of the
Independent and Interdependent
Self-Construal Questionnaire:
Evidence From the Czech Republic.
Front. Psychol. 12:564011.
doi: 10.3389/fpsyg.2021.564011

This article introduces a validation study of the Czech version of an independent and interdependent self-construal questionnaire (SCS, Vignoles et al., 2016) conducted on 330 Czech subjects. In this study, the reliability, convergent validity and factor validity were verified. However, the confirmatory factor analysis revealed unsatisfactory factor structure (*RMSEA* = 0.053 [0.048, 0.057], *SRMR* = 0.080, *CFI* = 0.775, *TLI* = 0.755). These results are discussed with respect to other adaptations of individualism/collectivism scales in countries beyond typical West-East dichotomy. Hence, the article not only critically discusses the shortcoming of the Czech and original versions of the questionnaires, but also the general issues of the individualism-collectivism construct in the cross-cultural context as a whole.

Keywords: individualism, collectivism, independent self-construal, interdependent self-construal, confirmatory factor analysis, psychometric properties, factor structure

INTRODUCTION

Formulated in the 1970s by Hofstede, the cultural dimension of individualism/collectivism (I/C) has become a popular theoretical concept in cross-cultural psychology and a useful tool to structure and measure the psychological characteristics of members of various cultures (Bond, 2002). Consequently, I/C is used as a predictor for many other psychological and behavioral variables (Oyserman et al., 2002). The I/C dimension was originally defined at a national level as a single bipolar dimension. Hofstede (1983) defined individualism as the quality of a relationship between an individual and his or her immediate social environment (family, friends, community, etc.). The theory of independent (*de facto* individualism) and interdependent (*de facto* collectivism) self-construal (Markus and Kitayama, 1991) later became the dominant approach in individual I/C research. It is based on the "social orientation hypothesis" (Varnum et al., 2010), which states that cultures differ in social orientations and their development. While some (individualistic) cultures adopt an independent social orientation and tend to emphasize self-direction, self-expression and autonomy, other (collectivistic) cultures endorse the development of interdependent social orientation and emphasize harmony, relatedness and connection with others. Even though individualism is currently rising in most societies, the mentioned cross-cultural differences remain detectable (Santos et al., 2017).

Despite the popularity of the construct, there is unfortunately no widely accepted method of measuring the individual level of I/C. Oyserman et al. (2002) identified 27 I/C scales and performed a content analysis of I/C domains. They found seven individualism and eight collectivism components accounting for 88% of the items across the scales. None of these 27 scales can be considered a single standard of measurement. The overall agreement on I/C operationalization differs in the selected scales from component to component, which has drawn attention to the fragmentation of the concept of I/C and its operationalization.

The debate on the number of factors and their structure is still ongoing and the existing research has suggested that the concept of an independent and interdependent self might be one-dimensional (Hofstede, 1983), two-dimensional (Markus and Kitayama, 1991; Lu and Gilmour, 2007), three-dimensional (Kashima and Hardie, 2000; Noguchi, 2007), four-dimensional (Singelis et al., 1995), five-dimensional (Bartoš, 2010) or possibly even seven-dimensional (Vignoles et al., 2016).

Such ambiguity raises a question about the true underlying factor structure and therefore calls for further investigation on independent samples (Bollen, 1989b). The importance of this step is even more crucial in cross-cultural research, where securing the equivalence of constructs as well as the measurement invariance is necessary to be able to compare country or cultural group means (van de Vijver and Tanzer, 1997; van de Vijver and Leung, 2001; Čeněk and Urbánek, 2019). In order to acquire some evidence of structural equivalence, statistical methods such as Confirmatory Factor Analysis (CFA), Multi-Group Confirmatory Factor Analysis (MG-CFA), Measurement Invariance (MI), etc., need to be applied (Fischer and Karl, 2019). Performing CFA is vital in cross-cultural research as a first step in verifying the possibility of comparing results across different questionnaire translations. This article is focused mainly on this step. The next steps after setting the configural model should lie in constraining the factor structure, factor loadings and intercepts and thus verifying the configural, metric and scalar invariance measurement across various cultural groups.

INDIVIDUALISM AND COLLECTIVISM IN THE CZECH REPUBLIC

Several previous studies have already been conducted on Czech participants with mixed results. Hofstede's (Hofstede et al., 2010) approach assigned an index of 58 to the Czech Republic, suggesting a slightly above-average level of individualism. In cross-cultural comparisons, Czechs have been shown to be more individualistic than the rest of countries in Central Europe (Kolman et al., 2003). Dumetz and Gáboriková (2017) study confirmed that the Czech Republic is more individualistic than Slovakia, but others found the exact opposite result (Bašnáková et al., 2016). Furthermore, the Czech Republic seems to be less individualistic than the Netherlands (Bašnáková et al., 2016) and less collectivistic and, similarly, individualistic as East Asians (Lacko et al., 2020). Another study found that Czechs are not only more individualistic, but simultaneously also more collectivistic than Czech Vietnamese (Čeněk, 2015).

Unfortunately, these mixed results might be caused by the lack of valid and reliable tools to measure I/C, because adaptation attempts are relatively sparse for the Czech population. The first exception is an adaptation of INDCOL (Singelis et al., 1995) done by Bartoš (2010). However, his validation has a factor structure that does not fully correspond to the original scale, and its results are therefore not fully comparable in cross-cultural research. The psychometric properties of INDCOL are described in the next chapter.

The second exception, an attempt to adapt an I/C questionnaire into Czech, is a translation and cross-cultural verification of the Independent and Interdependent Self Scale (IISS; Lu and Gilmour, 2007) performed by Lacko and Čeněk (2020), who compared Czech and East Asian university students. Although the scale exhibited satisfactory reliability for both independent self-construal (Czech α = 0.815, Chinese α = 0.929) and interdependent self-construal (Czech α = 0.795, Chinese α = 0.906), the IISS showed configural non-invariance ($RMSEA$ = 0.043 [0.025, 0.057], $SRMR$ = 0.144, CFI = 0.636, TLI = 0.617). In their second study, performed on a larger sample consisting of only Czechs, they found similar unsatisfactory fit indices of IISS ($RMSEA$ = 0.064 [0.061, 0.066], $SRMR$ = 0.104, CFI = 0.460, TLI = 0.432).

The third exception is found in several adaptations of traditional methods measuring the cultural values, where individualism represents one of the cultural values. These Czech adaptations include VSM-94 (Value Survey Module 1994; Hofstede, 1994; adapted by Kolman et al., 2003), SVS (Schwartz Value Survey; Schwartz, 1992; adapted by Hnilica et al., 2006) and PVQ (Portraits Value Questionnaire; Schwartz et al., 2001; adapted in two studies by Řeháková, 2006; Anýžová, 2014). However, neither the original questionnaire manuals, nor the Czech adaptations of VSM-94 and SVS reported any relevant psychometric properties. The first study of PVQ reports only unsatisfactory internal reliability coefficients (α = 0.35–0.70; Řeháková, 2006). Even though the second study revealed acceptable MG-CFA fit indices (CFI = 0.924, $RMSEA$ = 0.017), it suggested insufficient MI results across countries (metric MI ΔCFI = 0.009, scalar MI ΔCFI = 0.155) and these results were applied only to 10 out of 23 countries in total (Anýžová, 2014), which therefore raises doubts about the validity and reliability of those instruments.

As mentioned above, the two-dimensional model of the IISS failed in the factor structure validation on the Czech sample, and Hofstede's one-dimensional model is claimed to be outdated and considered obsolete and invalid by many scholars (e.g., Singelis et al., 1995; McSweeney, 2002; Blodgett et al., 2008). Hence, the goal of this paper is to conduct an adaptation and psychometric analysis of a relatively new tool for individual level I/C measurement, the Self-Construal Scale (SCS; Vignoles et al., 2016; for a description see the Method section). Furthermore, we tried to verify its convergent validity with the current Czech adaptation of the Individualism-Collectivism Scale (INDCOL; Singelis et al., 1995; adapted by Bartoš, 2010).

MATERIALS AND METHODS

Participants

Data were collected from 330 Czech participants. This number of participants should be satisfactory for several reasons: (1) our proposed models are simple and they are composed only of several first-order factors and indicators (Kline, 1998); (2) even though the rules of thumb are not fully reliable while planning research, they usually indicate an amount of 100–200 participants as an absolute minimum (Little, 2013; Brown, 2015) and moreover, as Kline (1998) pointed out, the number of about 200 participants is not only very often used in the SEM framework, but it might also be reliable in certain circumstances; (3) no missing values were observed in our dataset (Brown, 2015; Kline, 1998); (4) I/C scales usually yield high internal-consistency reliability which decreases demands on sample size (Brown, 2015; Kline, 1998); (5) in single group models standard errors are significantly reduced with more than 150 responses (Little, 2013); and (6) our models yielded a huge number of degrees of freedom (Hoyle, 2012; Kline, 1998).

The research sample was 77% (n = 254) female. The participants were 18–65 years old (M = 24.29, SD = 6.536). Regarding the field of study, or vocation, 34.5% (n = 114) of the participants were psychologists/students of psychology, followed by students of/employees in the field of languages and history and of international studies (both 8.8%). Concerning the level of education, 58.8% (n = 194) had completed high school and 39.7% (n = 131) had achieved a university degree. Regarding the participants' religion, political preference and family status, most of them identified themselves as an atheist (43.6%), had no political preference (40.6%; or were liberals 31.2%) and were single (58.5%). The comprehensive descriptive statistics are shown in **Table 1**:

Procedure

Data collection was conducted between December 2018 - February 2019. The participants were mostly gathered through university social groups and social websites (i.e., the non-probability convenience sampling method) which resulted in a research sample with an over-represented student part of the population compared to other groups. The participants were informed about the ethical aspects of the research, especially data anonymization, their voluntary participation and the option to end the questionnaire at any time without giving a reason. In order to proceed further with the administration, they had to consent with their participation in the research. All items were administered randomly to avoid possible response biases caused by context influences and preceding questions (for review, see Uskul and Oyserman, 2006). The whole testing procedure took approximately 20 min.

In order to minimize any potential method bias caused by an imperfect translation procedure (van de Vijver and Hambleton, 1996) a parallel translation method was applied. The English original Self-Construal Scale (SCS) was independently translated by two bilinguals with backgrounds in social sciences. Both translations were subsequently compared and, in the case of any inconsistencies, discussed by the authors of the study until an agreement on the formulation was reached. We put a special emphasis on minimizing any potential shifts of meaning between the English and Czech versions of the scale. The Individualism-Collectivism Scale (INDCOL) was used in the original Czech version (Bartoš, 2010). Both scales were administered online. In addition to SCS and INDCOL,

TABLE 1 | Demographic characteristic of sample.

Variable	Choice	Frequency
Gender	Man	76 (23.03%)
	Woman	254 (76.97%)
Age	Range	18–65
	Mean (SD)/median (IQR)	24.293 (6.536)/23 (4)
Family status	Single	193 (58.485%)
	In partnership	105 (31.818%)
	Married	24 (7.273%)
	Divorced	7 (2.121%)
	Widow	1 (0.303%)
Education	Primary school	2 (0.606%)
	High school	194 (58.788%)
	Higher vocational school	3 (0.909%)
	University	131 (39.697%)
Field of study/occupation	Psychology	114 (34.545%)
	International studies	29 (8.788%)
	IT	12 (3.636%)
	Pedagogy	19 (5.758%)
	Regional development	16 (4.848%)
	Information studies and librarianship	9 (2.727%)
	Languages and history	29 (8.788%)
	Other	102 (30.909%)
Salary of family during childhood	1300 CZK and less	21 (6.364%)
	1300–6500 CZK	84 (25.455%)
	6500–13000 CZK	130 (39.394%)
	13000–33000 CZK	79 (23.939%)
	More than 33000 CZK	16 (4.848%)
Religion	Atheist	144 (43.636%)
	Christianity	79 (23.939%)
	Spiritually based person	92 (27.879%)
	Other	15 (4.545%)
Political opinions	No preference	134 (40.606%)
	Liberalism	103 (31.212%)
	Environmentalism and green politics	45 (13.636%)
	Conservatism	24 (7.273%)
	Socialism	10 (3.030%)
	Nationalism	5 (1.515%)
	Anarchy	4 (1.212%)
	Other	5 (1.515%)
Number of siblings	0	51 (15.455%)
	1	161 (48.788%)
	2	78 (23.636%)
	3	26 (7.879%)
	4 and more	14 (4.242%)

all relevant socio-demographic variables were collected (see **Table 1**).

Measures

The Self-Construal Scale (SCS)

The SCS was developed by Vignoles et al. (2016) and validated on 9573 (Study 1, n = 2294; Study 2, n = 7279) participants across 55 cultural groups in 33 nations. The SCS is primarily based on the concept of independent and interdependent self (Markus and Kitayama, 1991). The authors built on other traditional I/C scales during the formulation of its items (e.g., Singelis et al., 1995).

The SCS consists of thirty-eight, nine-point, Likert-type numerical items scaled from 1 (not at all) to 9 (exactly), with three intermediate anchor-points (3 – a little, 5 – moderately, 7 – very well). The SCS contains half reversed items, which should enhance the validity of the factor structure and minimize the acquiescence bias (Smith et al., 2013).

The authors used exploratory and confirmatory factor analytic techniques, MG-CFA, multilevel analysis and other statistical procedures (such as modeling acquiescence as a common method factor in CFA or ipsatization for reliability estimation) in their validation study. Even though the authors did not perform an analysis of MI, they discussed it in relation to the items' factor loadings, which in their opinion suggested a satisfactory invariance. However, no reliability estimation was performed in the validation study. The SCS was also tested for response biases in their follow-up study (cf. Smith et al., 2016) as well as concurrent validity with the I/C dimensions measured by individualism values and in-group collectivism practices, where the *r* coefficients were between 0.425 and 0.752.

Using Principal Components Analysis (PCA) and Principal Axis Factoring (PAF) in the first study, the authors identified seven dimensions of the SCS, namely "Self-reliance vs. Dependence on others," "Self-containment vs. Connection to others," "Difference vs. Similarity," "Commitment to others vs. Self-interest," "Consistency vs. Variability," "Self-direction vs. Receptiveness to influence" and "Self-expression vs. Harmony." The CFA partially confirmed the factor structure in the second study. The authors presented two respective models: model 1, which was comprised of 38 items, and model 2, with 26 items. The first model yielded good fit indices despite an insufficient *CFI* (*SRMR* = 0.050; *RMSEA* = 0.046; *CFI* = 0.790). However, the authors claimed that a 0.90 threshold for CFI is often unreachable and unrealistic in multidimensional questionnaires used in various cultural samples and they justified a *CFI* value near 0.80 as acceptable for cross-cultural multidimensional questionnaires. The second model showed better fit indices (*SRMR* = 0.033, *RMSEA* = 0.033, *CFI* = 0.922) and can be considered valid from the point of view of factor structure. Nevertheless, this model only has 26 items and one subscale is consequently comprised of only two items, whereas an unabbreviated model contains four to six items per subscale. Our view is that 26 items are insufficient for a seven-dimensional questionnaire, and we therefore focused on the longer version of the SCS. At the same time, we concur with the authors that the SCS is currently one of the most comprehensive tests available for I/C dimension measurement.

The Individualism-Collectivism Scale (INDCOL)

The INDCOL was introduced by Singelis et al. (1995) and later improved by Triandis and Gelfand (1998). The original scale contains 32 items, and the improved and shortened version contains 27 items. All items are nine-point, Likert-type questions. Both questionnaires measure four dimensions: horizontal collectivism (HC – empathy, cooperation, sociability), horizontal individualism (HI – independence, uniqueness, self-sufficiency), vertical collectivism (VC – submissiveness) and vertical individualism (VI – competitiveness). The validation study was conducted on 267 participants by Singelis et al. (1995). The reliability of scales was not ideal (HI α = 0.67, VI α = 0.74, HC α = 0.74, VC α = 0.68) nor were the CFA fit indices [$\chi 2$ (458) = 898.88, *GFI* = 0.79, *AGFI* = 0.75, *RMSR* = 0.089]. Based on the CFA results, the item pool was reduced from 94 to 32 items. The questionnaire was improved by Triandis and Gelfand (1998) on 543 participants in total (Study 1, n = 326; Study 2, n = 127; Study 3, n = 90). They selected 27 items with the highest factor loadings and also reported higher reliability coefficients (HI α = 0.81, VI α = 0.82, HC α = 0.80, VC α = 0.73). The 27 INDCOL items also showed good convergent and divergent validity through correlations with I/C scenarios. However, they did not repeat the CFA for the 27-item questionnaire, nor did they perform MI.

The Czech validation study was conducted by Bartoš (2010) on 1081 participants. He modified the nine-point, Likert type items to seven-points and reduced the number of items to 24. He applied Exploratory Factor Analysis (EFA) to examine the factor structure of the Czech version of the INDCOL and found five factors. He separated HI into HI1 (uniqueness; 3 items) and HI2 (independence; 2 items). VI (7 items), HC (7 items) and VC (5 items) remained the same as in the original study. He also conducted reliability estimates, but two scales did not meet the minimum criteria (VI α = 0.79, HI1 α = 0.71, HI2 α = 0.60, VC α = 0.63, HC α = 0.76). Although the Czech version of the scale seems to have limited psychometric properties, we decided to use it in this study for two reasons: (1) to verify its factor structure as reported by Bartoš (2010) using CFA on an independent Czech sample, and (2) to test its convergent validity with SCS, because it is, despite its limitations, the only available criteria for Czech samples.

Analytical Procedure

In order to examine the factor structure of both questionnaires, we performed a CFA with a robust, weighted, least square mean and variance (WLSMV) estimator, which is suitable for ordinal and non-Gaussian distributed data from Likert-type scales (Finney and DiStefano, 2013), because according to multivariate Henze-Zirkler tests, data were non-normally distributed at the subscale level (univariate Shapiro-Wilk tests confirmed these findings at the item level) for both questionnaires, and which is also less biased than robust maximum likelihood (MLR; Li, 2016). As the criteria for evaluating a good model fit, many more or less strict cut-offs are used. We used the Tucker-Lewis Index (*TLI*) $\geq$ 0.95, Comparative Fit Index (*CFI*) $\geq$ 0.95, Root Mean Square Error of Approximation (*RMSEA*) $\leq$ 0.60, Standardized Root Mean Square Residual (*SRMR*) $\leq$ 0.80 (Hu and Bentler, 1999) and Adjusted Goodness-of-Fit Index (*AGFI*) $\geq$ 0.90 (Hooper et al., 2008) fit indices for the evaluation of a good model fit in this study.

Internal consistency of subscales was assessed with Cronbach's α and McDonald's ω. We used the 0.70 threshold of internal consistency as a satisfactory indicator of reliability. We also performed a reliability analysis with ipsatization in order to reduce culture-specific response and acquiescence biases (Fischer, 2004; Fischer and Milfont, 2010). Standardized within-subject ipsative scores were calculated for each item of each individual according to the following formula:

$$ipsative\ score = \frac{response - \ M\ of\ scale\ for\ each\ individual}{SD\ of\ scale\ or\ each\ individual}.$$

Convergent validity between and within measures was verified with nonparametric Spearman's correlation analyses, while each subscale score was entered into analysis as arithmetic mean. We interpreted correlation coefficients higher than 0.50 as indicators of minimally acceptable convergent validity and coefficients higher than 0.70 as sufficient evidence for convergent validity (Carlson and Herdman, 2010). The statistical analysis was conducted in *R* (v 3.6.1; R Core Team, 2020), specifically the packages *lavaan* (Rosseel, 2012), *semTools* (Jorgensen et al., 2018), *psych* (Revelle, 2020), *ShinyItemAnalysis* (Martinkova and Drabinova, 2018), and *MVN* (Korkmaz et al., 2014).

RESULTS

The descriptive statistics (means, standard deviations, skewness and kurtosis) of all scales are shown in **Table 2**. The item analysis within classical test theory approach (i.e., descriptive statistics of items, several types of discrimination, etc.) is reported in **Supplementary Appendix II**.

Factor Structure

The Czech version of the SCS showed satisfactory *RMSEA* and *SRMR*. The relative chi-square (χ^2/df) was 1.913, which suggested a good global fit of the model (Kline, 1998). However, the model showed unsatisfactory *CFI* and *TLI* values. Nevertheless, we would like to emphasized that *CFI* of the Czech version of the SCS was almost the same as the *CFI* of the Vignoles et al. (2016) original version (see **Table 3**). A common factor with acquiescence as a common method factor was used on the reversed items following the procedure used by Vignoles et al. (2016) in order to reduce acquiescence bias (see Welkenhuysen-Gybels et al., 2003). This model also did not yield satisfactory fit indices.

Almost all of the items' factor loadings besides three instances were above the recommended 0.40 threshold (Fornell and Larcker, 1981). Even if we take into consideration the stricter thresholds, for instance 0.50 (Hair et al., 2018), we would obtain only four more such instances. Furthermore, the current factor loadings often being higher than the originals obtained by Vignoles et al. (2016; see **Supplementary Appendix I**). All item parameters, covariances (with two exceptions) and variances were statistically significant). Therefore, no items had to be removed from the model in order to improve its fit.

An analysis of the potential cross-loadings with a modification index (*mi*) and expected parameter change (*epc*) could bring deeper insight into model misfit. We found that item 32 (for items wording see **Supplementary Appendix I**) from "Self-expression vs. Harmony" had potential cross-loadings on subscales "Self-containment vs. Connection to others" (*mi* = 155.075, *epc* = 0.869), "Self-interest vs. Commitment to others" (*mi* = 148.989, *epc* = 1.048) and "Consistency vs. Variability" (*mi* = 87.303, *epc* = −0.627). Analogously, item 15 from Self-direction vs. Receptiveness to influence had potential cross-loadings on subscales "Self-containment vs. Connection to others" (*mi* = 73.090, *epc* = 1.079), "Consistency vs. Variability" (*mi* = 68.539, *epc* = −0.746) and "Self-interest vs. Commitment to others" (*mi* = 57.856, *epc* = 1.239). Item 35 from "Self-interest vs. Commitment to others" had potential cross-loadings on "Self-direction vs. Receptiveness to influence"

TABLE 2 | The descriptive statistics of subscale scores.

Scale	Subscale	*M* [95% *CI*]	*SD*	Skewness	Kurtosis
SCS	Difference vs. Similarity	5.59 [5.43, 5.75]	1.45	−0.143	−0.423
	Self-containment vs. Connection to others	4.28 [4.14, 4.43]	1.37	0.501	−0.122
	Self-direction vs. Receptiveness to influence	6.03 [5.88, 6.18]	1.40	−0.036	−0.726
	Self-reliance vs. Dependence on others	6.61 [6.44, 6.77]	1.50	−0.658	0.521
	Consistency vs. Variability	5.09 [4.91, 5.27]	1.67	−0.044	−0.460
	Self-expression vs. Harmony	5.02 [4.89, 5.16]	1.28	−0.121	−0.179
	Self-interest vs. Commitment to others	4.66 [4.51, 4.81]	1.36	0.412	−0.167
INDCOL	Vertical individualism	3.81 [3.68, 3.93]	1.15	0.240	−0.273
	Horizontal collectivism	5.21 [5.12, 5.31]	0.86	−0.498	0.247
	Vertical collectivism	3.41 [3.30, 3.52]	1.01	0.136	−0.244
	Horizontal individualism 1	4.88 [4.75, 5.01]	1.21	−0.623	−0.114
	Horizontal individualism 2	4.21 [4.07, 4.36]	1.34	0.107	−0.556

M, mean; SD, standard deviation; CI, confidence intervals.

TABLE 3 | The SCS and SCS modified model fit indices compared to the original version by Vignoles et al. (2016).

Model	Chi-Square	*p*	*RMSEA* [90% *CI*]	*SRMR*	*CFI*	*TLI*	*AGFI*
CZ SCS 1	χ^2 (644) = 1232.107	<0.001	0.053 [0.048, 0.057]	0.080	0.775	0.755	0.920
CZ SCS 2	χ^2 (625) = 1125.036	<0.001	0.049 [0.045, 0.054]	0.073	0.809	0.785	0.931
Original SCS	NR	NR	0.046 [NR]	0.050	0.790	NR	NR
CZ SCS Mod.	χ^2 (632) = 1011.010	<0.001	0.043 [0.038, 0.048]	0.066	0.855	0.839	0.943
CZ SCS (1 second-order factor)	χ^2 (658) = 1392.214	<0.001	0.058 [0.054, 0.062]	0.093	0.720	0.700	0.893
CZ SCS (bifactor)	χ^2 (606) = 975.433	<0.001	0.043 [0.038, 0.048]	0.060	0.859	0.836	0.951
CZ SCS (2 factors)	χ^2 (664) = 2013.118	<0.001	0.079 [0.075, 0.083]	0.121	0.485	0.454	0.824
CZ SCS 3	χ^2 (605) = 975.774	<0.001	0.043 [0.038, 0.048]	0.059	0.858	0.835	0.953

NR, not reported; CZ SCS 1, Czech version of SCS; CZ SCS 2, Czech version of SCS with one common factor with acquiescence as a common method factor on reversed items; Original SCS, original version of SCS; CZ SCS Mod., Czech version of SCS with allowed cross-loadings; CZ SCS (1 second-order factor), Czech version of SCS with one higher-order factor; CZ SCS (bifactor), Czech version of SCS with one general factor; CZ SCS (2 factors), Czech version of SCS two-dimensional model; CZ SCS 3, Czech version of SCS with two common factors with acquiescence as a common method factor separately on reversed and positive items; p, p-value; χ^2, chi-square; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; CFI, Comparative Fit Index; TLI, Tucker-Lewis Index; CI, confidence intervals; AGFI, adjusted goodness of fit index.

TABLE 4 | The INDCOL fit indices compared to the original version by Singelis et al. (1995).

Model	Chi-Square	*p*	*RMSEA* [90% CI]	*SRMR*	*CFI*	*TLI*	*AGFI*
CZ INDOCL (5 factors)	χ^2(242) = 521.126	<0.001	0.059 [0.052, 0.066]	0.074	0.779	0.748	0.932
CZ INDOCL (4 factors)	χ^2(246) = 537.226	<0.001	0.060 [0.053, 0.067]	0.077	0.769	0.741	0.929
Original INDCOL	χ^2(458) = 898.88	NR	0.089 [NR]	NR	NR	NR	0.75

NR, not reported; CZ INDCOL, Czech version of INDCOL; Original INDCOL, original version of INDCOL; p, p-value; χ^2, chi-square; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; CFI, Comparative Fit Index; TLI, Tucker-Lewis Index; CI, confidence intervals; ECVI, expected cross-validation index; AGFI, adjusted goodness of fit index.

(*mi* = 61.102, *epc* = 1.124), "Self-expression vs. Harmony" (*mi* = 43.283, *epc* = 0.789) and "Self-containment vs. Connection to others" (*mi* = 39.786, *epc* = −1.486). Finally, item 25 from "Consistency vs. Variability" had potential cross-loadings on subscales "Self-direction vs. Receptiveness to influence" (*mi* = 55.498, *epc* = 0.563), "Self-interest vs. Commitment to others" (*mi* = 49.360, *epc* = 0.453) and "Self-containment vs. Connection to others" (*mi* = 48.192, *epc* = 0.485). These findings are discussed in detail in the Discussion section.

The analysis of modification indices showed that the SCS contained multiple cross-loaded items. If all of the above-mentioned cross-loadings are included in the model (CZ SCS Mod.; see **Table 2** above), the majority of fit indices would be better than the original study by Vignoles et al. (2016). However, even these improved fit indices could still be considered unsatisfactory. Furthermore, using an exploratory approach (i.e., not confirmatory, cf. Bollen, 1989a; Byrne, 2010) we proposed four alternative models with individualism dimension as a one second-order factor of all subscales (i.e., CZ SCS 1 second-order factor), with individualism as a general factor in the bifactor model (i.e., CZ SCS bifactor), with individualism (non-reversed items) and collectivism subscales (reversed items in two-dimensional model (i.e., CZ SCS 2 factors) and with two common factors with acquiescence as a common method factor separately on reversed and positive items (i.e., CZ SCS 3, see **Table 2**). None of these models fit the data well.

The same CFA procedure was applied for the INDCOL scale with similar results as the SCS. The relative chi-square (χ^2/df) was 2.118, which suggested a good global fit. The *RMSEA* and *SRMR* were satisfactory, however, the *CFI* and *TLI* were not. Hence, the current CFA results did support neither the 5-factor configural model provided by Bartoš (2010) nor the original 4-factor model provided by Singelis et al. (1995; see **Table 4**).

Reliability

Concerning the reliability of the SCS, Cronbach's α varied between 0.667 and 0.855, while McDonald's ω fell between 0.651 and 0.854 (see **Table 5**). The values of both coefficients showed satisfactory internal consistency in most of the subscales. Three subscales were slightly below the minimum threshold of 0.70. In case of ipsatization, α varied between 0.265 and 0.786. The results with ipsative scores were less satisfactory than the raw score results. This suggested that the questionnaire's items might have been potentially influenced by a response bias (especially

TABLE 5 | Reliability estimations of the Czech version of the SCS subscales.

Dimension	α *SCS* (ipsatized)	ω *SCS*
Difference vs. Similarity	0.759 (0.571)	0.771
Self-containment vs. Connection to others	0.697 (0.634)	0.707
Self-direction vs. Receptiveness to influence	0.670 (0.265)	0.674
Self-reliance vs. Dependence on others	0.772 (0.624)	0.774
Consistency vs. Variability	0.855 (0.782)	0.854
Self-expression vs. Harmony	0.651 (0.291)	0.651
Self-interest vs. Commitment to others	0.763 (0.463)	0.764

α, *Cronbach's alpha;* ω, *McDonald's omega.*

TABLE 6 | Reliability estimations of the Czech version of the INDCOL subscales.

Dimension	α INDCOL (ipsatized)	ω INDCOL
Vertical individualism (VI)	0.812 (0.642)	0.814
Horizontal individualism 1 (HI1)	0.564 (0.432)	0.483
Horizontal individualism 2 (HI2)	0.502 (0.510)	0.510
Vertical collectivism (VC)	0.589 (0.435)	0.577
Horizontal collectivism (HC)	0.739 (0.632)	0.737

α, Cronbach's alpha; ω, McDonald's omega.

the subscales "Self-direction vs. Receptiveness to influence" and "Self-expression vs. Harmony").

Concerning the reliability of the INDCOL, α varied between 0.502 and 0.812 (for ipsative α between 0.432 and 0.642), while ω fell between 0.483 and 0.814 (see **Table 6**). The values of both coefficients demonstrated satisfactory internal consistency only for the VI and HC subscales. VC and HI demonstrated unsatisfactory internal consistency, and the ipsative scores suggested that they might have been influenced by a response bias.

Convergent Validity Between Measures

In the following section, the results of the correlation analyses between subscales of the original and adapted version of SCS, between subscales of the original and adapted version of INDCOL, and between scales of SCS and INDOL are reported. A comparison of Spearman's ρ to the original correlation coefficients (by Vignoles et al., 2016) in the subscales is shown in **Table 7**. All differences between obtained and original coefficients were smaller than 0.250. Besides a few exceptions, the Czech version showed relatively similar patterns of correlations to the original version. All of these associations were statistically significant and ranged from small to medium effect sizes (with exceptions of two insignificant associations and one association with high effect size).

A comparison of the Spearman's ρ correlations and original correlations of the Czech INDCOL version among the subscales is shown in **Table 8**. Even though current correlations were generally higher than correlations reported by Bartoš (2010), the corelations coefficients were still rather small. We also observed three differences between original coefficients and coefficients obtained in this study which were higher than 0.250. However, our results appear to be more in line with the I/C theory, because negative correlations between the HC (collectivism) and individualistic subscales (VI and HI2) were observed (instead of positive as reported by Bartoš, 2010).

Relationships were also expected between the SCS subscales and the INDCOL subscales as a demonstration of convergent validity. We assumed that HC and VC (i.e., collectivism) should be negatively correlated with all SCS subscales, whereas HI1, HI2 and VI (i.e., individualism) should correlate positively. As shown in **Table 9**, our expectations about directions were confirmed. However, these r_s coefficients were relatively small, and some of them non-significant. Only four associations were higher than 0.50 threshold ("difference vs. similarity" and "horizontal individualism: uniqueness"; "self-direction vs. receptiveness

TABLE 7 | Comparison of the estimated correlations in the Czech version (below diagonal) and the original version (above diagonal) of the SCS by Vignoles et al. (2016).

Dimension	1	2	3	4	5	6	7
1. Difference vs. Similarity	–	0.112	0.288	0.436	0.136	0.401	0.214
2. Self-containment vs. Connection to others	0.143** [0.035, 0.247]	–	0.625	−0.075	−0.219	0.330	0.557
3. Self-direction vs. Receptiveness to influence	0.299*** [0.197, 0.394]	0.405*** [0.311, 0.492]	–	0.328	−0.002	0.417	0.435
4. Self-reliance vs. Dependence on others	0.240*** [0.135, 339]	0.125* [0.017, 0.229]	0.479*** [0.391, 0.558]	–	0.301	0.132	0.104
5. Consistency vs. Variability	0.267*** [0.164, 0.364]	−0.008 [−0.116, 0.100]	0.243*** [0.139, 0.342]	0.193*** [0.086, 0.294]	–	0.252	−0.141
6. Self-expression vs. Harmony	0.450*** [0.360, 0.532]	0.246*** [0.142, 0.345]	0.424*** [0.331, 0.509]	0.180** [0.074, 0.283]	0.299*** [0.198, 0.394]	–	0.366
7. Self-interest vs. Commitment to others	0.334*** [0.234, 0.426]	0.619*** [0.547, 0.681]	0.468*** [0.379, 0.548]	0.300*** [0.199, 0.395]	0.043 [−0.065, 0.150]	0.443*** [0.352, 0.526]	–

$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

TABLE 8 | Comparison of the estimated correlations in the current (below diagonal) and original study (above diagonal) of the Czech version of INDCOL by Bartoš (2010).

Dimension	VI	HI1	HI2	VC	HC
VI	–	NR	NR	NR	0.19***
HI1	0.292*** [0.190, 0.388]	–	NR	−0.13***	−0.12***
HI2	0.241*** [0.136, 0.340]	0.112* [0.004, 0.217]	–	−0.13***	0.11***
VC	−0.083 [−0.189, 0.025]	−0.214*** [−0.314, −0.108]	−0.214*** [−0.315, −0.109]	–	NR
HC	−0.217*** [−0.318, −0.112]	−0.157** [−0.261, −0.050]	−0.308*** [−0.403, −0.207]	0.402*** [0.308, 0.489]	–

**p < 0.05, **p < 0.01, ***p < 0.001; NR, not reported.*

to influence" and "vertical collectivism"; "self-containment vs. connection to others" and "horizontal collectivism"; and "self-interest vs. commitment to others" and "horizontal collectivism"); none of them were above the recommended 0.70 value. Hence, these results do not support the assumption of the convergent validity of the SCS and INDCOL. It appears that both scales measure slightly different constructs.

DISCUSSION

The Psychometric Properties of the SCS

A validation study of the SCS was conducted and the psychometric properties of the SCS were examined. As Bollen (1989b) pointed out, the replication on independent samples is the only way to check whether original associations are a sampling fluke or not. He also emphasized the necessity of such research, because despite the fact that replications are often considered very valuable, such studies appear far too seldom. This type of research therefore serves as a contribution to the cross-cultural examination of the I/C concept and its results are necessary for a deeper understanding of I/C across various cultures.

In summary, both questionnaires demonstrated limitations in their reliability and validity. These shortcomings could have stemmed from the lack of reliability and validity of the original versions (Singelis et al., 1995; Triandis and Gelfand, 1998; Bartoš, 2010; Vignoles et al., 2016) rather than our cross-cultural adaptation.

In more detail, despite that Czech version of the SCS showed satisfactory reliability in four subscales (the rest of the subscales were only slightly below the 0.70), similar correlations between subscales were observed and similar fit indices were obtained in comparison with the original study (Vignoles et al., 2016). Additionally, our study revealed several crucial psychometric shortcomings of the scale which suggest its insufficient validity and reliability. This might be to some extent a consequence of the psychometric properties of the original instrument.

Four main issues were identified. First, the reliability estimation with ipsative scores showed poor internal consistency suggesting that the SCS may be influenced by response biases. Second, the CFA results were unsatisfactory, and therefore cross-cultural comparisons using this questionnaire might be biased, non-invariant and invalid (Fischer and Karl, 2019). Both the original and Czech versions of the SCS have serious shortcomings in their factor structure, as suggested by, for example, the *CFI*.

Consequently, the third issue stemmed from an analysis of the modification indices, which revealed some cross-loadings. For example, the item "*You follow your personal goals even if they are very different from the goals of your family*" might saturate not only "Self-interest vs. Commitment to others," but also "Self-containment vs. Connection to others," "Self-direction vs. Receptiveness to influence" and "Self-expression vs. Harmony." This finding seems logical, because a person who answers negatively to the mentioned item is probably not only more committed to others, but also leans toward harmony and receptiveness to an influence and is more connected to others. The presented analysis suggests that many items have similar cross-loadings. We believe that this is probably more likely caused by the poor theoretical background in the latent variables than vague and ambiguous item wording. Consequently, SCS factors are vaguely defined and lack divergent validity because they are based primarily on psychometric results. Therefore, even simply worded items (e.g., the item "*You always ask your family for advice before making a decision*") have potential cross-loadings on other subscales. Furthermore, the semantic qualities of some factors seem to be quite similar (e.g., "Self-direction vs. Receptiveness to influence" and "Self-interest vs. Commitment to others"), and perhaps may be adequate to reduce the number of I/C factors. Although the process of validation of a multidimensional cross-cultural questionnaire like SCS is very tedious, it should not be limited just to the psychometric evaluation of factor structure, model fit, reliability, etc., but it should also be theoretically well grounded.

And four, despite that all directions of relationships between SCS and INDCOL were as expected, i.e., the dimensions of horizontal and vertical collectivism correlated with interdependent self, whereas the dimensions of horizontal and vertical individualism correlated with independent self, the correlation coefficients were mostly small or moderate. This suggests, that both scales probably measure slightly different and insufficiently related constructs The above-mentioned issues with the SCS lead us to questions about the I/C concept itself, because similar issues were observed in multiple previous studies (see below).

General Issues of the I/C Concept

Research of I/C has been criticized by many scholars. Generally, there is no questionnaire in the literature measuring I/C that repeatedly meets the demanding requirements of cross-cultural research (i.e., CFA, MG-CFA, MI across different cultures, controlling for response bias, etc.). Many studies do

TABLE 9 | Coefficients of Spearman's rho in the SCS and INDCOL.

Dimension	VI	HI1	HI2	VC	HC
Difference vs. Similarity	0.269*** [0.166, 0.366]	0.586*** [0.510, 0.652]	0.053 [−0.055, 0.160]	−0.347*** [−0.439, −0.249]	−0.098 [−0.204, 0.010]
Self-containment vs. Connection to others	0.154** [0.047, 0.258]	0.160*** [0.053, 0.263]	0.273*** [0.170, 0.370]	−0.389*** [−0.477, −0.293]	−0.561*** [−0.631, −0.482]
Self-direction vs. Receptiveness to influence	0.086 [−0.023, 0.192]	0.234*** [0.130, 0.334]	0.290*** [0.186, 0.386]	−0.552*** [−0.623, −0.473]	−0.285*** [−0.381, −0.182]
Self-reliance vs. Dependence on others	0.210*** [0.104, 0.311]	0.277*** [0.174, 0.374]	0.263*** [0.159, 0.361]	−0.324*** [−0.417, −0.223]	−0.155*** [−0.258, −0.047]
Consistency vs. Variability	−0.061 [−0.168, 0.047]	0.090 [−0.018, 0.196]	0.062 [−0.046, 0.169]	−0.139* [−0.244, −0.032]	0.099 [−0.009, 0.205]
Self-expression vs. Harmony	0.204*** [0.098, 0.305]	0.228*** [0.123, 0.328]	0.139* [0.032, 0.243]	−0.496*** [−0.573, −0.410]	−0.162** [−0.265, −0.055]
Self-interest vs. Commitment to others	0.370*** [0.273, 0.459]	0.336*** [0.237, 0.429]	0.325*** [0.225, 0.419]	−0.493*** [−0.571, −0.407]	−0.513*** [−0.589, −0.429]

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

not sufficiently conduct or report these important psychometric properties, and do not conduct any adequate multi-level analysis (Oyserman et al., 2002; Levine et al., 2003a,b; Chen and West, 2008; Cozma, 2011). Another relevant critique argues that the I/C research ignores or even lacks concurrent and discriminant validity of the scales (Oyserman et al., 2002; Levine et al., 2003b; Bresnahan et al., 2005; Schimmack et al., 2005). Furthermore, the conceptual unclarity of the I/C research is also often criticized (e.g., Oyserman et al., 2002; Voronov and Singer, 2002; Brewer and Chen, 2007; Oyserman and Lee, 2008).

As mentioned above, the original validation studies of INDCOL (Singelis et al., 1995; Triandis and Gelfand, 1998), SCS (Vignoles et al., 2016), VSM (Hofstede, 1994), SVS (Schwartz, 1992), and IISS (Lu and Gilmour, 2007) did not meet the minimum criteria of CFA or did not even perform such a procedure. Similar patterns can be found in other I/C questionnaires, for example in the W-M (Wagner and Moch, 1986), COS (Communal Orientation Scale; Clark et al., 1987), RISC (Relational-Interdependent Self-Construal; Cross et al., 2000), AICS (Auckland Individualism and Collectivism Scale; Shulruf et al., 2007) and ICIA scales (Collectivism Interpersonal Assessment Inventory; Matsumoto et al., 1997).

On the other hand, there are also several exceptions. For example, the Human Relations Questionnaire (HRQ; Noguchi, 2007) identifies three factors, namely the focus on others, helping others and self-focus, with satisfactory CFA indices: $\chi^2(24) = 49.93$, $GFI = 0.973$, $CFI = 0.964$, $RMSEA = 0.052$. Nevertheless, the author did not perform MG-CFA and MI analysis across the United States and Japanese versions, and its internal consistency was also not sufficient (α = 0.44-0.76). Three factors were also identified in the RIC scale (Relational, Individual and Collective Self-Aspects; Kashima and Hardie, 2000), namely relational, individual and collective self-aspects. The RIC scale showed satisfactory CFA fit indices [$\chi^2(24) = 79.26$, $CFI = 0.94$, $NFI = 0.91$]; however, the CFA was supported by a nine-item model (only 3 items per subscale), not the original 30-item model ($CFI = 0.72$).

Other examples can be the PVQ (Schwartz et al., 2001), or the Collective Self-Esteem Scale (CSES, Luhtanen and Crocker, 1992). The PVQ provided an acceptable MG-CFA but insufficient MI results (Anýžová, 2014), while the four-dimensional CSES reached close to satisfactory CFA results (NFI = 0.74–0.91, TLI = 0.84–0.92, CFI = 0.87–0.93). However, the CSES factor structure was not confirmed in the African American ethnic group, and therefore we can doubt its usability in a cross-cultural equivalence (Utsey and Constantine, 2006). Additionally, both questionnaires also focus primarily on different constructs that only possess a partial overlap with the I/C concept, i.e., cultural values in the case of PVQ and collective self-esteem in the case of CSES.

In this paper we conducted an attempt to validate an adaptation of a relatively new I/C scale on a Czech sample which is a sample not fitting into the group of West and East countries (such as the United States, England, Japan, China) that are studied the most often in the field. In this section we want to provide information about other similar research going beyond this dichotomy, both successful and unsuccessful ones. We omit adaptations without a CFA procedure or its adequate equivalent, which unfortunately represents the vast majority of studies (Chen and West, 2008). I/C scales were already successfully adapted for instance in Turkey (e.g., Li and Aksoy, 2006; Akın et al., 2010), Jordan, Lebanon, and Syria (e.g., Harb and Smith, 2008), Hong Kong and Ghana (e.g., Affum-Osei et al., 2019) or Switzerland and South Africa (results were satisfactory only for one of two used I/C scales; see Györkös et al., 2012).

Nevertheless, these successful attempts are relatively rare compared to the amount of studies that failed to do so. In spite of the fact that the authors themselves often interpreted the quality of the adaptations as satisfactory and part of studies is indeed methodologically and statistically sound, a deeper inspection reveals issues in the factor structure of the adapted scales. The adaptations usually did not yield satisfactory fit indices. In some cases, the number of factors and items were substantially changed compared to the original scales. Despite the fact that these model modifications and changes lead in some cases to the satisfactory fit indices of the "new" scales, this rather data-driven approach needs to be considered exploratory (see e.g., Bollen, 1989a; Byrne, 2010).

Similar psychometric problems with the adaptation of the I/C scale as in the current study were observed for instance in Poland (cf. Pilarska, 2011), Spain (cf. Gouveia et al., 2003), India (cf. Sivadas et al., 2008), Malaysia (cf. Miramontes, 2011; Ramley et al., 2020), Mexico (cf. Miramontes, 2011), Singapore (cf. Soh and Leong, 2002), Italy (cf. Bobbio and Sarrica, 2009; D'Amico and Scrima, 2015; Germani et al., 2020), France (cf. Gibas et al., 2016), Philippines (cf. Miramontes, 2011; Bernardo et al., 2012; Datu, 2014), Australia (cf. Freeman and Bordia, 2001; Miramontes, 2011), Portugal (cf. Gonçalves et al., 2017), Thailand (cf. Christopher et al., 2011) or Argentina (cf. Chiou, 2001).

The current, rather unsatisfactory results might have deeper causes than just the psychometric quality of the original SCS scale. Even though past studies assumed I/C being a stable cross-cultural construct with an ambition to categorize nations along the collectivistic and individualistic spectrum, these assumptions were not entirely confirmed. It seems that the relatively simplistic East-West dichotomy doesn't truly exist (Matsumoto, 1999; Takano and Osaka, 1999, 2018; Heine et al., 2002; Oyserman et al., 2002; Levine et al., 2003a,b), and the I/C construct is far less stable than assumed (Yamagishi, 1988; Gardner et al., 1999; Oyserman and Lee, 2008). Consequently, some authors with respect to the previously mentioned shortcomings and critiques of I/C research came to the conclusion that the concept of I/C itself does not exist and suggest not using it in research (e.g., Levine et al., 2003a,b). Therefore, doubts about the validity of using I/C as a predictor of other constructs in current cross-cultural studies should be raised.

Limitations and Future Directions

The results of our study are based on the unrepresentative sample gained through the non-probability convenience sampling method which resulted in various imbalances of demographic characteristics, especially in the overrepresentation of women, young participants and participants with a university education. An analysis performed on different populations might result in a different factor structure. However, I/C research usually validates the scales on samples of university students; for example, INDCOL (Singelis et al., 1995; Triandis and Gelfand, 1998) was created on the basis of such samples and SCS (Vignoles et al., 2016) used this population in the first phase of their validation study. Furthermore, despite the fact that sample size was considered satisfactory, SEM usually needs large samples and therefore is an *a priori* power analysis based on model simulations highly recommended (Hoyle, 2012; Little, 2013; Brown, 2015; Kline, 1998), which was not performed in the current study.

The future research should, in the first step, focus on a redefinition and reconceptualization of the I/C construct (e.g., Oyserman et al., 2002; Voronov and Singer, 2002; Brewer and Chen, 2007; Oyserman and Lee, 2008), while it is quite possible that such redefinition would not be universal for different cultural groups. Consequently, after the theoretical clarification of the I/C concept, the main aim of the future research should lie in a sounder methodological and statistical approach such as routinely using SEM techniques.

One of the possible ways to achieve this is the development of a new self-report instrument with satisfactory psychometric properties with the potential to be adapted in multiple cultures (Schimmack et al., 2005; Chen and West, 2008; Cozma, 2011). An important characteristic of this instrument would be its resistance to a reference-group effect (see Heine et al., 2002). Additionally, such an instrument would need to yield satisfactory results in repeated replications on independent samples from both the same and other cultures (Bollen, 1989b). Furthermore, adequate statistical approaches need to be used while comparing means across various cultural groups, such as MG-CFA with scalar measurement invariance (see Fischer and Karl, 2019). The research needs to be robust enough to cover the whole spectrum of variables that can potentially affect the level of I/C in order to reduce the cultural attribution fallacy (i.e., unpacking studies; see Matsumoto and Juang, 2013). Since the validation procedure usually does not end with one (un)successful validation study, but it represents an iterative process of bringing new evidence of validity and reliability the research in the field is far from concluded. However, we believe that without such an approach it is not possible to bring valid information about the real nature of I/C in culturally diverse populations via self-report scales.

The second possible approach to solve the current unsatisfactory situation in I/C research could lie in a shift from quantitative self-report questionnaires based on verbal responses to the usage of entirely different group of methods (e.g., Matsumoto, 1999; Bond, 2002; Fiske, 2002; Heine et al., 2002). For example, Talhelm et al. (2018) observed the differences in I/C with an observational design of the real-life behavior of participants; Partikova (2019) used interpretative phenomenological analysis of semi-structured interviews, and Hsu and Barker (2013) identified differences in I/C using the content analysis of TV advertisements. Furthermore, meta-analysis of other "cultural products" by Morling and Lamoreaux (2008) revealed higher effect sizes than meta-analyses performed on self-report scales. Another example can be found in the work of Klein et al. (2018) who created a new "WEIRDness score" (WEIRD: Western, educated, industrialized, rich and democratic; see Henrich et al., 2010) which might be included into multi-group statistical analysis as a predictor in a similar fashion as Hofstede's dimensions. We believe that it might be possible to create a similar country-level index specifically related to the I/C. Maybe not on the basis of self-report questionnaire data, but rather from an in-depth qualitative analysis of several indicators and consequent inter-rater agreement of experts from various cultures.

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 The Research Ethics Committee of Masaryk University (Ref. No.: EKV-2018-011, Proposal No.: 0257/2018). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was implied via completion of the questionnaires.

AUTHOR CONTRIBUTIONS

DL contributed to data collection, article drafting, and data analysis. JČ contributed to questionnaire adaptation, data collection, and article drafting. TU contributed to article revisions and data analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Czech Science Foundation under Grant GC19-09265J (The influence of socio-cultural factors and writing system on perception and cognition of complex visual stimuli).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.564011/full#supplementary-material

REFERENCES

Affum-Osei, E., Aboagye, M. O., Antwi, C. O., and Asante, E. A. (2019). Validating the Auckland individualism–collectivism scale (AICS): testing factor structure and measurement invariance in Hong Kong and Ghanaian Samples. *Psychol. Stud.* 64, 187–199. doi: 10.1007/s12646-019-00494-2

Akın, A., Eroğlu, Y., Kayış, A. R., and Satıcı, S. A. (2010). The validity and reliability of the Turkish version of the relational-interdependent self-construal scale. *Proc. Soc. Behav. Sci.* 5, 579–584. doi: 10.1016/j.sbspro.2010.07.145

Anýžová, P. (2014). SrovnatelnostSchwartzovyhodnotovéškály v mezinárodníchdatech [The Comparability of Schwartz's Human Values Scale in International Data]. *Czech. Sociol. Rev.* 50, 547–580. doi: 10.13060/00380288.2014.50.4.108

Bartoš, F. (2010). Individualismus a kolektivismus v česképopulaci a jejichsouvislosti s narcismem [Individualism and Collectivism of the Czech Population and their Relation to Narcissism]. *Sociol. Slovak Sociol. Rev.* 42, 134–161.

Bašnáková, J., Brezina, I., and Masaryk, R. (2016). Dimensions of culture: the case of Slovakia as an outlier in Hofstede's research. *Cesk. Psychol.* 60, 13–25.

Bernardo, A. B. I., Lising, R. L. S., and Shulruf, B. (2012). Validity of two language versions of the Auckland individualism and collectivism scale with Filipino-english bilinguals. *Psychol. Stud.* 58, 33–37. doi: 10.1007/s12646-012-0172-8

Blodgett, J. G., Bakir, A., and Rose, G. M. (2008). A test of the validity of Hofstede's cultural framework. *J. Consum. Mark.* 25, 339–349. doi: 10.1108/07363760810902477

Bobbio, A., and Sarrica, M. (2009). Horizontal and vertical individualism and collectivism: an Italian adaptation of Singelis et al.'s scale and its relations with conflict management and leadership styles. *TPM Test. Psychom. Methodol. Appl. Psychol.* 16, 209–226.

Bollen, K. A. (1989b). *Structural Equations with Latent Variables*. New York, NY: John Wiley & Sons.

Bollen, K. A. (1989a). A new incremental fix index for general structural equation models. *Sociol. Methods. Res.* 17, 303–316. doi: 10.1177/0049124189017003004

Bond, M. (2002). Reclaiming the individual from Hofstede's ecological analysis: a 20-year odyssey: comment on Oyserman et al. (2002). *Psychol. Bull.* 128, 73–77. doi: 10.1037/0033-2909.128.1.73

Bresnahan, M. J., Levine, T. R., Shearman, S. M., Lee, S. Y., Park, C.-Y., and Kiyomiya, T. (2005). A multimethod multitrait validity assessment of self-construal in Japan, Korea, and the United States. *Hum. Commun. Res.* 31, 33–59. doi: 10.1111/j.1468-2958.2005.tb00864.x

Brewer, M. B., and Chen, Y.-R. (2007). Where (Who) are collectives in collectivism? Toward conceptual clarification of individualism and collectivism. *Psychol. Rev.* 114, 133–151. doi: 10.1037/0033-295x.114.1.133

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research*. 2nd Edn. New York, NY: The Guilford Press.

Byrne, B. M. (2010). *Structural Equation Modeling with Amos: Basic Concepts, Applications and Programming*. London: Routledge/Taylor & Francis Group.

Carlson, K. D., and Herdman, A. O. (2010). Understanding the impact of convergent validity on research results. *Organ. Res. Methods* 15, 17–32. doi: 10.1177/1094428110392383

Čeněk, J. (2015). Cultural dimension of individualism and collectivism and its perceptual and cognitive correlates in cross-cultural research. *J. Educ. Cult. Soc.* 2015, 210–225. doi: 10.15503/jecs20152.210.225

Čeněk, J., and Urbánek, T. (2019). The adaptation and equivalence of test methods: an inspiration for psychological assessment of minorities in the Czech Republic. *Cesk. Psychol.* 63, 42–54.

Chen, F. F., and West, S. G. (2008). Measuring individualism and collectivism: the importance of considering differential components, reference groups, and measurement invariance. *J. Res. Pers.* 42, 259–294. doi: 10.1016/j.jrp.2007.05.006

Chiou, J. S. (2001). Horizontal and vertical individualism and collectivism among college students in the United States, Taiwan, and Argentina. *J. Soc. Psychol.* 141, 667–678. doi: 10.1080/00224540109600580

Christopher, M. S., Norris, P., D'Souza, J. B., and Tiernan, K. A. (2011). A test of the multidimensionality of the self-construal scale in Thailand

and the United States. *J. Cross. Cult. Psychol.* 43, 758–773. doi: 10.1177/0022022111406119

Clark, M., Ouellette, R., Powell, M., and Milberg, S. (1987). Recipient's mood, relationship type, and helping. *J. Pers. Soc. Psychol.* 53, 94–103. doi: 10.1037/0022-3514.53.1.94

Cozma, I. (2011). How are individualism and collectivism measured? *Rom. J. Appl. Psychol.* 13, 11–17.

Cross, S. E., Bacon, P. L., and Morris, M. L. (2000). The relational-interdependent self-construal and relationships. *J. Pers. Soc. Psychol.* 78, 791–808. doi: 10.1037/0022-3514.78.4.791

Datu, J. A. D. (2014). Validating the revised self-construal scale in the Philippines. *Curr. Psychol.* 34, 626–633. doi: 10.1007/s12144-014-9275-9

Dumetz, J., and Gáboriková, E. (2017). The Czech and Slovak Republics: a cross-cultural comparison. *Marketing Sci. Inspir.* 11, 2–13.

D'Amico, A., and Scrima, F. (2015). The Italian validation of Singelis's self-construal scale (SCS): a short 10-item version shows improved psychometric properties. *Curr. Psychol.* 35, 159–168. doi: 10.1007/s12144-015-9378-y

Finney, S. J., and DiStefano, C. (2013). "Nonnormal and categorical data in structural equation modelling," in *Quantitative Methods in Education and the Behavioral Sciences: Issues, Research, and Teaching. Structural Equation Modeling: A Second Course*, eds G. R. Hancock and R. O. Mueller (Charlotte: IAP Information Age Publishing), 439–492.

Fischer, R., and Karl, J. A. (2019). A primer to (Cross-Cultural) multi-group invariance testing possibilities in R. *Front. Psychol.* 10:1507. doi: 10.3389/fpsyg.2019.01507

Fischer, R., and Milfont, T. L. (2010). Standardization in psychology research. *Int. J. Psychol. Res.* 3, 88–96. doi: 10.21500/20112084.852

Fischer, R. (2004). Standardization to account for cross-cultural response bias: a classification of score adjustment procedures and review of research in JCCP. *J. Cross. Cult. Psychol.* 35, 263–282. doi: 10.1177/0022022104264122

Fiske, A. P. (2002). Using individualism and collectivism to compare cultures–a critique of the validity and measurement of the constructs: comment on Oyserman et al. (2002). *Psychol. Bull.* 128, 78–88. doi: 10.1037/0033-2909.128.1.78

Fornell, C., and Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18, 39–50. doi: 10.2307/3151312

Freeman, M. A., and Bordia, P. (2001). Assessing alternative models of individualism and collectivism: a confirmatory factor analysis. *Eur. J. Pers.* 15, 105–121. doi: 10.1002/per.398

Gardner, W., Gabriel, S., and Lee, A. (1999). "I" value freedom, but "we" value relationships: self-construal priming mirrors cultural differences in judgment. *Psychol. Sci.* 10, 321–326. doi: 10.1111/1467-9280.00162

Germani, A., Delvecchio, E., Li, J. B., and Mazzeschi, C. (2020). The horizontal and vertical individualism and collectivism scale: early evidence on validation in an Italian sample. *J. Child Fam. Stud.* 29, 904–911. doi: 10.1007/s10826-019-01571-w

Gibas, D., Giraud, T., Le Conte, J., Rubens, L., Martin, J. C., and Isableu, B. (2016). Attempt to validate the self-construal scale in French: systematic approach and model limitation. *Eur. Rev. Appl. Psychol.* 66, 85–93. doi: 10.1016/j.erap.2016.02.001

Gonçalves, G., Sousa, C., Santos, J., Gomes, A., Santa-Rita, A., Hipólito, S., et al. (2017). Psychometric properties of the Portuguese version of the relational interdependent self–construal scale. *Psicol. Rev. Associação Portuguesa Psicol.* 31, 121–136. doi: 10.17575/rpsicol.v31i2.1264

Gouveia, V. V., Clemente, M., and Espinosa, P. (2003). The horizontal and vertical attributes of individualism and collectivism in a Spanish population. *J. Soc. Psychol.* 143, 43–63. doi: 10.1080/00224540309598430

Györkös, C., Becker, J., Massoudi, K., Antonietti, J.-P., Pocnet, C., de Bruin, G. P., et al. (2012). Comparing the horizontal and vertical individualism and collectivism scale and the Auckland individualism and collectivism scale in two cultures. *Cross. Cult. Res.* 47, 310–331. doi: 10.1177/1069397112470371

Hair, J. F., Babin, B. J., Anderson, R. E., and Black, W. C. (2018). *Multivariate Data Analysis*, 8th Edn. Boston, MA: Cengage.

Harb, C., and Smith, P. B. (2008). Self-construals across cultures. *J. Cross. Cult. Psychol.* 39, 178–197. doi: 10.1177/0022022107313861

Heine, S. J., Lehman, D. R., Peng, K., and Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *J. Pers. Soc. Psychol.* 82, 903–918. doi: 10.1037/0022-3514.82.6.903

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain. Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X

Hnilica, K., Rendlová, M., Bariekzahyová, T., and Hnilica, M. (2006). Životní standard, individualistickéhodnoty a spokojenost se životem [Life standard, individualistic values, and life satisfaction]. *Cesk. Psychol.* 50, 201–217.

Hofstede, G., Hofstede, G. J., and Minkov, M. (2010). *Cultures and Organizations: Software of the Mind*, Third Edn. New York, NY: McGraf-Hill.

Hofstede, G. (1983). The cultural relativity of organizational practices and theories. *J. Int. Bus. Stud.* 14, 75–89. doi: 10.1057/palgrave.jibs.8490867

Hofstede, G. (1994). *Value Survey Module 1994 Manual*. Maastricht: Institute for Research on Intercultural Cooperation.

Hooper, D., Coughlan, J., and Mullen, M. R. (2008). Structural equation modelling: guidelines for determining model fit. *Electron. J. Bus. Res. Methods* 6, 53–60. doi: 10.21427/D7CF7R

Hoyle, R. H. (2012). *Handbook of Structural Equation Modeling*. New York, NY: The Guilford Press.

Hu, L., 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

Hsu, S. Y., and Barker, G. G. (2013). Individualism and collectivism in Chinese and American television advertising. *Int. Commun. Gaz.* 75, 695–714. doi: 10.1177/1748048513482543

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., and Rosseel, Y. (2018). *semTools: Useful Tools for Structural Equation Modeling*. R Package Version 0.5-3. Available online at: https://CRAN.R-project.org/package=semTools

Kashima, E., and Hardie, E. (2000). The development and validation of the relational, individual, and collective self-aspects (RIC) scale. *Asian J. Soc. Psychol.* 3, 19–48. doi: 10.1111/1467-839X.00053

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., et al. (2018). Many labs 2: investigating variation in replicability across samples and settings. *Adv. Methods Pract. Psychol. Sci.* 1, 443–490. doi: 10.1177/2515245918810225

Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling*, 4th Edn. New York, NY: The Guilford Press.

Kolman, L., Noorderhaven, N., Hofstede, G., and Dienes, E. (2003). Cross—cultural differences in Central Europe. *J. Manag. Psychol.* 18, 76–88. doi: 10.1108/02683940310459600

Korkmaz, S., Goksuluk, D., and Zararsiz, G. (2014). MVN: an R package for assessing multivariate normality. *R J.* 6, 151–162. doi: 10.32614/RJ-2014-031

Lacko, D., and Čeněk, J. (2020). "Factor structure of two individualism and collectivism scales," in *PhD Existence 10; Česko-slovenskápsychologickákonference (nejen) Pro Doktorandy a o Doktorandech*, eds E. Maierová, L. Viktorová, M. Dolejš, and T. Dominik (Olomouc: Palacký University), 18–24.

Lacko, D., Šašinka, Č, Čeněk, J., Stachoň, Z., and Lu, W. (2020). Cross-cultural differences in cognitive style, individualism/collectivism and map reading between Central European and East Asian University students. *Stud. Psychol. (Bratisl)* 62, 23–43. doi: 10.31577/sp.2020.01.789

Levine, T. R., Bresnahan, M. J., Park, H. S., Lapinski, M. K., Wittenbaum, G., Shearman, S., et al. (2003b). Self-construal scales lack validity. *Hum. Commun. Res.* 29, 210–252. doi: 10.1111/j.1468-2958.2003.tb00837.x

Levine, T. R., Bresnahan, M. J., Park, H. S., Lapinski, M. K., Lee, T. S., and Lee, D. W. (2003a). The (In)validity of self-construal scales revisited. *Hum. Commun. Res.* 29, 291–308. doi: 10.1111/j.1468-2958.2003.tb00840.x

Li, C. (2016). Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* 48, 936–949. doi: 10.3758/s13428-015-0619-7

Li, F., and Aksoy, L. (2006). Dimensionality of individualism–collectivism and measurement equivalence of Triandis and Gelfand's scale. *J. Bus. Psychol.* 21, 313–329. doi: 10.1007/s10869-006-9031-8

Little, T. D. (2013). *Longitudinal Structural Equation Modeling*. New York, NY: The Guilford Press.

Lu, L., and Gilmour, R. (2007). Developing a new measure of independent and interdependent views of the self. *J. Res. Pers.* 41, 249–257. doi: 10.1016/j.jrp.2006.09.005

Luhtanen, R., and Crocker, J. (1992). A collective self-esteem scale: self-evaluation of one's social identity. *Pers. Soc. Psychol. Bull.* 18, 302–318. doi: 10.1177/0146167292183006

Markus, H., and Kitayama, S. (1991). Culture and the self: implications for cognition, emotion, and motivation. *Psychol. Rev.* 98, 224–253. doi: 10.1037/0033-295X.98.2

Martinkova, P., and Drabinova, A. (2018). ShinyItemAnalysis for teaching psychometrics and to enforce routine analysis of educational tests. *R J.* 10, 503–515. doi: 10.32614/RJ-2018-074

Matsumoto, D., Weissman, M., Preston, K., Brown, B., and Kupperbush, C. (1997). Context-specific measurement of individualism-collectivism on the individual level: the individualism-collectivism interpersonal assessment inventory. *J. Cross. Cult. Psychol.* 28, 743–767. doi: 10.1177/0022022197286006

Matsumoto, D. (1999). Culture and self: an empirical assessment of Markus and Kitayama's theory of independent and interdependent self-construals. *Asian J. Soc. Psychol.* 2, 289–310. doi: 10.1111/1467-839x.00042

Matsumoto, D., and Juang, L. (2013). *Culture & Psychology*. Belmont: Wadsworth Publishing.

Morling, B., and Lamoreaux, M. (2008). Measuring culture outside the head: a meta-analysis of individualism-collectivism in cultural products. *Pers. Soc. Psychol. Rev.* 12, 199–221. doi: 10.1177/1088868308318260

McSweeney, B. (2002). Hofstede's model of national cultural differences and their consequences: A triumph of faith -a failure of analysis. *Hum. Relat.* 55, 89–118. doi: 10.1177/0018726702551004

Miramontes, L. G. (2011). *The Structure and Measurement of Self-Construals: A Cross-Cultural Study of the Self-Construal Scale*. dissertation thesis. Washington, DC: Washington State University.

Noguchi, K. (2007). Examination of the content of individualism/collectivism scales in the cultural comparisons of the USA and Japan. *Asian J. Soc. Psychol.* 10, 131–144. doi: 10.1111/j.1467-839X.2007.00220.x

Oyserman, D., and Lee, S. (2008). Does culture influence what and how we think? Effects of priming individualism and collectivism. *Psychol. Bull.* 134, 311–342. doi: 10.1037/0033-2909.134.2.311

Oyserman, D., Coon, H., and Kemmelmeier, M. (2002). Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychol. Bull.* 128, 3–72. doi: 10.1037/0033-2909.128.1.3

Partikova, V. (2019). Exploring the self-perception of kung fu teachers. An interpretative phenomenological analysis. *Eur. J. Sport Soc.* 16, 247–267. doi: 10.1080/16138171.2019.1661143

Pilarska, A. (2011). PolskaadaptacjaSkaliKonstruktówJa [Polish adaptation of Self-Construal Scale]. *Studia Psychol.* 49, 21–34. doi: 10.2478/v10167-011-0002-y

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ramley, F., Karnilowicz, W., Yasin, M. A. S. M., Balqis, S., and Nor, M. (2020). An analysis of cross-cultural equivalence of self-construal scale in Malaysia. *Int. J. Adv. Appl. Sci.* 7, 91–98. doi: 10.21833/ijaas.2020.02.013

Revelle, W. (2020). *psych: Procedures for Psychological, Psychometric, and Personality Research*. R Package Version 2.1.3. Evanston, IL: Northwestern University.

Řeháková, B. (2006). Měřeníhodnotovýchorientacímetodouhodnotovýchportrétů S. H. Schwartze [Measuring Value Orientations with the Use of S. H. Schwartz's Value Portraits]. *Czech. Sociol. Rev.* 42, 107–128. doi: 10.13060/00380288.2006.42.1.07

Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. doi: 10.18637/jss.v048.i02

Santos, H. C., Varnum, M. E. W., and Grossmann, I. (2017). Global increases in individualism. *Psychol. Sci.* 28, 1228–1239. doi: 10.1177/0956797617700622

Schimmack, U., Oishi, S., and Diener, E. (2005). Individualism: a valid and important dimension of cultural differences between nations. *Pers. Soc. Psychol. Rev.* 9, 17–31. doi: 10.1207/s15327957pspr0901_2

Schwartz, S. (1992). Universals in the content and structure of values: theoretical advances and empirical tests in 20 Countries. *Adv. Exp. Soc. Psychol.* 22, 1–65. doi: 10.1016/S0065-2601(08)60281-6

Schwartz, S., Melech, G., Lehrnami, A., Burgess, S., Harris, M., and Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *J. Cross. Cult. Psychol.* 32, 519–542. doi: 10.1177/0022022101032005001

Shulruf, B., Hattie, J., and Dixon, R. (2007). Development of a new measurement tool for individualism and collectivism. *J. Psychoeduc. Assess.* 25, 385–401. doi: 10.1177/0734282906298992

Singelis, T., Triandis, H., Bhawuk, D., and Gelfand, M. (1995). Horizontal and vertical dimensions of individualism and collectivism: a theoretical and measurement refinement. *Cross Cult. Res.* 29, 240–275. doi: 10.1177/106939719502900302

Sivadas, E., Bruvold, N. T., and Nelson, M. R. (2008). A reduced version of the horizontal and vertical individualism and collectivism scale: a four-country assessment. *J. Bus. Res.* 61, 201–210. doi: 10.1016/j.jbusres.2007.06.016

Smith, P., Fischer, R., Vignoles, V., and Bond, M. (2013). *Understanding Social Psychology Across Cultures: Engaging With Others in a Changing World*. London: Sage.

Smith, P., Vignoles, V., Becker, M., Owe, E., Easterbrook, M., Brown, R., et al. (2016). Individual and culture-level components of survey response styles: a multi-level analysis using cultural models of selfhood. *Int. J. Psychol.* 51, 453–463. doi: 10.1002/ijop.12293

Soh, S., and Leong, F. T. L. (2002). Validity of vertical and horizontal individualism and collectivism in Singapore. *J. Cross. Cult. Psychol.* 33, 3–15. doi: 10.1177/0022022102033001001

Talhelm, T., Zhang, X., and Oishi, S. (2018). Moving chairs in starbucks: observational studies find rice-wheat cultural differences in daily life in China. *Sci. Adv.* 4:eaa8469. doi: 10.1126/sciadv.aap8469

Takano, Y., and Osaka, E. (1999). An Unsupported common view: comparing Japan and the US on individualism/collectivism. *Asian J. Soc. Psychol.* 2, 311–341. doi: 10.1111/1467-839X.00043

Takano, Y., and Osaka, E. (2018). Comparing Japan and the United States on individualism/ collectivism: a follow-up review. *Asian J. Soc. Psychol.* 21, 301–316. doi: 10.1111/ajsp.12322

Triandis, H., and Gelfand, M. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *J. Pers. Soc. Psychol.* 74, 118–128. doi: 10.1037/0022-3514.74.1.118

Uskul, A. K., and Oyserman, D. (2006). "Question comprehension and response: implications of individualism and collectivism," in *National Culture and Groups*, Vol. 9, ed. Y.-R. Chen (Bingley: Emerald Publishing), 173–201. doi: 10.1016/S1534-0856(06)09008-6

Utsey, S., and Constantine, M. (2006). A confirmatory test of the underlying factor structure of scores on the collective self-esteem scale in two independent samples of black Americans. *J. Pers. Assess.* 86, 172–179. doi: 10.1207/s15327752jpa8602_06

van de Vijver, F., and Hambleton, R. (1996). Translating tests: some practical guidelines. *Eur. Psychol.* 1, 89–99. doi: 10.1027/1016-9040.1.2.89

van de Vijver, F., and Leung, K. (2001). Personality in cultural context: methodological issues. *J. Pers.* 69, 1007–1031. doi: 10.1111/1467-6494.696173

van de Vijver, F., and Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: an overview. *Eur. Rev. of App. Psychol.* 47, 263–280. doi: 10.1016/j.erap.2003.12.004

Varnum, M., Grossmann, I., Kitayama, S., and Nisbett, R. (2010). The origin of cultural differences in cognition: evidence for the social orientation hypothesis. *Curr. Dir. Psychol. Sci.* 19, 9–13. doi: 10.1177/0963721409359301

Vignoles, V., Owe, E., Becker, M., Smith, P., Easterbrook, M., Brown, R., et al. (2016). Beyond the 'East-West' dichotomy: global variation of cultural models of selfhood. *J. Exp. Psychol. Gen.* 145, 966–1000. doi: 10.1037/xge0000175

Voronov, M., and Singer, J. (2002). The myth of individualism-collectivism: a critical review. *J. Soc. Psychol.* 142, 461–480. doi: 10.1080/00224540209603912

Wagner, J., and Moch, K. (1986). Individualism-collectivism: concept and measure. *Group Organ. Stud.* 11, 280–304. doi: 10.1177/105960118601100309

Welkenhuysen-Gybels, J., Billiet, J., and Cambré, B. (2003). Adjustment for acquiescence in the assessment of the construct equivalence of Likert type score items. *J. Cross. Cult. Psychol.* 34, 702–722. doi: 10.1177/0022022103257070

Yamagishi, T. (1988). The provision of a sanctioning system in the United States and Japan. *Soc. Psychol. Q.* 51, 265–271. doi: 10.2307/2786924

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 in Psychology

ORIGINAL RESEARCH
published: 09 June 2021
doi: 10.3389/fpsyg.2021.681058

Check for updates

Development of a Clinic Screening Tool to Identify Burdensome Health-Related Issues Affecting People Living With HIV in Spain

Maria José Fuster-RuizdeApodaca[1,2], Kelly Safreed-Harmon[3], Marta Pastor de la Cal[1,4], Ana Laguia[2], Denise Naniche[3] and Jeffrey V. Lazarus[3]*

[1] Sociedad Española Interdisciplinaria del Sida (SEISIDA), Madrid, Spain, [2] Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain, [3] Barcelona Institute for Global Health (ISGlobal), Hospital Clínic, University of Barcelona, Barcelona, Spain, [4] Bizkaisida, Bilbao, Spain

OPEN ACCESS

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

Reviewed by:
Javier Fenollar Cortés,
Loyola University Andalusia, Spain
Luis Botella,
Ramon Llull University, Spain

***Correspondence:**
Kelly Safreed-Harmon
kelly.safreed-harmon@isglobal.org

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

Received: 15 March 2021
Accepted: 03 May 2021
Published: 09 June 2021

Citation:
Fuster-RuizdeApodaca MJ,
Safreed-Harmon K, Pastor de la
Cal M, Laguia A, Naniche D and
Lazarus JV (2021) Development of a
Clinic Screening Tool to Identify
Burdensome Health-Related Issues
Affecting People Living With HIV
in Spain. Front. Psychol. 12:681058.
doi: 10.3389/fpsyg.2021.681058

Background: Numerous health-related issues continue to undermine the health and health-related quality of life (HRQoL) of people living with HIV (PLHIV). We developed a clinic screening tool (CST-HIV) for the purpose of identifying these issues in routine specialist clinical care in Spain.

Methods: We used the following established instrument development procedures: (1) a literature review; (2) four focus group discussions (FGDs), two that convened 16 expert HIV care providers, and two that convened 15 PLHIV; (3) prioritisation, selection and definition of constructs (health-related issues) to include in the CST-HIV and drafting of initial item pool; and (4) a pilot study to analyse psychometric properties and validity of items and to determine which to retain in the final CST-HIV. The FGD interview scripts incorporated an exercise to prioritise the health-related issues perceived to have the greatest negative effect on HRQoL. The online questionnaire used for the pilot study included the pool of CST-HIV items and validated measures of each construct.

Results: We identified 68 articles that reported on factors associated with the HRQoL of PLHIV. The most burdensome health-related issues identified in the FGDs related to stigma, socioeconomic vulnerability, sleep/fatigue, pain, body changes, emotional distress, and sexuality. Based on the literature review and FGD findings, we selected and defined the following constructs to include in the initial CST-HIV: anticipated stigma, emotional distress, sexuality, social support, material deprivation, sleep/fatigue, cognitive problems, and physical symptoms. Two researchers wrote six to eight items for each construct. Next, 18 experts rated 47 items based on their clarity, relevance, and representativeness. Pilot testing was carried out with 226 PLHIV in Spain. We retained 24 items based on empirical criteria that showed adequate psychometric properties. Confirmatory factor analysis confirmed the eight-factor structure with a good fit to the data (RMSEA = 0.035, AGFI = 0.97, CFI = 0.99). We found strong positive correlations between the instrument's eight dimensions and validated measures of the same constructs. Likewise, we found negative associations between the dimensions of the CST-HIV and HRQoL.

Conclusion: The CST-HIV is a promising tool for use in routine clinical care to efficiently identify and address health-related issues undermining the HRQoL of PLHIV.

Keywords: HIV, patient-reported outcome measure (PROM), health-related quality of life, symptom assessment, health measurement instrument, psychometrics, Spain

INTRODUCTION

Widespread access to antiretroviral therapy (ART) has enabled many people living with HIV (PLHIV) to control their infection on a long-term basis. The life expectancy of PLHIV now approaches that of the general population in resource-rich settings and has greatly increased in resource-poor settings as well (Antiretroviral Therapy Cohort Collaboration, 2017; Teeraananchai et al., 2017). However, numerous issues undermine the well-being of PLHIV, including PLHIV who are stable on ART.

Multimorbidity is more prevalent among PLHIV than members of the general population, with commonly occurring comorbidities including mental health disorders and ageing-related non-communicable diseases such as cardiovascular, liver and kidney disease (Chuah et al., 2017; Maciel et al., 2018; Smit et al., 2015). PLHIV have a high burden of symptoms of ill health such as pain, fatigue and gastrointestinal problems (Harding et al., 2010; Wilson et al., 2016; Ibarra-Barrueta et al., 2019). They furthermore face an array of challenges to their psychosocial and material well-being (Bristowe et al., 2019; Public Health England, 2020). HIV-related stigma and discrimination have far-reaching ramifications in terms of mental health, medication adherence, health-seeking behaviour, social relationships, employment and other areas of people's lives (Sweeney and Vanable, 2016; Wagener et al., 2017; Ikeda et al., 2019). PLHIV also must grapple with the emotional and practical demands of living with a complex chronic health condition that requires lifelong ongoing treatment.

In this context, it is notable that a large study in the United Kingdom found poorer health-related quality of life (HRQoL) outcomes among PLHIV than among the general population (Miners et al., 2014). This difference persisted even for the subgroup of PLHIV who were virally suppressed. Other research has found poor HRQoL outcomes in PLHIV populations to be associated with a wide range of factors, including pain, insomnia, mental health disorders and HIV-related stigma (Degroote et al., 2014; Sabin et al., 2018; Andersson et al., 2020; Kunisaki et al., 2021).

People living with HIV who have responded well to ART typically are advised to see their healthcare providers for clinical monitoring two to four times per year. These routine clinic visits present an important window of opportunity for healthcare providers to identify and address some of the problems that can contribute to poor HRQoL. However, PLHIV often encounter communication barriers with their healthcare providers and may not feel that providers are responsive to their healthcare priorities (Antunes et al., 2020; Fredericksen et al., 2020a; Okoli et al., 2020). Furthermore, providers may overlook important symptoms (Edelman et al., 2011).

In recent years, the World Health Organization (WHO) and many health systems increasingly have promoted person-centred care, which WHO describes as being "organised around the comprehensive needs of people rather than individual diseases" (McCormack et al., 2015; World Health Organization (WHO), 2016). One means of promoting good communication about people's healthcare needs is to ask patients to complete surveys known as patient-reported outcome measures (PROMs) (Wheat et al., 2018; Fredericksen et al., 2020b). There are validated PROMs focusing on numerous aspects of health and well-being, including generic PROMs designed for all patient populations as well as PROMS that reflect the concerns of patients with specific diseases and conditions including HIV.

A 2017 review of HIV-specific PROMs identified 117 validated instruments for measuring patients' perceptions of their health and related issues in areas such as medication adherence, symptoms, psychological challenges, HIV-related stigma, social support, and sexual and reproductive health (Engler et al., 2017). Because these instruments typically focus on narrow topics, it would be necessary to use multiple instruments to learn about different aspects of a patient's well-being. The time-intensive nature of such an approach points to a need for broadly focused PROMs that are short enough to be easily integrated into routine clinic visits, enabling healthcare providers to quickly determine which of many potential health-related problems should be addressed in these visits. Despite the contribution that this type of PROM could make to routine clinical care, this remains an area under development. The only such instrument that we are aware of in the HIV field is currently being developed by Bristowe et al. (2019, 2020), with the content of the instrument guided by qualitative research involving PLHIV and other key stakeholders in England and Ireland.

The present study is part of a broader research project to improve the HRQoL and the long-term health of PLHIV in Spain and Italy. It constitutes the first stage of the research, and its aim is to develop a brief Spanish clinic screening tool (CST-HIV) that can be used in routine clinical care to identify problems that undermine the HRQoL of PLHIV. This paper reports the process of developing the instrument to ensure its content and face validity, describes the psychometric properties of the instrument, and presents the evidence of construct and criterion validity that we obtained when we piloted the instrument.

MATERIALS AND METHODS

Study Design

This study comprised several steps, including a literature review, a qualitative study, an item design process, a cognitive debriefing study, and a pilot cross-sectional *ex post-facto* study to analyse the psychometric properties of the initial version of the CST-HIV. **Table 1** summarises the research design, procedures and participants involved. All of these steps will be detailed in the following sections.

Participants

A total of 31 persons participated in the qualitative study to identify the initial dimensions of the CST-HIV. Sixteen of them were expert service providers from diverse disciplines (physicians, nurses, psychiatrists, psychologists, and staff of non-governmental organisations [NGOs]). The remaining 15 were PLHIV. Six of the experts and also six of the PLHIV were cis-women. Among PLHIV, one transgender woman also participated. The other participants were cis-men.

A total of 18 multidisciplinary experts from diverse disciplines and areas of expertise, three of whom were PLHIV, participated in the expert assessment and inter-rater process to develop the initial pool of items.

Eight PLHIV, five men and three women, participated in the cognitive debriefing of the CST-HIV items. Next, we conducted the pilot study investigating the item pool's psychometric properties in a sample of 226 PLHIV from different regions of Spain. The sample size was determined in accordance with the sample size requirements for carrying out confirmatory factor analysis (Bentler and Chou, 1987). Since these requirements call for 10 participants per item, and we anticipated that the final number of CST-HIV items would be between 21 and 24, our target sample size was between 210 and 240 PLHIV. The inclusion criteria were having an HIV-positive diagnosis, being at least 18 years old, and not having any severe psychiatric or cognitive disorders. Excluding people with such disorders is standard in studies in which participants complete self-administered surveys since the presence of such disorders could affect one's cognitive capacity to understand questions and provide reliable responses.

TABLE 1 | Summary of research design.

Steps		Procedures	Participants involved
Step 1	Literature review	Identification of initial domains	Authors
Step 2	Qualitative study with focus groups	Identification of initial domains	N = 15 PLHIV N = 16 experts N = 31 Total
Step 3	Development of initial pool of items	Definition of constructs and drafting of items	N = 3
		Expert assessment and inter-rater process	N = 18
		Cognitive debriefing	N = 8
Step 4	Pilot study	Assessment of psychometric properties and validity of items	N = 226

PLHIV, people living with HIV.

Table 2 shows the sociodemographic and clinical characteristics of pilot study participants. Most of them were male and homosexual, and the most commonly reported mode of HIV infection was sexual intercourse. The mean age was 44. Approximately one-third of the participants had a university degree, and 39% were employed. Sixty-eight percent reported having a personal monthly income of €900 or less. The immunological and virological HIV status of most participants were good.

Procedure

This research took place from April 2019 to October 2020. The Ethics Committee of the Hospital Clínic of Barcelona, Spain, approved all research procedures. Participants in all phases signed informed consent forms before data collection began.

The HIV Clinic Screening Tool (CST-HIV) was developed through the following well-established methodological steps (Eignor, 2001; Revicki et al., 2007).

Firstly, we conducted an exploratory literature review to obtain information about issues that undermine the well-being of PLHIV and to identify themes that would warrant further exploration in focus group discussions (FGDs). We identified English-language peer-reviewed articles and conference abstracts indexed in PubMed using search strings that addressed two major lines of research: HIV symptom burden and predictors of HRQoL in PLHIV. We used appropriate selection criteria to identify the studies of greatest relevance to our study (e.g., studies reporting on adult PLHIV who are taking ART and studies reporting on the symptom burden in PLHIV from 2010 onward, in recognition that the symptom profile has changed in accordance with ART improvements). We used Scopus and ResearchGate to identify articles that cited a key source about the widely used HIV Symptom Index (Justice et al., 2001). Selected references were compiled in tables to identify evidence regarding burdensome symptoms and predictors of HRQoL in PLHIV.

Drawing on literature review findings, we conducted a qualitative study using the FGD methodology to obtain the perspectives of PLHIV and other key informants regarding the most burdensome health-related problems facing PLHIV. We carried out four FGDs. Two of them enrolled HIV service providers (n = 8 per FGD), and the other two enrolled PLHIV (n = 8 and n = 7). Participants in the service provider FGDs were selected via purposive sampling to ensure the representation of different types of providers such as physicians, nurses, psychiatrists, psychologists, and NGO staff. Service providers worked in Barcelona, Bilbao, Madrid, Seville, and Valencia. Participants in the PLHIV FGDs were selected via purposive sampling to ensure diverse epidemiological profiles in terms of age, sex, sexual orientation, and drug use history. One PLHIV FGD was comprised of clients of an NGO providing HIV services in Barcelona, and the other PLHIV FGD was comprised of patients at the HIV outpatient clinic of a large Barcelona university hospital. FGDs took place in April and May 2019, with each one lasting approximately two hours. Facilitators used

TABLE 2 | Characteristics of pilot study participants (*N* = 226).

Sociodemographic and clinical variables	% (*n*)
Age, mean ($M \pm SD$)	43.81 ± 11.15
Gender	
Male	75.7 (171)
Female	21.7 (49)
Transgender	1.3 (3)
Other	1.3 (3)
Sexual orientation	
Heterosexual	41.2 (93)
Homosexual	54.9 (124)
Bisexual	2.2 (5)
Other	1.3 (3)
No answer	0.4 (1)
Education level	
No education	2.2 (5)
Elementary school	19.9 (45)
High school	44.2 (100)
University degree	32.3 (73)
Other	1.3 (3)
Work situation	
Working	38.9 (88)
Unemployed	28.3 (64)
Retired/on disability	21.2 (48)
Other	11.5 (26)
Personal monthly income	
None	12.8 (29)
≤ 300 €	13.7 (31)
301–600 €	17.3 (39)
601–900 €	24.3 (55)
901–1200 €	13.7 (31)
1201–1800 €	11.9 (27)
1801–2400 €	3.1 (7)
2401–3000 €	0.4 (1)
3001–4500 €	0.9 (2)
No answer	1.8 (4)
Housing	
Own home (rent or own)	56.6 (128)
Family home	12.8 (29)
Shared home	16.8 (38)
Someone else's home	1.3 (3)
Shelter/institution	6.6 (15)
Other	5.8 (13)
HIV transmission route	
Sexual intercourse	78.3 (177)
Sharing injection materials	10.6 (24)
Unknown	8.8 (20)
Other	2.2 (5)
CD4 cell count, cells/mm^3	
≤ 200	7.5 (17)
201–400	7.1 (16)
> 400	53.5 (121)
Unknown	31.9 (72)
Duration of infection, years, mean ($M \pm SD$)	14.18 ± 10.47
Undetectable plasma viral load	92.5 (209)

Data in percentages unless otherwise stated.

semi-structured scripts with open-ended questions and prompts to guide the discussions.

The next step in the development of the CST-HIV consisted of developing a pool of potential items. Based on findings from the FGDs and the literature reviews, three members of the research team selected the most prevalent and burdensome health-related problems undermining the HRQoL of PLHIV. Also, they defined the constructs (the health-related problems) after deliberation (Nunnally and Bernstein, 1994). Items were developed to measure each construct, following psychometric recommendations (Osterlind, 1989; Haladyna et al., 2002), and the response format for the items was decided. A team of 18 multidisciplinary experts rated the items based on their clarity, relevance and representativeness. Based on the experts' ratings and comments, items were selected and reworded as appropriate to create the initial item pool. A cognitive debriefing study was then carried out, in which eight PLHIV rated the understandability of the items. These participants were members of the NGO collaborators in the research.

Finally, we conducted a pilot study to assess the initial items' psychometric properties and to select those that would be part of the final CST-HIV. We recruited participants through NGO collaborators, and we asked those who agreed to participate to complete an online questionnaire using Qualtrics[1], a private online survey development platform.

Measures

For the qualitative study, we designed a semi-structured FGD script addressing two central questions: (1) "In your opinion, what are the health-related problems that have the most significant negative effect on the quality of life of PLHIV?"; and (2) "Among the problems that you have identified, what do you think are the most important ones to include in a short diagnostic questionnaire?" All FGD participants were also asked to carry out a prioritisation exercise in which they selected what they believed to be the most burdensome issues from among all issues identified during the discussions.

The online questionnaire used for the pilot study included the 40 items selected after the inter-rater process. We selected brief instruments to measure preliminary evidence of the convergent validity of each CST-HIV dimension. We chose instruments according to their psychometric properties, validity, and availability of cut-off points. When a Spanish version of an instrument was available, we used it. When it was not, we conducted a backward translation of the instrument. The questionnaire included the following instruments:

Anticipated Stigma

The factors of disclosure concerns and public stigma of the Spanish Stigma Scale measured through 13 items were used (Fuster-RuizdeApodaca et al., 2015). Results of the Spanish adaptation of the instrument indicated that these two factors could be grouped in a latent second-order dimension related to internalised stigma (Fuster-RuizdeApodaca et al., 2015). The scale is rated on a four-point response format (*1 = strongly*

[1] www.qualtrics.com

disagree, 4 = strongly agree), with higher scores indicating greater concerns.

Emotional Distress

We used the Patient Health Questionnaire-4 (PHQ-4) (Kroenke et al., 2009) and the Spanish version of the Hospital Anxiety and Depression Scale (HADS) (Tejero et al., 1986). The PHQ-4 is a validated ultra-brief screening tool that has a two-factor structure, one containing two anxiety items (GAD-2) and the other containing two depression items (PHQ-2). Responses are scored from 0 (*not at all*) to 3 (*nearly every day*). The total score on this measure ranges from 0 to 12. The HADS is a 14-item, self-reporting screening scale that contains two seven-item Likert scales, one for anxiety and one for depression. Each item is answered by the patient on a four-point (0–3) response category, and thus the possible scores range from 0 to 21 for anxiety and 0 to 21 for depression.

Sexuality

We used the PROMIS V2.0 Satisfaction with Sex Life scale (Weinfurt et al., 2015), which is part of the modular and customisable PROMIS Sexual Function and Satisfaction 2.0 measures that assess multiple components of sexual functioning. The Satisfaction with Sex Life module assesses how satisfying and pleasurable the person regards his or her sexual activities, with no constraints on how the person defines "sex life". Items are gender-non-specific. Higher scores indicate more satisfying sexual experiences.

Social Support

The Duke-UNC Functional Social Support Questionnaire was selected (Broadhead et al., 1988). It is an 11-item scale measuring two dimensions of social support: confidant support and affective support. Items have a five-point Likert format response. Higher scores indicate higher social support.

Material Deprivation

We used the Social Exclusion Index for Health Surveys (SEI-HS) (Van Bergen et al., 2017). This instrument contains 17 items that measure four dimensions: (1) social participation; (2) normative integration; (3) material deprivation; and (4) access to basic social rights.

Sleep Problems

We used the Spanish version of the Insomnia Severity Index (ISI) (Bastien et al., 2001; Fernandez-Mendoza et al., 2012). This seven-item index is a reliable measure for evaluating perceived sleep difficulties. Each item is rated on a 5-point Likert scale (*0 = no problem, 4 = very severe problem*), yielding a total score ranging from 0 to 28.

Fatigue

We used the seven-item version of the Fatigue Severity Scale (FSS), which has demonstrated good psychometric properties in PLHIV (Lerdal et al., 2011). Each item is rated on a seven-point Likert-scale (*1 = strongly disagree, 7 = strongly agree*). The mean score is used to estimate fatigue severity.

Cognitive Problems

The Neuro-QoL V2.0 Cognitive Function measure was used for cognitive assessment (Lai et al., 2014). This eight-item scale measures both cognitive function concerns and abilities.

HRQoL

We used the HIV-specific WHOQoL-HIV-BREF measure that has been validated in Spanish (Fuster-RuizdeApodaca et al., 2019). The instrument has 31 items covering six domains: physical health; psychological health; level of independence; environmental health; social relationships; and spirituality, religion and personal beliefs (SRPB). It additionally has a general health dimension assessing one's overall perception of one's health and HRQoL. All items use a five-point scale. Negative items are reverse-coded for scoring. Thus, higher scores for all items indicate better HRQoL.

We also used the generic HRQoL measure EQ-5D-5L, which has five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Responses are provided on a five-point scale ranging from "no problems" to "extreme problems" (Herdman et al., 2011).

The online questionnaire also included a section that requested health and sociodemographic data.

Data Analysis

To analyse the qualitative data, we performed directed content analysis (Mayring, 2000) using MAXQDA 12 software. The FGDs were transcribed, reviewed for accuracy, and coded. Inductive and deductive coding were used to identify relevant concepts, and an analysis of these concepts led to the identification of key categories and subcategories of health-related problems. We also performed a quantitative analysis of the qualitative data to determine the number of times each code and category was used. Two analysts discussed and agreed on the data categorisation, with inconsistencies resolved by consensus. Following the coding of the FGD content, all research team members reviewed and approved the final categorisation of data.

To analyse the content validity of the initial pool of items evaluated in the inter-rater process, we calculated the Osterlind Index (Osterlind, 1989) for the items' representativeness and relevance scores. Representativeness and relevance items had a three-point ordinal response (high, medium, low). There is no clear criterion regarding a cut-off point for this index; some use 0.5 and others 0.75 depending on the objective. We used a strict criterion in most dimensions, selecting items with an Osterlind Index of up to 0.75.

In the pilot study, we assessed the psychometric properties of the initial CST-HIV item pool based on empirical criteria. We assessed the floor and ceiling effects, the internal consistency, the reliability, and the validity index of each dimension (Kline, 2013). Most items in the online questionnaire in the Qualtrics survey platform were programmed for compulsory completion. Thus, there were no missing values in the variables collected.

Next, to test the construct validity, first-order confirmatory factor analysis (CFA) was used to assess the retained CST-HIV items' fit with the theoretical proposed structure. Due to the ordinal nature of our data and the sample size, we chose the robust unweighted least-squares extraction method (ULS) (Batista-Foguet and Coenders, 2000; Holgado-Tello et al., 2009; Holgado–Tello et al., 2010). Although the weighted least squares method also could be used, we did not use it because of the instability of its inverse matrix when the models have more than ten variables or a moderate sample size (Holgado-Tello et al., 2018; Holgado-Tello et al., 2009; Satorra, 1990). The goodness of fit was evaluated using several absolute and relative fit indices, including the goodness of fit index (GFI), the adjusted goodness of fit index (AGFI), the comparative fit index (CFI), the standardised root mean square residual (SRMR) and the standardised root mean square error of approximation (RMSEA). A model is considered to have a good fit when the goodness of fit indices (GFI and AGFI) and CFI are greater than 0.90, RMSEA is lower than 0.08, and SRMR is lower than 0.08 (Hu and Bentler, 1995).

We then calculated reliability and construct statistics of the CST-HIV including the Cronbach's alpha coefficient to assess internal consistency, the average extracted variance (AVE) to assess convergent validity, and the Jöreskog rho (Omega) to assess construct reliability (Fornell and Larcker, 2016). Cronbach's alpha coefficients between 0.70 and 0.90 are adequate, and between 0.60 and 0.70 are acceptable (Kline, 2013). AVE values greater than 0.50 indicate convergent validity, and Omega coefficients between 0.70 and 0.90 are considered to represent acceptable construct reliability (Campo-Arias and Oviedo, 2008), although in some circumstances, values higher than 0.65 can be accepted (Katz, 2006).

Convergent and concurrent validity were analysed by calculating the Pearson correlation between each CST-HIV dimension and the validated instruments used to measure the constructs and HRQoL. We expected each dimension to correlate positively with its convergent criterion measure and negatively with HRQoL.

Regarding the data analysis software, LISREL (LInear Structural RELations) 8.7 and its companion preprocessor programme PRELIS for Windows were used for the CFAs (Jöreskog and Sörbom, 1996). IBM SPSS Statistics 22 (IBM Corp, 2013) was used for the remaining analyses.

RESULTS

Step One – Identification of Dimensions to Include in the CST-HIV: Literature Review

The literature review on the HIV symptom burden identified five articles and two conference abstracts that were relevant to the current study. The symptoms that were most commonly reported to be highly prevalent in PLHIV were sleep-related problems, fatigue, and muscle/joint pain (Erdbeer et al., 2014; McGowan et al., 2014; Wilson et al., 2016; Schnall et al., 2018; Cioe et al., 2019; Ibarra-Barrueta et al., 2019; Schnall et al., 2019). Other highly prevalent symptoms observed in some studies included anxiety, depression, sexual dysfunction, changes in body appearance, and gastrointestinal problems (Erdbeer et al., 2014; Wilson et al., 2016; Schnall et al., 2018; Ibarra-Barrueta et al., 2019).

The HRQoL literature review identified a large body of relevant research on factors associated with HRQoL outcomes in PLHIV, including a 2014 review article (Degroote et al., 2014). We analysed the findings of the review article and 68 additional articles that reported on more recent studies. We observed that one of the factors most commonly reported to be associated with positive HRQoL outcomes in PLHIV is social support (Bekele et al., 2013; Emlet et al., 2013; Slater et al., 2013; Dalmida et al., 2015; George et al., 2016; Nideröst and Imhof, 2016; den Daas et al., 2019). Two factors associated with negative HRQoL outcomes in many studies are depression and material insecurity (e.g., unemployment, financial problems, unmet needs for food and housing) (Douab et al., 2014; Dalmida et al., 2015; Ballester-Arnal et al., 2016; George et al., 2016; Nideröst and Imhof, 2016; Catalan et al., 2017; Logie et al., 2018; Sok et al., 2018; Olson et al., 2019). Other factors associated with negative HRQoL outcomes in some studies included comorbidity, stigma and HIV disclosure concerns (Emlet et al., 2013; Slater et al., 2013; Fekete et al., 2016; George et al., 2016; Nideröst and Imhof, 2016; Logie et al., 2018; Reinius et al., 2018). A high symptom burden was also associated with negative HRQoL outcomes, as were specific symptoms such as body disfigurement, memory difficulties and sexual functioning (Ballester-Arnal et al., 2016; George et al., 2016; Brandt et al., 2017; den Daas et al., 2019; Olson et al., 2019).

Step Two – A Qualitative Study With Focus Groups to Identify the Most Burdensome Health-Related Problems Undermining HRQoL in PLHIV

Focus group discussion participants identified many issues that impact the HRQoL of PLHIV. The issue raised most frequently by both PLHIV and healthcare providers was stigma/discrimination (n = 150 segments coded), with people commenting far more on this issue than on physical symptoms or emotional problems. The category of physical symptoms was the second-most frequently discussed (n = 83 segments coded). The physical symptom noted most often was sleep problems. Other physical symptoms that were frequently mentioned included fatigue, pain, body fat changes, and neurocognitive problems. Both PLHIV and healthcare providers emphasised the importance of psychological well-being (n = 67 segments coded). They often commented on emotional distress in general terms rather than naming specific disorders, although depression and anxiety were mentioned numerous times. Healthcare providers, and to a lesser extent PLHIV, called attention to sexuality-related problems such as lack of libido, sexually transmitted infections and general sexual dissatisfaction. When PLHIV addressed sexuality-related problems, they often linked these problems to their perceptions about HIV-related stigma.

Step Three – Development of Potential CST-HIV Items

The initial item pool was developed through the following steps:

(a) Selection and definition of the constructs to include in the CST-HIV. A theoretical conceptualisation of the selected health-related problems undermining HRQoL was carried out, taking into account the content analysis of the FGDs and the literature review. A total of eight constructs were selected: anticipated stigma, emotional distress, sexuality, social support, material deprivation, sleep/fatigue, cognitive problems, and physical symptoms. Three members of the research team wrote independent definitions for the constructs. They then met to reach agreement about definitions and about the essential components that should be included in the instrument.

(b) Development and writing of items. First, we conducted a review of validated instruments measuring the constructs selected for inclusion in the CST-HIV. The same three researchers selected the items that most closely represented the components of each construct. Drawing on these items and the definitions of constructs, two Spanish researchers adapted or wrote six to eight items for each construct. Psychometric recommendations for the development of items were followed (Nunnally and Bernstein, 1994), with the following criteria taken into account: clarity (i.e., items should be written in short, simple and intelligible sentences, and should avoid excessive generality); relevance (i.e., content should be clearly related to the construct); and representativeness (i.e., items should be representative of the construct). This process yielded an initial pool of 47 items.

(c) Expert assessment and inter-rater process. The 18 participating experts rated the items based on their clarity, relevance and representativeness. They also assessed whether the items required modification, and provided further input in comments. This process led to the elimination of seven items. Sixteen other items were modified in response to suggestions from experts. The item pool to be evaluated in the pilot study was comprised of 40 items. **Table 3** shows the items and their Osterlind Index scores for representativeness/relevance. All of the experts agreed on the five-point response format that was proposed for the items.

(d) Cognitive debriefing interview. Eight PLHIV completed a questionnaire containing the selected items, then reported to a member of the research team about possible difficulties in understanding the questionnaire. The items were generally regarded as relevant, accessible, and easy to understand and answer.

Step Four – A Pilot Study to Analyse the Psychometric Properties of the CST-HIV Items

The pilot study enrolled 226 PLHIV. Data collection was carried out with the collaboration of NGOs from the following Spanish cities: Alicante, Barcelona, Bilbao, Madrid, Malaga, and Seville.

Assessment and Selection of the Items

Because our goal was to create a brief instrument that was feasible to use in clinical practice, we had previously decided that no more than three items should be selected for each construct. Any item was eliminated because of ceiling or floor effects. We considered each item's reliability and validity indices to select the three items that would maximise the reliability and representation of each construct. **Table 3** presents all piloted items, indicating their psychometric properties and the retained items. The Spanish wording of items is provided in **Supplementary Table 1**.

Construct Validity: Confirmatory Factor Analysis Results

The Confirmatory Factor Analysis (CFA) results confirmed the eight-factor structure with a good fit to the data. All of the standardised loadings were higher than 0.5, the level considered adequate (Green, 1978). The results of the fully standardised solution including fit indices of the model are displayed in **Table 4**. **Table 5** reports the covariance among factors. The highest covariance was found between the physical symptoms dimension and three other dimensions: emotional distress, sleep/fatigue, and cognitive problems.

Internal Consistency

Despite the low number of items, most of the dimensions presented an alpha index of close to or $\geq$ 0.70, with the notable exception of the physical symptoms dimension (**Table 6**). However, since the number of items is crucial for Cronbach's alpha, values lower than 0.70 for scales with only two or three items may not be considered an indicator of low consistency. As can be seen in **Table 6**, estimates of reliability were higher using the Jöreskog rho (omega) coefficient because the Cronbach's alpha underestimates reliability in ordinal data (Bentler, 2009). Omega is based on the loadings rather than the correlations between the observed variables.

Regarding validity, we calculated the Average Variance Extracted (AVE) values for all variables. All of them except for physical symptoms were above the critical threshold of 0.50, indicating good convergent validity. The AVE measures the amount of variance that is captured by the construct in relation to the amount of variance due to measurement error (Fornell and Larcker, 2016); thus, an AVE value greater than 0.50 indicates that the variance captured by the construct is larger than the amount of variance due to measurement error.

Convergent and Concurrent Validity

We found high positive correlations between the CST-HIV dimensions and the validated measures of the same constructs (**Table 7**). Also, we found correlations in the expected direction between each CST-HIV dimension and the validated instruments used to assess the convergent validity of the other CST-HIV dimensions.

We found negative associations between the eight dimensions of the CST-HIV and the dimensions of HRQoL measured using the disease-specific instrument WHOQOL-HIV-BREF. As can be seen in **Table 8**, most of the correlations were moderate to high. We also found negative associations between most of the

TABLE 3 | Psychometric properties of the initial pool of items of the CST-HIV.

CST-HIV dimensions and items	Osterlind Index (representativeness/ relevance)	Mean ± SD	Skewness	Kurtosis	Corrected item-domain correlation	Cronbach's α if item is deleted	Reliability Index	Validity Index
Anticipated Stigma[a] (α = 0.877)								
During the past month, to what extent have you been worried...								
S1... about telling someone you have HIV?[1]	0.94/1.00	2.62 ± 1.51	0.40	−1.30	0.694	0.859	1.04	0.84
S3... about people judging you if they learn you have HIV?[1]	1.00/0.94	2.95 ± 1.45	0.06	−1.39	0.824	0.807	1.19	0.98
S4... about the idea that you can't find a partner because you have HIV?	0.94/0.94	2.66 ± 1.43	0.32	−1.24	0.618	0.886	0.89	0.74
S5m... about people rejecting you for having HIV?[1,2]	1.00/1.00	2.86 ± 1.43	0.12	−1.33	0.815	0.812	1.16	0.97
Emotional distress[b] (α = 0.901)								
During the past month, how often...								
E1m... have you had negative feelings? (for example, sadness, despair, low spirits, or anxiety?[2])	1.00/1.00	3.04 ± 1.01	−0.35	−0.76	0.789	0.872	0.79	0.64
E2m... have you felt anxiety?[1]	0.88/0.89	2.94 ± 1.11	−0.33	−0.82	0.800	0.869	0.88	0.75
E3m... have you felt sadness or discouraged?[1,2]	0.88/0.83	2.95 ± 1.06	−0.33	−0.85	0.825	0.866	0.86	0.67
E5... have you felt fearful of the future?[1]	0.76/0.83	3.01 ± 1.24	−0.09	−0.99	0.744	0.881	0.92	0.69
E6m... have you been concerned for your future because of having HIV?[2]	0.81/0.88	2.74 ± 1.33	0.14	−1.19	0.647	0.907	0.86	0.66
Sexuality[a] (α = 0.734)								
During the past month...								
Sx1m... how satisfied have you felt with your sex life?[1,2,3]	0.82/0.94	3.03 ± 1.21	−0.23	−0.96	0.310	0.736	0.37	0.88
Sx2... has your sex drive or interest in sex decreased?[1]	1.00/1.00	2.72 ± 1.26	0.02	−1.10	0.380	0.720	0.48	0.52
Sx3m... how difficult has it been to start an intimate or sexual relationship with a new partner?[2]	0.88/0.94	2.83 ± 1.47	0.10	−1.37	0.510	0.685	0.75	0.48
Sx4m... how fearful have you been of being rejected by a sexual partner for having HIV?[2]	0.94/0.94	2.96 ± 1.52	0.04	−1.46	0.616	0.649	0.94	0.29
Sx5... how worried have you been about transmitting HIV to a sexual partner?	1.00/1.00	2.84 ± 1.66	0.14	−1.64	0.458	0.700	0.76	0.06
Sx6... has HIV negatively affected your sex life?[1]	0.94/0.94	2.52 ± 1.33	0.39	−1.03	0.572	0.670	0.76	0.50
Social Support[b] (α = 0.837)								
During the past month, how often...								
SS1m... have you had people around you whom you can lean on in case of need?[1,2,3]	0.88/0.82	3.62 ± 1.18	−0.39	−0.85	0.699	0.787	0.82	0.61
SS2... have you had someone you trust to speak to about your problems?[1,3]	1.00/1.00	3.61 ± 1.21	−0.45	−0.75	0.657	0.800	0.79	0.58
SS3... have people made you feel loved?[1,2]	0.82/0.89	3.83 ± 1.08	−0.67	−0.27	0.739	0.777	0.80	0.63
SS4... have you felt isolated from other people?	1.00/1.00	2.49 ± 1.01	0.13	−0.81	0.486	0.843	0.49	0.48
SS6... have you felt alone?	0.82/0.83	2.83 ± 1.15	−0.10	−0.89	0.622	0.809	0.71	0.55
Material deprivation[a] (α = 0.774)								
During the past month...								
Ex1... how concerned have you been about your economic situation?	0.94/0.89	3.45 ± 1.31	−0.44	0.96	0.580	0.720	0.76	0.46
Ex3... have you had enough money to meet your needs?[1]	0.94/0.94	3.01 ± 1.01	−0.03	−0.80	0.646	0.703	0.65	0.58
Ex4m... how satisfied have you been with the quality of the place where you live?[1,2,3]	0.47/0.61	3.52 ± 1.10	−0.49	−0.48	0.344	0.792	0.38	0.60

(Continued)

TABLE 3 | Continued

CST-HIV dimensions and items	Osterlind Index (representativeness/ relevance)	Mean ± SD	Skewness	Kurtosis	Corrected item-domain correlation	Cronbach's α if item is deleted	Reliability Index	Validity Index
Ex5m... have you had money for leisure activities?[1,2,3]	0.47/0.72	2.71 ± 1.21	0.15	−0.97	0.717	0.672	0.87	0.75
Ex6m... how worried have you been about keeping your home in the short term?[2]	0.88/0.88	2.81 ± 1.39	0.12	−1.26	0.483	0.759	0.66	0.44
Sleep and fatigue*[a]* (α = 0.827)								
During the past month...								
SF1... have you had sleep problems?[1]	1.00/1.00	3.12 ± 1.31	−0.20	−1.04	0.571	0.815	0.75	0.89
SF3... have you had enough energy to do the things you would like to?[3]	1.00/0.94	3.22 ± 1.06	−0.16	−0.73	0.586	0.803	0.62	0.40
SF5m... have you had enough energy for your daily life activities?[2,3]	0.94/0.94	3.29 ± 0.97	−0.08	−0.62	0.627	0.794	0.61	0.44
SF7... how satisfied have you felt with the quality of your sleep?[1,3]	0.94/0.94	2.89 ± 1.14	0.01	−0.82	0.698	0.770	0.79	0.65
SF8... how tired have you felt?[1]	0.88/0.94	3.30 ± 1.01	−0.17	−0.42	0.663	0.782	0.67	0.54
Cognitive problems*[a]* (α = 0.924)								
During the past month...								
CG1... do you feel you've lost memory or capacity to focus or to organise yourself?	0.94/1.00	2.80 ± 1.22	0.09	−0.96	0.769	0.911	0.93	0.71
CG2m... how difficult has it been for you to remember things?[2]	1.00/1.00	2.77 ± 1.13	0.09	−0.73	0.822	0.905	0.93	0.71
CG3... how difficult has it been for you to make decisions?	0.88/0.89	2.71 ± 1.14	0.03	−0.96	0.737	0.916	0.84	0.71
CG4... have you had difficulty thinking clearly?[1]	0.88/0.89	2.55 ± 1.15	0.23	−0.94	0.795	0.908	0.91	0.76
CG7... have you had difficulty paying attention?[1]	0.65/0.67	2.65 ± 1.17	0.12	−0.96	0.836	0.902	0.98	0.76
CG8 ... do you think that it has been harder for you to learn new things?[1]	1.00/1.00	2.62 ± 1.22	0.24	−1.05	0.729	0.917	0.89	0.79
Physical symptoms*[a]* (α = 0.729)								
During the past month...								
PS1m... have you experienced unpleasant body changes such as fat accumulation, weight gain, or weight loss?[1,2]	0.94/0.89	2.68 ± 1.26	0.19	−1.01	0.550	0.649	0.69	0.49
PS2... how worried have you been about experiencing future body changes?	0.76/0.78	3.11 ± 1.27	−0.22	−1.04	0.586	0.627	0.74	0.49
PS3... have you felt pain somewhere in your body? (for example, headache, joint pain, muscle cramps)[1]	0.89/1.00	3.04 ± 1.20	−0.24	−0.84	0.504	0.677	0.60	0.66
PS5... have you suffered digestive problems? (stomach pain, flatulence, diarrhea, nausea, or vomiting)[1]	0.94/0.89	2.62 ± 1.31	0.26	−1.13	0.440	0.715	0.58	0.58

[1] *Item selected for final CST-HIV.*
[2] *Item slightly reworded after inter-judgement process.*
[3] *Reverse item.*
[a] *Response category labels: 1, None; 2, Slightly; 3, Somewhat; 4, Quite; 5, Extremely.*
[b] *Response category labels: 1, Never; 2, Rarely; 3, Sometimes; 4, Frequently; 5, Always.*

CST-HIV dimensions and the generic measure of HRQoL EQ-5D-5L, with the exception of the anticipated stigma and sexuality dimensions (**Table 7**).

CST-HIV Scores

Table 5 reports the CST-HIV dimension scores. These were calculated by adding the values corresponding to each response after recoding the positive items. Thus, higher scores indicate a higher burden in the construct measured in the dimension. All scores were higher than the theoretical mean of the scale except for social support (M = 6.95, SD = 3.10), although that score was close to it. The highest score was found in the sleep/fatigue dimension (M = 9.52, SD = 2.97), followed by emotional distress (M = 8.90, SD = 3.03) and material deprivation (M = 8.76, SD = 2.80).

TABLE 4 | Standardised estimations for the first-order confirmatory factor analysis model.

CST-HIV dimensions and items	Lambda (λ)
Anticipated stigma	
S1	0.70
S3	0.97
S5m	0.98
Emotional distress	
E2m	0.90
E3m	0.93
E5	0.79
Sexuality	
Sx1m[a]	0.57
Sx2	0.76
Sx6	0.82
Social support	
SS1m[a]	0.83
SS2[a]	0.85
SS3	0.93
Material deprivation	
Ex3[a]	0.71
Ex4m[a]	0.61
Ex5m[a]	0.97
Sleep and fatigue	
SF1	0.75
SF7[a]	0.72
SF8	0.88
Cognitive problems	
CG4	0.94
CG7	0.89
CG8	0.79
Physical symptoms	
PS1m	0.60
PS3	0.69
PS5	0.62
SB-χ2	285.09
Degrees of freedom	224
p	0.0036
RMSEA [90% CI]	0.035 (0.021;0.046)
SRMR	0.053
GFI	0.98
AGFI	0.97
CFI	0.99
NFI	0.96

N = 226. Estimation of the robust unweighted least squares. SB-χ^2, Satorra-Bentler chi-square; RMSEA, root mean square error of approximation; CI, confidence interval; SRMR, standardised root mean square residual; GFI, goodness of fit index; AGFI, adjusted goodness of fit index; CFI, comparative fit index; NFI, normed fit index.
[a] Reversed items recoded.
All factor loadings $p < 0.05$.

DISCUSSION

The present paper described the development and psychometric properties of a clinic screening tool to facilitate the rapid identification of problems that undermine the HRQoL of PLHIV in Spain. The results indicate that this new measure could be useful for achieving the intended objective. The CST-HIV showed adequate psychometric properties and evidence of content, face, construct and criterion validity. Although this preliminary evidence of validity should be confirmed in a broad validation study, the results enable us to state that a new brief PROM to identify burdensome problems experienced by PLHIV in routine clinical care is now available.

This new instrument has several strengths. It was developed following a robust methodological process that used both qualitative and quantitative data, in accordance with best practices for ensuring content validity (Pedrosa et al., 2014). The selection of the instrument's content was based on a relevant literature review and on the findings of a qualitative study that included PLHIV and multidisciplinary experts. These procedures allowed us to learn firsthand and from multiple perspectives the problems that undermine the HRQoL of PLHIV in Spain. Findings guided us in determining which issues to prioritise for inclusion in the CST-HIV. The selected issues – anticipated stigma, emotional distress, sexuality, social support, material deprivation, sleep/fatigue, cognitive problems, and physical symptoms – are consistent with research findings about social, psychological, and symptom issues prevalent in Spain (Muñoz-Moreno et al., 2013; Fuster-RuizdeApodaca et al., 2015; Fuster-RuizdeApodaca et al., 2019).

The selected issues are quite similar to those chosen for inclusion in a recent PROM developed by Bristowe et al. (2020) and colleagues on the basis of research conducted in England and Ireland. Those authors have reported the content and face validity of their new instrument. They defined six initial dimensions – physical, cognitive, psychological, welfare, social/relational, and information – and their final version of the instrument is comprised of 23 items. Many of the items are similar to CST-HIV items. However, the other instrument includes some issues that were not considered high priorities by our study participants. These issues included information needs, conception and contraception issues, immigration problems, and alcohol and drug use. Most of these issues also arose during our FGDs, but were not emphasised to the same degree as other issues that we selected for inclusion in our CST-HIV. Several reasons could explain this, such as differences in the epidemiological and socioeconomic profiles of PLHIV whose experiences informed instrument development, differences in the nature of the health-related issues that impose the greatest burden in different settings, and cultural differences that affect how these issues are conceptualised by PLHIV and service providers (Regnault and Herdman, 2015; Nobre et al., 2016). A potential avenue of future research is to explore whether new CST-HIV modules might be developed to add dimensions that are relevant to PLHIV in Spain if this can be done without making the length of the instrument overly burdensome.

After we defined the constructs and drafted the items in accordance with psychometric recommendations, we conducted an inter-judgement process with the participation of 18 multidisciplinary experts, including PLHIV. This

TABLE 5 | Descriptive statistics and covariances (ϕ) between the CST-HIV dimensions

			Covariances (ϕ)							
CST-HIV dimensions	**M**	**SD**	**STG**	**EMD**	**SEX**	**SS**	**MAD**	**SF**	**CG**	**PHYS**
Anticipated stigma	8.43	4.00	1							
Emotional distress	8.90	3.03	0.48	1						
Sexuality	8.21	3.00	0.32	0.55	1					
Social support	6.95	3.10	0.34	0.29	0.39	1				
Material deprivation	8.76	2.80	0.26	0.38	0.40	0.52	1			
Sleep and fatigue	9.52	2.97	0.25	0.67	0.37	0.22	0.36	1		
Cognitive problems	7.82	3.17	0.19	0.63	0.30	0.23	0.26	0.59	1	
Physical symptoms	8.34	2.87	0.23	0.71	0.29	0.22	0.36	0.83	0.78	1

Overall scores for each dimension, comprised by three items, ranged from 3 to 15. M, mean; SD, standard deviation; STG, anticipated stigma; EMD, emotional distress; SEX, sexuality; SS, social support; MAD, material deprivation; SF, sleep and fatigue; CG, cognitive problems; PHYS, physical symptoms.

process guided us to select and reword items with consideration for clarity, relevance and representativeness. Furthermore, we conducted cognitive debriefing interviews that allowed us to test the face validity of the instrument. According to the previous procedures, the CST-HIV seems to be relevant to, and representative of, the targeted constructs that it is designed to measure, and it is subjectively viewed as covering the concepts that it purports to measure.

The pilot study results enabled us to select a 24-item scale considering both the reliability and validity indices of the items. We were able to estimate the validity indices because our study, despite its pilot nature, included convergent measures for each CST-HIV dimension. We selected three items per dimension, ensuring that both consistency and representation of the construct were fulfilled.

This study also provided preliminary evidence of the validity of the internal structure of the instrument. The results confirmed the eight-factor structure that was theoretically proposed. These factors were related to each other with different magnitudes. The highest covariances were found between the physical symptoms dimension and the dimensions of emotional distress, sleep/fatigue, and cognitive problems. Several studies have found relationships between these issues (Muñoz-Moreno et al., 2014; Tedaldi et al., 2015; Uebelacker et al., 2015; Allavena et al., 2016; Redman et al., 2018; Ren et al., 2018; Nogueira et al., 2019; Sabin et al., 2020). The size of the covariances suggests that these four dimensions could be grouped in a second-order latent dimension that encompasses physical, emotional and cognitive concerns. A second validation study is planned using a larger sample, and in that study it will be feasible to analyse the instrument's potential second-order structure.

The results showed that most CST-HIV dimensions presented adequate-to-good internal consistency and construct validity. The physical symptoms dimension was the one that showed the lowest internal consistency and construct validity. This result was not surprising because the dimension included three different physical symptoms, with each measured through one item (body changes, pain, and gastrointestinal problems). We decided not to eliminate the dimension for several reasons. The reliability and validity coefficients were not far from the values considered adequate (Bentler, 2009). Moreover, the construct is theoretically relevant. Several studies have shown that the symptoms included in the dimension are prevalent and burdensome (Edelman et al., 2011; Erdbeer et al., 2014; Wilson et al., 2016; Schnall et al., 2018; Ibarra-Barrueta et al., 2019). Additionally, the size of the correlations found between this dimension and other constructs such as HRQoL and psychological well-being endorse its relevance.

This study also provided preliminary evidence of criterion validity of the CST-HIV. We found high correlations between its dimensions and the measures of the convergent constructs. Furthermore, most dimensions presented moderate-to-high correlations with the HRQoL dimensions, providing evidence of concurrent validity. The anticipated stigma dimension was the one that presented the lowest correlations with the criterion measures. The anticipated stigma dimension includes items measuring HIV disclosure concerns and anticipatory fear of being rejected. Previous research on multiple dimensions of stigma has found that the disclosure concerns dimension was less correlated with HRQoL than other dimensions (Franke et al., 2010; Fuster-RuizdeApodaca et al., 2015). A potential explanation for this finding is the mediating role of other variables such as self-efficacy or coping strategies on the negative impact of some stigma dimensions on HRQoL (Fuster-RuizdeApodaca

TABLE 6 | Construct and reliability statistics of the CST-HIV dimensions.

CST-HIV dimension	Cronbach's alpha	Average variance extracted	Jöreskog rho (omega)
Anticipated stigma	0.800	0.797	0.920
Emotional distress	0.866	0.766	0.907
Sexuality	0.698	0.525	0.764
Social support	0.869	0.759	0.904
Material deprivation	0.765	0.606	0.816
Sleep and fatigue	0.800	0.618	0.828
Cognitive problems	0.874	0.767	0.907
Physical symptoms	0.627	0.407	0.672

TABLE 7 | Correlations between the CST-HIV dimensions and the criterion variables.

	Criterion measures											
CST-HIV dimension	**SSS**	**HADS-A**	**HADS-D**	**PHQ-4**	**PROMIS-SEX**	**DUKE-AC**	**DUKE-AA**	**SEI-HS**	**ISI**	**FSS**	**PROMIS-NQ**	**EQ-5D-5L**
STG	**0.63****	0.33**	0.19**	0.29**	−0.09	−0.15*	−0.19**	0.21**	0.19*	0.16*	0.12	0.09
EMD	0.35**	**0.74****	**0.59****	**0.71****	−0.32**	−0.32**	−0.36**	0.41**	0.54**	0.42**	0.50**	0.47**
SEX	0.23**	0.34**	0.33**	0.33**	**−0.64****	−0.28**	−0.29**	0.35**	0.19**	0.22**	0.20**	0.11
SS	0.23**	0.33**	0.49**	0.28**	−0.38**	**−0.59****	**−0.64****	0.55**	0.16*	0.11	0.19**	0.27**
MAD	0.22**	0.41**	0.42**	0.37**	−0.30**	−0.36**	−0.42**	**0.72****	0.33**	0.19**	0.31**	0.34**
SF	0.19**	0.63**	0.45**	0.52**	−0.30**	−0.25**	−0.28**	−0.32**	**0.71****	**0.36****	0.41**	0.46**
CG	0.10	0.63**	0.54**	0.58**	−0.26**	−0.33**	−0.32**	0.53**	0.46**	0.46**	**0.71****	0.46**
PHYS	0.13	0.59**	0.49**	0.56**	−0.19**	−0.22**	−0.25**	0.31**	0.52**	0.43**	0.52**	**0.52****

CST dimensions: STG, anticipated stigma; EMD, emotional distress; SEX, sexuality; SS, social support; MAD, material deprivation; SF, sleep and fatigue; CG, cognitive problems; PHYS, physical symptoms. Criterion variables: SSS, Spanish Stigma Scale; HADS-A, Hospital Anxiety and Depression Scale–Anxiety Subscale; HADS-D, Hospital Anxiety and Depression Scale–Depression Subscale; PHQ-4, Patient Health Questionnaire-4; PROMIS-SEX, PROMIS V2.0 Satisfaction with Sex Life; Duke-AC, Duke-confidential support dimension; Duke-AA, Duke-affective support dimension; SEI-HS, Social Exclusion Index for Health Surveys; ISI, Insomnia Severity Index; FSS, Fatigue Severity Scale; PROMIS-NQ, PROMIS Neuro-QoL V2.0 Cognitive Function; EQ-5D-5L, generic health-related quality of life.
Correlations in bold: correlations with specific criterion variables.
**$p < 0.05$, **$p < 0.001$.*

TABLE 8 | Correlations between the CST-HIV dimensions and dimensions of health-related quality of life (WHOQOL-HIV-BREF).

	HRQoL dimensions						
CST-HIV dimensions	**General health**	**Physical health**	**Psychological health**	**Level of independence**	**Social relationships**	**Environmental health**	**SRPB**
Anticipated stigma	−0.18**	−0.25**	−0.27**	−0.17**	−0.27**	−0.28**	−0.52**
Emotional distress	−0.54**	-0.57**	−0.66**	−0.49**	−0.47**	−0.46**	−0.69**
Sexuality	−0.35**	−0.27**	−0.33**	−0.25**	−0.47**	−0.36**	−0.37**
Social support	−0.34**	−0.23**	−0.37**	−0.28**	−0.65**	−0.51**	−0.24**
Material deprivation	−0.46**	−0.33**	−0.35**	−0.42**	−0.44**	−0.68**	−0.27**
Sleep and fatigue	−0.51**	−0.62**	−0.54**	−0.48**	−0.37**	−0.39**	−0.48**
Cognitive problems	−0.47**	−0.58**	−0.64**	−0.57**	−0.42**	−0.44**	−0.49**
Physical symptoms	−0.44**	−0.59**	−0.52**	−0.52**	−0.29**	−0.42**	−0.41**

*HRQoL, health-related quality of life; SRPB, spirituality, religion, and personal beliefs. **$p < 0.001$.*

et al., 2015). Although most correlations were small, the anticipated stigma dimension showed a high correlation with the HRQoL domain for spirituality, religion and personal beliefs. This domain of the WHOQOL-HIV-BREF is the one that includes HIV-specific items assessing existential issues and concerns. A previous Spanish study found that the SRPB domain was the unique WHOQOL-HIV-BREF dimension significantly and negatively associated with disclosure concerns (Fuster-RuizdeApodaca et al., 2019). Thus, our current finding provides additional evidence about the relationship between stigma and HIV-specific existential concerns such as those related to fearing the future or feeling that one's life is meaningful. Correlations found between each CST-HIV dimension and HRQoL point to the relevance of the scale for both theory and intervention.

The present study showed that the scores obtained in most of the CST-HIV dimensions were higher than the theoretical mean of the scale, indicating a relevant burden in these dimensions. The highest scores were found in the sleep/fatigue dimension, followed by emotional distress and material deprivation. These results are consistent with a 2019 Spanish HRQoL study in a cohort of 1462 PLHIV who were demographically similar to the overall Spanish PLHIV population. In that study, sleep was the facet most impaired in the physical health HRQoL dimension, and the psychological HRQoL dimension was one of the most impaired dimensions. The financial resources facet had the lowest score of all facets (Fuster-RuizdeApodaca et al., 2019).

This study had several limitations. We conducted an exploratory but not systematic literature review. Further, our priority in designing the CST-HIV was to keep it brief in order to ensure the feasibility of integrating it into clinical practice. This forced us to prioritise the most prevalent and relevant problems according to our content validity sources. Other potentially relevant health-related issues that negatively impact the HRQoL of PLHIV may have been omitted. It is also possible that the most burdensome problems may change over time in accordance with changing factors such as improvements in ART and simplified ART dosing schedules. To offset these limitations, we recommend collecting HRQoL data periodically to assess whether other dimensions will emerge as more burdensome. The desired brevity of the measure led us to choose only three items in each dimension. This could result in low levels of reliability and low construct validity scores in some dimensions. It might also have an impact on the predictive validity of the tool. We plan to test it further in subsequent studies, and we anticipate that by defining risk cut-off points for the scores on all dimensions, we will be able to provide guidance to healthcare providers regarding when findings should be followed up with the administration of other validated PROMs to further investigate specific issues of concern. Moreover, PLHIV are a heterogeneous group, and there are specific sub-groups particularly vulnerable to poor HRQoL (Degroote et al., 2014; Fuster-RuizdeApodaca et al., 2019). Thus, we should analyse the scale invariance as a function of relevant sociodemographic or epidemiological characteristics. This would allow for the generalisation of the model (Vandenberg and Lance, 2000). Moreover, this scale was developed and tested in the Spanish context. Thus, the scale and its factor structure should be tested in samples from other cultures to investigate its applicability in different contexts. As a first step, we will perform the cross-cultural adaptation of the CST-HIV to another European country (Italy).

Despite these limitations, we can conclude that we have a new brief instrument to screen eight significant problems that undermine HRQoL and contribute to poor health outcomes in PLHIV. The CST-HIV appears to have good psychometric properties and good preliminary evidence of validity. We anticipate that our next validity study results will strengthen the present evidence, recommending its use in clinical care in Spain. In addition to conducting the CST-HIV validation study, our other planned research will involve assessing the usefulness, efficacy, feasibility, and acceptability of integrating the CST-HIV and related PROMs into clinical practice.

The use of PROMs has been associated with improvements in clinical care and in health outcomes in fields such as mental health and oncology, and there are unrealised opportunities for the HIV field to integrate PROMs into clinical care in ways that will benefit patients (Fredericksen et al., 2020b; Kall et al., 2020). This new instrument is particularly timely in light of growing interest in the objective of improving HRQoL in PLHIV (Lazarus et al., 2016; Guaraldi et al., 2019). Our research findings are novel because few studies focus on brief screening PROMs that cover the range of biological, psychological and social issues that impair the HRQoL of PLHIV, and the present study is unique in Spain. The clinical care challenges presented by the COVID-19 pandemic underscore the importance of implementing tools that will help PLHIV and their healthcare providers make the best use of limited consultation time (Guaraldi et al., 2020). Using the CST-HIV to gather information about patients' symptoms, concerns, and experiences in advance of clinical appointments could help determine individual consultation models, resulting in greater patient satisfaction and better health outcomes.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Figshare repository [https://doi.org/10.6084/m9.figshare.14216162].

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Hospital Clínic of Barcelona. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KS-H and MJFR conceptualised the study, prepared the focus group scripts, and prepared the first draft of the manuscript. KS-H, MJFR, and JL jointly acquired funding for the study and organised the focus group discussions. MJFR formulated the study design. KS-H conducted literature reviews. MJFR moderated the focus groups with assistance from JL. MJFR conducted the pilot study and was the main analyst with the support of AL. MJFR performed the qualitative analysis of focus group data and reported on results. KS-H, MJFR, and MP contributed to the interpretation of results and participated in the development of the pool of items. MJFR, MP, and DN were involved in the inter-judgement process. All authors participated in revising the manuscript and all authors approved the final manuscript.

FUNDING

Gilead Sciences provided funding to ISGlobal and ViiV Healthcare provided funding to SEISIDA for the present study.

ACKNOWLEDGMENTS

We are grateful to Nuria Balanza and Samantha Duckworth for providing research assistance in early stages of instrument development. We thank SEISIDA, GTt, Hernando Knobel, and Emili Letang for their assistance in organising the focus group discussions and recruiting participants. We also thank the following experts who participated in the focus group discussions: Maribel Alonso, Victor Baceiredo, Jordi Blanch, Alfonso Cabello, Diego García, Emma Fernández, Ramón Ferrando, Juan Sebastián Hernández, Mónica López, Esteban Martínez, Pilar Miralles, Ana Palmar, Marta Pastor de la Cal, Jordi Puig, and Joaquim Roqueta. We also thank the following experts who participated in the inter-judgement process: Arrate Agirrezabal, Victor Baceiredo, Alejandro Bertó, Ramón Espacio, Ramón Ferrando, Almudena García, Jorge García, María Asunción García, Lorena Gil de Montes, Juan Sebastián Hernández, Carlos Iniesta, Cristina Landín, Francesc Martínez, Denise Naniche, Patricia Recio, Jorge Saz, Jordi Puig, and Jorge Valencia. We also thank the NGOs that collaborated in the recruitment and data collection for the pilot study: ACCAVIH, Adhara, Bizkaisida, Apoyo Positivo, GTt, and Fundación 26 de Diciembre. Finally, we thank all of the people living with HIV who participated in all stages of this study. KS-H, DN, and JL acknowledge support to ISGlobal from the Spanish Ministry of Science, Innovation and Universities through the "Centro de Excelencia Severo Ochoa 2019–2023" Programme (CEX2018-000806-S), and from the Government of Catalonia through the CERCA Programme.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.681058/full#supplementary-material

The Spanish Clinic Screening Tool (CST-HIV) items are presented in both Spanish and English in **Supplementary Table 1**.

REFERENCES

Allavena, C., Guimard, T., Billaud, E., De la Tullaye, S., Reliquet, V., Pineau, S., et al. (2016). Prevalence and risk factors of sleep disturbance in a large hiv-infected adult population. *AIDS Behav.* 20, 339–344. doi: 10.1007/s10461-015-1160-5

Andersson, G. Z., Reinius, M., Eriksson, L. E., Svedhem, V., Esfahani, F. M., Deuba, K., et al. (2020). Stigma reduction interventions in people living with HIV to improve health-related quality of life. *Lancet HIV* 7, e129–e140. doi: 10.1016/S2352-3018(19)30343-1

Antiretroviral Therapy Cohort Collaboration. (2017). Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *Lancet HIV* 4, e349–e356. doi: 10.1016/S2352-3018(17)30066-8

Antunes, A., Augusto, I., Parada, P., Okoli, C., Appiah, A., and De los Rios, P. (2020). Treatment challenges, priorities, and relationship with healthcare providers in HIV care: a cross-sectional survey of portuguese adults living with HIV. *Eur. J. Person Center. Healthcare* 8:282. doi: 10.5750/ejpch.v8i3.1873

Ballester-Arnal, R., Gómez-Martínez, S., Fumaz, C. R., González-García, M., Remor, E., and Fuster, M. J. (2016). A Spanish Study on psychological predictors of quality of life in people with HIV. *AIDS Behav.* 20, 281–291. doi: 10.1007/s10461-015-1208-6

Bastien, C. H., Vallières, A., and Morin, C. M. (2001). Validation of the insomnia severity index as an outcome measure for insomnia research. *Sleep Med.* 2, 297–307. doi: 10.1016/S1389-9457(00)00065-4

Batista-Foguet, J. M., and Coenders, G. (2000). Modelos de ecuaciones estructurales. Madrid: La Muralla

Bekele, T., Rourke, S. B., Tucker, R., Greene, S., Sobota, M., Koornstra, J., et al. (2013). Direct and indirect effects of perceived social support on health-related quality of life in persons living with HIV/AIDS. *AIDS Care Psychol. Socio Med. Aspects AIDS/HIV* 25, 337–346. doi: 10.1080/09540121.2012.701716

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika* 74, 137–143. doi: 10.1007/s11336-008-9100-1

Bentler, P. M., and Chou, C. P. (1987). Practical issues in structural modeling. *Socio. Methods Res.* 16:78. doi: 10.1177/0049124187016001004

Brandt, C. P., Jardin, C., Sharp, C., Lemaire, C., and Zvolensky, M. J. (2017). Main and interactive effects of emotion dysregulation and HIV symptom severity on quality of life among persons living with HIV/AIDS. *AIDS Care Psychol. Socio Med. Aspects AIDS/HIV* 29, 498–506. doi: 10.1080/09540121.2016.1220484

Bristowe, K., Clift, P., James, R., Josh, J., Platt, M., Whetham, J., et al. (2019). Towards person-centred care for people living with HIV: what core outcomes matter, and how might we assess them? A cross-national multi-centre qualitative study with key stakeholders. *HIV Med.* 20, 542–554. doi: 10.1111/hiv.12758

Bristowe, K., Murtagh, F. E. M., Clift, P., James, R., Josh, J., Platt, M., et al. (2020). The development and cognitive testing of the positive outcomes HIV PROM: a brief novel patient-reported outcome measure for adults living with HIV. *Health and Qual. Life Outcomes* 18:214. doi: 10.1186/s12955-020-01462-5

Broadhead, W. E., Gehlbach, S. H., de Gruy, F. V., and Kaplan, B. H. (1988). The Duke-UNC functional social support questionnaire: measurement of social support in family medicine patients. *Med. Care* 26, 709–723. doi: 10.1097/00005650-198807000-00006

Campo-Arias, A., and Oviedo, H. C. (2008). Propiedades psicométricas de una escala: la consistencia interna. *Rev. Salud Pública* 10, 831–839.

Catalan, J., Tuffrey, V., Ridge, D., and Rosenfeld, D. (2017). What influences quality of life in older people living with HIV? *AIDS Res. Ther.* 14:22. doi: 10.1186/s12981-017-0148-9

Chuah, F. L. H., Haldane, V. E., Cervero-Liceras, F., Ong, S. E., Sigfrid, L. A., Murphy, G., et al. (2017). Interventions and approaches to integrating HIV and mental health services: a systematic review. *Health Policy Plan.* 32(suppl 4), iv27–iv47. doi: 10.1093/heapol/czw169

Cioe, P. A., Gordon, R. E. F., Williams, D. M., and Kahler, C. W. (2019). The effect of increased physical activity on symptom burden in older persons living with HIV. *AIDS Care* 31, 1548–1554. doi: 10.1080/09540121.2019.1601675

Dalmida, S. G., Koenig, H. G., Holstad, M. M., and Thomas, T. L. (2015). Religious and psychosocial covariates of health-related quality of life in people living with HIV/AIDS. *HIV/AIDS Res. Treat. Open J.* 1:1000HARTOJ1101. doi: 10.17140/HARTOJ-1-101

Degroote, S., Vogelaers, D., and Vandijck, D. M. (2014). What determines health-related quality of life among people living with HIV: An updated review of the literature. *Arch. Public Health* 72, 1–10. doi: 10.1186/2049-3258-72-40

den Daas, C., van den Berk, G. E. L., Kleene, M. J. T., de Munnik, E. S., Lijmer, J. G., and Brinkman, K. (2019). Health-related quality of life among adult HIV positive patients: assessing comprehensive themes and interrelated associations. *Qual. Life Res.* 28, 2685–2694. doi: 10.1007/s11136-019-02203-y

Douab, T., Marcellin, F., Vilotitch, A., Protopopescu, C., Préau, M., Suzan-Monti, M., et al. (2014). Health-related quality of life of people living with HIV followed up in hospitals in France: comparing trends and correlates between 2003 and 2011 (ANRS-VESPA and VESPA2 national surveys). *AIDS Care Psychol. Socio Med. Aspects AIDS/HIV* 26(Suppl 1), 29–40. doi: 10.1080/09540121.2014.906553

Edelman, E., Gordon, K., and Justice, A. C. (2011). Patient and provider-reported symptoms in the post-cART Era. *AIDS Behav.* 15, 853–861. doi: 10.1007/s10461-010-9706-z

Eignor, D. R. (2001). Standards for the development and use of tests: the standards for educational and psychological testing. *Eur. J. Psychol. Assess.* 17, 157–163. doi: 10.1037/14047-013

Emlet, C. A., Fredriksen-Goldsen, K. I., and Kim, H. J. (2013). Risk and protective factors associated with health-related quality of life among older gay and bisexual men living with HIV disease. *Gerontologist* 53, 963–972. doi: 10.1093/geront/gns191

Engler, K., Lessard, D., and Lebouché, B. (2017). A review of HIV-specific patient-reported outcome measures. *Patient* 10, 187–202. doi: 10.1007/s40271-016-0195-7

Erdbeer, G., Sabranski, M., Sonntag, I., Stoehr, A., Horst, H.-A., Plettenberg, A., et al. (2014). Everything fine so far? Physical and mental health in HIV-infected patients with virological success and long-term exposure to antiretroviral therapy. *J. Int. AIDS Soc.* 17(4 Suppl 3), 19673. doi: 10.7448/IAS.17.4.19673

Fekete, E. M., Williams, S. L., Skinta, M. D., and Bogusch, L. M. (2016). Gender differences in disclosure concerns and HIV-related quality of life. *AIDS Care* 28, 450–454. doi: 10.1080/09540121.2015.1114995

Fernandez-Mendoza, J., Rodriguez-Muñoz, A., Vela-Bueno, A., Olavarrieta-Bernardino, S., Calhoun, S. L., Bixler, E. O., et al. (2012). The Spanish version of the Insomnia severity index: a confirmatory factor analysis. *Sleep Med.* 13, 207–210. doi: 10.1016/j.sleep.2011.06.019

Fornell, C., and Larcker, D. F. (2016). Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18, 39–50.

Franke, M. F., Muñoz, M., Finnegan, K., Zeladita, J., Sebastian, J. L., Bayona, J. N., et al. (2010). Validation and abbreviation of an HIV stigma scale in an adult spanish-speaking population in Urban Peru. *AIDS Behav.* 14, 189–199. doi: 10.1007/s10461-008-9474-1

Fredericksen, R. J., Crane, H. M., Lober, W., Fitzsimmons, E., and Short, D. (2020a). *PROgress Evidence Review and Summary: Impact of Administering Patient-Reported Measures and Outcomes (PROs) Within HIV Routine Care*. Available online at: https://progresshivcare.org/assets/PROgress%20Review%20and%20Summary%20FINAL%20Nov2020.pdf (accessed May 21, 2021).

Fredericksen, R. J., Fitzsimmons, E., Gibbons, L. E., Loo, S., Dougherty, S., Avendano-Soto, S., et al. (2020b). How do treatment priorities differ between patients in hiv care and their providers? A mixed-methods study. *AIDS Behav.* 24, 1170–1180. doi: 10.1007/s10461-019-02746-8

Fuster-RuizdeApodaca, M. J., Laguía, A., Safreed-Harmon, K., Lazarus, J. V., Cenoz, S., and Del Amo, J. (2019). Assessing quality of life in people with HIV in Spain: psychometric testing of the Spanish version of WHOQOL-HIV-BREF. *Health Qual. Life Outcomes* 17:144. doi: 10.1186/s12955-019-1208-8

Fuster-RuizdeApodaca, M. J., Molero, F., Holgado, F. P., and Ubillos, S. (2015). Adaptation of the HIV stigma scale in spaniards with HIV. *Spanish J. Psychol.* 18:E66. doi: 10.1017/S1138741615000694

George, S., Bergin, C., Clarke, S., Courtney, G., and Codd, M. B. (2016). Health-related quality of life and associated factors in people with HIV: an Irish cohort study. *Health Q. Life Outcomes* 14, 1–12. doi: 10.1186/s12955-016-0517-4

Green, B. (1978). In defense of measurement. *Am. Psychol.* 33, 664–670. doi: 10.1037/0003-066X.33.7.664

Guaraldi, G., Milic, J., Martinez, E., Kamarulzaman, A., Mussini, C., Waters, L., et al. (2020). Human Immunodeficiency Virus (HIV) Care models during the coronavirus disease 2019 (COVID-19) Era. *Clin. Infect. Dis.* ciaa1864. doi: 10.1093/cid/ciaa1864 [Epub ahead of print].

Guaraldi, G., Milic, J., and Wu, A. W. (2019). What is the measure of success in HIV? The fourth 90: quality of life or healthy aging? *Eur. Geriatr. Med.* 10, 267–274. doi: 10.1007/s41999-019-00168-2

Haladyna, T. M., Downing, S. M., and Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Appl. Measure. Educ.* 15, 309–334. doi: 10.1207/s15324818ame1503_5

Harding, R., Lampe, F. C., Norwood, S., Date, H. L., Clucas, C., Fisher, M., et al. (2010). Symptoms are highly prevalent among HIV outpatients and associated with poor adherence and unprotected sexual intercourse. *Sex. Transmit. Infect.* 86, 520–524. doi: 10.1136/sti.2009.038505

Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual. Life Res.* 20, 1727–1736. doi: 10.1007/s11136-011-9903-x

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

Holgado-Tello, F. P., Carrasco-Ortiz, M. Ã, Del Barrio-Gándara, M. V., and Chacón-Moscoso, S. (2009). Factor analysis of the Big Five Questionnaire using polychoric correlations in children. *Qual. Quant.* 43, 75–85. doi: 10.1007/s11135-007-9085-3

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–García, I., and Vila–Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Qual. Quant.* 44, 153–166. doi: 10.1007/s11135-008-9190-y

Hu, L.-T., and Bentler, P. M. (1995). "Evaluating model fit," in *Structural Equation Modeling. Concepts, Issues, and Applications*, ed. R. H. Hoyle (Thousand Oaks, CA: Sage Publications), 76–99.

Ibarra-Barrueta, O., Mora-Atorrasagasti, O., Legarreta, M. J., Illaro-Uranga, A., and Morillo-Verdugo, R. (2019). Pilot study to develop and validate a questionnaire based on HIV symptoms index. *Farm. Hosp.* 43, 87–93. doi: 10.7399/fh.11127

IBM Corp (2013). *IBM SPSS for Windows, Version 22.0*. Armonk, NY: IBM doi: 10.1080/09064710.2013.867070.

Ikeda, D. J., Nyblade, L., Srithanaviboonchai, K., and Agins, B. D. (2019). A quality improvement approach to the reduction of HIV-related stigma and discrimination in healthcare settings. *BMJ Glob. Health* 4:e001587. doi: 10.1136/bmjgh-2019-001587

Jöreskog, K., and Sörbom, D. (1996). *LISREL 8 User's Reference Guide*. Lincolnwood, IL: Scientific Software International.

Justice, A. C., Holmes, W., Gifford, A. L., Rabeneck, L., Zackin, R., Sinclair, G., et al. (2001). Development and validation of a self-completed HIV symptom index. *J. Clin. Epidemiol.* 54(12 Suppl 1), 8–10. doi: 10.1016/S0895-4356(01)00449-8

Kall, M., Marcellin, F., Harding, R., Lazarus, J. V., and Carrieri, P. (2020). Patient-reported outcomes to enhance person-centred HIV care. *Lancet HIV* 7, e59–e68. doi: 10.1016/S2352-3018(19)30345-5

Katz, M. H. (2006). *Multivariable Analysis*, 2nd Edn. Cambridge: Cambridge University Press.

Kline, P. (2013). *Handbook of Psychological Testing*. London: Routledge.

Kroenke, K., Spitzer, R. L., Williams, J. B. W., and Löwe, B. (2009). An Ultra-brief screening scale for anxiety and depression: the PHQ–4. *Psychosomatics* 50, 613–621. doi: 10.1016/s0033-3182(09)70864-3

Kunisaki, K. M., De Francesco, D., Sabin, C. A., Winston, A., Mallon, P. W. G., Anderson, J., et al. (2021). Sleep disorders in human immunodeficiency virus: a substudy of the pharmacokinetics and clinical observations in People Over Fifty (POPPY) Study. *Open Forum Infect. Dis.* 8:ofaa561. doi: 10.1093/ofid/ofaa561

Lai, J. S., Wagner, L. I., Jacobsen, P. B., and Cella, D. (2014). Self-reported cognitive concerns and abilities: two sides of one coin? *Psycho Oncol.* 23, 1133–1141. doi: 10.1002/pon.3522

Lazarus, J. V., Safreed-Harmon, K., Barton, S. E., Costagliola, D., Dedes, N., Del Amo Valero, J., et al. (2016). Beyond viral suppression of HIV – the new quality of life frontier. *BMC Med.* 14:94. doi: 10.1186/s12916-016-0640-4

Lerdal, A., Kottorp, A., Gay, C., Aouizerat, B. E., Portillo, C. J., and Lee, K. A. (2011). A 7-item version of the fatigue severity scale has better psychometric properties among HIV-infected adults: an application of a Rasch model. *Qual. Life Res.* 20, 1447–1456. doi: 10.1007/s11136-011-9877-8

Logie, C. H., Wang, Y., Kazemi, M., Hawa, R., Kaida, A., Conway, T., et al. (2018). Exploring social ecological pathways from resilience to quality of life among women living with HIV in Canada. *AIDS Care* 30(Suppl 5), S67–S75. doi: 10.1080/09540121.2018.1488031

Maciel, R. A., Klück, H. M., Durand, M., and Sprinz, E. (2018). Comorbidity is more common and occurs earlier in persons living with HIV than in HIV-uninfected matched controls, aged 50 years and older: a cross-sectional study. *Int. J. Infect. Dis.* 70, 30–35. doi: 10.1016/j.ijid.2018.02.009

Mayring, P. (2000). Qualitative content analysis. *Forum Qual. Soc. Res.* 1, 1–10. doi: 10.1111/j.1365-2648.2007.04569.x

McCormack, B., Borg, M., Cardiff, S., Dewing, J., Jacobs, G., Janes, N., et al. (2015). Person-centredness – the "state" of the art. *Int. Pract. Dev. J.* 5, 1–15.

McGowan, J., Sherr, L., Rodger, A., Fisher, M., Miners, A., Johnson, M., et al. (2014). Effects of age on symptom burden, mental health and quality of life amongst people with HIV in the UK. *J. Int. AIDS Soc.* 17(4 Suppl 3):19511. doi: 10.7448/ias.17.4.19511

Miners, A., Phillips, A., Kreif, N., Rodger, A., Speakman, A., Fisher, M., et al. (2014). Health-related quality-of-life of people with HIV in the era of combination antiretroviral treatment: a cross-sectional comparison with the general population. *Lancet. HIV* 1, e32–e40. doi: 10.1016/S2352-3018(14)70018-9

Muñoz-Moreno, J. A., Fuster-Ruiz de Apodaca, M. J., Fumaz, C. R., Ferrer, M. J., Molero, F., Jaen, À, et al. (2013). Quejas cognitivas en personas con infección por el virus de la inmunodeficiencia humana en España: prevalencia y variables relacionadas. *Med. Clín.* 142, 438–444. doi: 10.1016/j.medcli.2013.02.037

Muñoz-Moreno, J. A., Fuster-Ruiz de Apodaca, M. J., Fumaz, C. R., Ferrer, M. J., Molero, F., Jaen, À, et al. (2014). [Cognitive complaints in people with human immunodeficiency virus in Spain: prevalence and related variables]. *Med. Clín.* 142, 438–444. doi: 10.1016/j.medcli.2013.02.037

Nideröst, S., and Imhof, C. (2016). Aging With HIV in the era of antiretroviral treatment: living conditions and the quality of life of people aged above 50 living with HIV/AIDS in Switzerland. *Gerontol. Geriatr. Med.* 2:2333721416636300. doi: 10.1177/2333721416636300

Nobre, N., Pereira, M., Sutinen, J., Canavarro, M. C., Sintonen, H., and Roine, R. P. (2016). Quality of life of people living with HIV/AIDS: a cross-country comparison study of Finland and Portugal. *AIDS Care* 28, 873–877. doi: 10.1080/09540121.2016.1147016

Nogueira, L. F. R., da Fonseca, T. C., Paterlini, P. H., Duarte, A., de, S., Pellegrino, P., et al. (2019). Influence of nutritional status and gastrointestinal symptoms on sleep quality in people living with HIV. *Int. J. STD AIDS* 30, 885–890. doi: 10.1177/0956462419846723

Nunnally, J., and Bernstein, I. (1994). *Psychometric Theory*, 3rd Edn. New York, NY: McGraw-Hill.

Okoli, C., Brough, G., Allan, B., Castellanos, E., Young, B., Eremin, A., et al. (2020). Shared decision making between patients and healthcare providers and its association with favorable health outcomes among people living with HIV. *AIDS Behav.* 25, 1384–1395. doi: 10.1007/s10461-020-02973-4

Olson, B., Vincent, W., Meyer, J. P., Kershaw, T., Sikkema, K. J., Heckman, T. G., et al. (2019). Depressive symptoms, physical symptoms, and health-related quality of life among older adults with HIV. *Qual. Life Res.* 28, 3313–3322. doi: 10.1007/s11136-019-02271-0

Osterlind, S. J. (1989). *Constructing Test Items. In Uma ética para quantos?*. Dordrecht: Springer. doi: 10.1007/s13398-014-0173-7.2

Pedrosa, I., Suárez-Álvarez, J., and García-Cueto, E. (2014). Evidencias sobre la Validez de Contenido: avances teóricos y métodos para su estimación [content validity evidences: theoretical advances and estimation methods]. *Acción Psicoló.* 10:3. doi: 10.5944/ap.10.2.11820

Public Health England (2020). *Positive Voices: the National Survey of People Living with HIV: Findings from the 2017 Survey*. London: Public Health England.

Redman, K. N., Karstaedt, A. S., and Scheuermaier, K. (2018). Increased CD4 counts, pain and depression are correlates of lower sleep quality in treated HIV positive patients with low baseline CD4 counts. *Brain Behav. Immun.* 69, 548–555. doi: 10.1016/j.bbi.2018.02.002

Regnault, A., and Herdman, M. (2015). Using quantitative methods within the Universalist model framework to explore the cross-cultural equivalence of patient-reported outcome instruments. *Qual. Life Res.* 24, 115–124. doi: 10.1007/s11136-014-0722-8

Reinius, M., Wiklander, M., Wettergren, L., Svedhem, V., and Eriksson, L. E. (2018). The relationship between stigma and health-related quality of life in people living with hiv who have full access to antiretroviral treatment: an assessment of Earnshaw and Chaudoir's HIV stigma framework using empirical data. *AIDS Behav.* 22, 3795–3806. doi: 10.1007/s10461-018-2041-5

Ren, J., Zhao, M., Liu, B., Wu, Q., Hao, Y., Jiao, M., et al. (2018). Factors associated with sleep quality in HIV. *J. Assoc. Nurs. AIDS Care* 29, 924–931. doi: 10.1016/j.jana.2018.04.006

Revicki, D. A., Gnanasakthy, A., and Weinfurt, K. (2007). Documenting the rationale and psychometric characteristics of patient reported outcomes for labeling and promotional claims: the Pro Evidence Dossier. *Qual. Life Res.* 16, 717–723. doi: 10.1007/s11136-006-9153-5

Sabin, C. A., Harding, R., Bagkeris, E., Nkhoma, K., Post, F. A., Sachikonye, M., et al. (2018). Pain in people living with HIV and its association with healthcare resource use, well being and functional status. *AIDS* 32, 2697–2706. doi: 10.1097/QAD.0000000000002021

Sabin, C. A., Harding, R., Doyle, N., Redline, S., de Francesco, D., Mallon, P. W. G., et al. (2020). Associations between widespread pain and sleep quality in people with HIV. *J.Acq. Immun. Def. Syndrom.* 85, 106–112. doi: 10.1097/QAI.0000000000002410

Satorra, A. (1990). Robustness issues in structural equation modeling: a review of recent developments. *Qual. Quant.* 24, 367–386. doi: 10.1007/BF00152011

Schnall, R., Jia, H., and Reame, N. (2019). Association between HIV symptom burden and inflammatory cytokines: an analysis by sex and menopause stage. *J. Women's Health* 29, 119–127. doi: 10.1089/jwh.2019.7749

Schnall, R., Siegel, K., Jia, H., Olender, S., and Hirshfield, S. (2018). Racial and socioeconomic disparities in the symptom reporting of persons living with HIV. *AIDS Care Psychol. Socio Med. Aspects AIDS/HIV* 30, 774–783. doi: 10.1080/09540121.2017.1417532

Slater, L. Z., Moneyham, L., Vance, D. E., Raper, J. L., Mugavero, M. J., and Childs, G. (2013). Support, stigma, health, coping, and quality of life in older gay men with HIV. *J. Assoc. Nurs. AIDS Care* 24, 38–49. doi: 10.1016/j.jana.2012.02.006

Smit, M., Brinkman, K., Geerlings, S., Smit, C., Thyagarajan, K., Sighem , Av, et al. (2015). Future challenges for clinical care of an ageing population infected with HIV: a modelling study. *Lancet Infect. Dis.* 15, 810–818. doi: 10.1016/S1473-3099(15)00056-0

Sok, P., Gardner, S., Bekele, T., Globerman, J., Seeman, M. V., Greene, S., et al. (2018). Unmet basic needs negatively affect health-related quality of life in people aging with HIV: results from the positive spaces, healthy places study. *BMC Public Health* 18:644. doi: 10.1186/s12889-018-5391-z

Sweeney, S. M., and Vanable, P. A. (2016). The association of HIV-related stigma to HIV medication adherence: a systematic review and synthesis of the literature. *AIDS Behav.* 20, 29–50. doi: 10.1007/s10461-015-1164-1

Tedaldi, E. M., Minniti, N. L., and Fischer, T. (2015). HIV-associated neurocognitive disorders: the relationship of HIV infection with physical and social comorbidities. *BioMed. Res. Int.* 2015:641913. doi: 10.1155/2015/641913

Teeraananchai, S., Kerr, S. J., Amin, J., Ruxrungtham, K., and Law, M. G. (2017). Life expectancy of HIV-positive people after starting combination antiretroviral therapy: a meta-analysis. *HIV Med.* 18, 256–266. doi: 10.1111/hiv.12421

Tejero, A., Guimerá, E. M., Farré, J. M., and Peri, J. M. (1986). Uso clínico del HAD (Hospital Anxiety and Depression Scale) en población psiquiátrica: Un estudio de su sensibilidad, fiabilidad y validez. *Rev. Psiquiat. Facult. Med. Barcelona* 8, 123–130.

Uebelacker, L. A., Weisberg, R. B., Herman, D. S., Bailey, G. L., Pinkston-Camp, M. M., and Stein, M. D. (2015). Chronic pain in HIV-infected patients: relationship to depression, substance use, and mental health and pain treatment. *Pain Med.* 16, 1870–1881. doi: 10.1111/pme.12799

Van Bergen, A. P. L., Hoff, S. J. M., Schreurs, H., Van Loon, A., and Van Hemert, A. M. (2017). Social Exclusion Index-for Health Surveys (SEI-HS): a prospective nationwide study to extend and validate a multidimensional social exclusion questionnaire. *BMC Public Health* 17:253. doi: 10.1186/s12889-017-4175-1

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, 4–70. doi: 10.1177/109442810031002

Wagener, M. N., van Opstal, S. E. M., Miedema, H. S., van Gorp, E. C. M., and Roelofs, P. D. D. M. (2017). Work-related stigma and disclosure: a daily challenge for people living with HIV. *Work* 58, 537–548. doi: 10.3233/WOR-172650

Weinfurt, K. P., Lin, L., Bruner, D. W., Cyranowski, J. M., Dombeck, C. B., Hahn, E. A., et al. (2015). Development and initial validation of the PROMIS® sexual function and satisfaction measures version 2.0. *J. Sex. Med.* 12, 1961–1974. doi: 10.1111/jsm.12966

Wheat, H., Horrell, J., Valderas, J. M., Close, J., Fosh, B., and Lloyd, H. (2018). Can practitioners use patient reported measures to enhance person centred coordinated care in practice? A qualitative study. *Health Qual. Life Outcomes* 16:223. doi: 10.1186/s12955-018-1045-1

Wilson, N. L., Azuero, A., Vance, D. E., Richman, J. S., Moneyham, L. D., Raper, J. L., et al. (2016). Identifying symptom patterns in people living with HIV disease. *J. Assoc. Nurs. AIDS Care* 27, 121–132. doi: 10.1016/j.jana.2015.11.009

World Health Organization (WHO) (2016). *Framework on Integrated, People Centred Health Services*. Available online at: https://apps.who.int/gb/ebwha/pdf_files/WHA69/A69_39-en.pdf?ua=1 (accessed May 21, 2021).

Conflict of Interest: MJFR and MP are members of the Spanish Interdisciplinary AIDS Society (SEISIDA) Executive Board. SEISIDA has received grants from Gilead, Janssen, MSD and ViiV, outside of the submitted work. MJFR has provided consultancy services to Gilead, Janssen, MSD and ViiV, and has received payments for lectures or educational presentations from Gilead, Janssen, MSD and ViiV, outside of the submitted work. JL reports grants and speaker fees from AbbVie and MSD, outside of the submitted work, speaker fees from Gilead Sciences and ViiV, and an institutional grant from Gilead Sciences for this study.

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 July 2021
doi: 10.3389/fpsyg.2021.615581

Having the Cake and Eating It Too: First-Order, Second-Order and Bifactor Representations of Work Engagement

Janos Salamon[1,2,3*†], *István Tóth-Király*[4†], *Beáta Bõthe*[5†], *Tamás Nagy*[2†] and *Gábor Orosz*[6†]

[1] Doctoral School of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary, [2] Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary, [3] Department of Ergonomics and Psychology, Budapest University of Technology and Economics, Budapest, Hungary, [4] Department of Psychology, Concordia University, Montreal, QC, Canada, [5] Département de Psychologie, Université de Montréal, Montreal, QC, Canada, [6] ULR 7369 -URePSSS - Unité de Recherche Pluridisciplinaire Sport Santé Société, Sherpas, Univ. Lille, Univ. Artois, Univ. Littoral Côte d'Opale, Lille, France

OPEN ACCESS

Edited by:
Holmes Finch,
Ball State University, United States

Reviewed by:
Susana Sanduvete-Chaves,
Seville University, Spain
José Antonio Lozano Lozano,
Universidad Autónoma de Chile, Chile

***Correspondence:**
Janos Salamon
salamon.jon@gmail.com;
salamon.janos@ppk.elte.hu

†ORCID:
Janos Salamon
orcid.org/0000-0002-4005-7090
István Tóth-Király
orcid.org/0000-0003-2810-3661
Beáta Bõthe
orcid.org/0000-0003-2718-4703
Tamás Nagy
orcid.org/0000-0001-5244-0356
Gábor Orosz
orcid.org/0000-0001-5883-6861

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

Received: *09 October 2020*
Accepted: *28 June 2021*
Published: *22 July 2021*

Citation:
Salamon J, Tóth-Király I, Bõthe B, Nagy T and Orosz G (2021) Having the Cake and Eating It Too: First-Order, Second-Order and Bifactor Representations of Work Engagement. Front. Psychol. 12:615581. doi: 10.3389/fpsyg.2021.615581

Even though work engagement is a popular construct in organizational psychology, the question remains whether it is experienced as a global construct, or as its three components (vigor, dedication, absorption). The present study thus contributes to the ongoing scientific debate about the dimensionality of work engagement systematically compared one-factor, first-order, higher-order, and bifactor confirmatory factor analytic (CFA) representations of work engagement measured by the short version of Utrecht Work Engagement Scale (UWES-9). We also documented the validity evidence of the most optimal representation based on its test-criterion relationship with basic psychological need fulfillment at work, turnover intentions, work addiction, and work satisfaction. Based on responses provided by two distinct samples of employees (N_1 = 242, N_2 = 505), our results supported the superiority of the bifactor-CFA representation including a global factor of work engagement and three co-existing specific factors of vigor, dedication, and absorption. This representation replicated well across the two samples through tests of measurement invariance. Finally, while global work engagement was substantially related to all correlates, the specific factors also demonstrated meaningful associations over and above the global levels of work engagement.

Keywords: work engagement, validity evidence based on test-criterion relationship, bifactor-CFA, work addiction, work satisfaction, basic psychological needs

INTRODUCTION

Following the changes in work conditions and technological advancements over the last decades, employees invest more and more time and energy in their work (van Beek et al., 2012). This heavy work investment can be conceptualized in the form of work engagement which has been described as a positive and fulfilling, work-related state of mind (Schaufeli et al., 2002) characterized by three components: vigor (i.e., having high levels of energy during work), dedication (i.e., perceiving work as being important and meaningful), and absorption (i.e., being immersed in work). Work engagement is thus a high activation state of mind that is associated with pleasant work-related emotions (Bakker and Oerlemans, 2011). Research has generally demonstrated that work engagement is a desirable state of mind that is positively associated with psychological health (Simbula et al., 2013; Gillet et al., 2019), psychological capital (Mills et al., 2012), occupational self-efficacy (Simbula et al., 2013; Villotti et al., 2014), passion at work (Tóth-Király et al., 2021), work performance (Gorgievski et al., 2010; Alessandri et al., 2015), personal development (Simbula et al., 2013), organizational commitment (Hallberg and Schaufeli, 2006), and job satisfaction (Wefald et al., 2012; Schaufeli et al., 2019).

Despite these findings, the dimensionality of work engagement remains questionable and is frequently investigated in the scientific literature, with two perspectives being prevalent. The first perspective (e.g., Balducci et al., 2010) proposes that the three specific components of work engagement are experienced separately, while the second perspective (e.g., Alessandri et al., 2015) proposes that work engagement is often experienced holistically, as a global construct. The present study was designed with the aim of bringing together these two diverging perspectives by showing that one can "have the cake and eat it too"; that is, one could simultaneously take into account the global and specific nature of work engagement. To achieve this goal, we first compared alternative first-order, second-order, and bifactor confirmatory factor analytic (CFA) models of the 9-item Utrecht Work Engagement Scale (UWES-9; Schaufeli et al., 2006) across two distinct samples of Hungarian[1] employees to identify the most adequate representation of work engagement. Second, via tests of measurement invariance, we investigated the generalizability of the most optimal representation across the two samples. Third, we investigated the relations between this improved representation and key work-related correlates of work engagement, namely basic psychological need fulfillment at work, turnover intentions, work addiction, and work satisfaction.

The Dimensionality of Work Engagement

While the 17-item Utrecht Work Engagement Scale (UWES-17) was developed first by Schaufeli et al. (2002) as a measure of work engagement, the present study focuses on the shorter, 9-item version (UWES-9, Schaufeli et al., 2006) whose factor structure was investigated in numerous studies and validated in many countries. We were able to identify a total of 33 independent studies that investigated the factor structure and reliability of the UWES-9 (more details are provided in **Supplementary Table 1** in the online supplements). These studies were conducted in a large variety of nations (e.g., Netherlands, Sweden, South Korea, United States, Italy) using samples that differed not just in size, but age composition as well. Generally speaking, these studies showed that the specific components of work engagement (i.e., vigor, dedication, and absorption) had at least moderate levels of internal consistency in some studies (e.g., Chaudhary et al., 2012), but also satisfactory levels of internal consistency in most studies ranging between 0.70 and 0.92.

Although studies supported the generally adequate reliability of the UWES-9, contradictory findings have been reported about its factor structure and, in turn, the dimensionality of work engagement. Findings in most of the studies (25 out of the 33) align with the first perspective about the specific work engagement components. Consequently, these studies reported support for the three-factor model as the most optimal solution, which incorporated the three intercorrelated specific components of work engagement, but not the global work engagement construct. Based on commonly used goodness-of-fit indices (such as CFI, TLI, and RMSEA), only nine out of the 25 studies (Schaufeli et al., 2006; Nerstad et al., 2009; Seppälä et al., 2009; Breevaart et al., 2012; Fong and Ng, 2012; Yusoff et al., 2013; Panthee et al., 2014; Lathabhavan et al., 2017; Moreira-Fontán et al., 2019) reported empirical support for the three-factor solution without any model modification. It is interesting to note that ten studies (Samples 1 and 2 of Littman-Ovadia and Balducci, 2013; Ho Kim et al., 2017; Kulikowski, 2019; Sample 1 of Mills et al., 2012; Wefald et al., 2012; Villotti et al., 2014; Vazquez et al., 2015; Petroviæ et al., 2017; Zeijen et al., 2018) chose the three-factor solution as the most optimal one even though the three-factor solution in these studies failed to achieve an acceptable level of fit. In the remaining six studies, the authors opted to modify the three-factor solution by including correlated uniquenesses between a subset of items (Samples 1 and 2 of Balducci et al., 2010; Chaudhary et al., 2012; Simbula et al., 2013; Zecca et al., 2015; Lovakov et al., 2017). However, the *ad hoc* inclusion of correlated uniquenesses for the artificial improvement of model fit is considered to be problematic without any substantive interpretation of why the uniquenesses of a particular subset of items should be allowed to correlate (Marsh, 2007; Marsh et al., 2010).

Despite studies supporting the relative adequacy of the three-factor solution, it has to be noted that the average correlation between vigor, dedication, and absorption was often so high (ranging from 0.57 to 0.97) that it questions the validity evidence based on relations to other variables, specifically discriminant evidence of these components. Consequently, it has been suggested in the literature that the global construct of work engagement, and not its specific components, should be in the focus of investigations. The presence of a global work engagement factor could be investigated in different ways, with the first being the estimation of a one-factor solution that only incorporates a single work engagement factor. Three studies reported this solution as the most optimal model. However, model fit indices were not unanimously adequate in these studies (study 2 of Mills et al., 2012; Vallières et al., 2017). Although the one-factor solution reported by Klassen et al. (2012) was adequate,

[1] We carried out this study in Hungary which provided us with a unique context for multiple reasons. First, recent national surveys show that Hungarian people spend a lot of time with work, around 43-44 hours per week (Kun et al., 2020; Urbán et al., 2019). Second, at the same time, Hungarian employees are substantially less engaged with their work when compared to other European countries (Schaufeli, 2018). This discrepancy (i.e., working a lot but not being engaged with it) thus creates a unique research environment that could provide further insights into the nature of work engagement.

the inclusion of correlated uniquenesses limits the adequacy of their findings. The fourth study that supported the one-factor solution (Hallberg and Schaufeli, 2006) simultaneously accepted the three-factor solution, while neither model reached an acceptable level of RMSEA.

As a second way of testing the presence of a global construct, Sinval et al. (2018) estimated a second-order model in which a global work engagement factor was responsible for the associations between the three first-order specific factors. However, the fit indices were marginally acceptable only in one of their samples, and not unanimously acceptable in another sample, suggesting that this particular representation might not be the most optimal.

Psychometrically, however, second-order models have one important limitation: they assume that the ratio of variance explained by the global factor relative to that explained by the specific factors is the same for all items related to the specific first-order factor (Reise, 2012; Gignac, 2016). This proportionality constraint, however, has been shown to be overly strict and rarely verified in practice (Gignac, 2016; Morin et al., 2016a). Alternatively, bifactor modeling has been proposed as flexible alternative that does not rely on such an unrealistic assumption. More importantly, bifactor modeling makes it possible to directly test the simultaneous presence of a global (G-) factor (i.e., global levels of work engagement underlying responses to all items) and co-existing specific (S-) factors (i.e., unique specificities not explained by the global factor).

To the best of our knowledge, there has only been a single study that tested the adequacy of bifactor solutions. de Bruin and Henn (2013) compared first-order and bifactor solutions and reported a partial bifactor solution (including 1 G- and 2 S-factors) as the most optimal. This partial bifactor model was characterized by a well-defined work engagement G-factor and two more weakly defined vigor and absorption S-factors. The authors did not estimate a third S-factor and argued that all the variance in the dedication items was absorbed by the G-factor, leaving no residual specificity to the dedication S-factor. Other studies relying on the longer version of the UWES also showed the added value of estimating a bifactor representation of work engagement (e.g., Gillet et al., 2018, 2019).

Based on these contradictory findings, there is still a debate on whether work engagement should be measured as a single overarching construct or via its three components. Bifactor modeling appears to be a promising avenue that could bring together the two diverging perspectives and show that work engagement might be characterized by a global dimension *and* co-existing specific components not explained by the global factor. The directly related findings of de Bruin and Henn (2013) and the indirectly related findings of Gillet et al. (2018, 2019) appear to lend support for our proposition, and allow us to propose the following hypothesis:

> **Hypothesis 1.** The bifactor representation of work engagement will be the most optimal compared to the alternative first-order and second-order representation and it will replicate well across the two independent samples.

Validity of Work Engagement Based on Its Test-Criterion Relationship

Beyond the structural analysis of work engagement, we also aimed to investigate its validity evidence based on test-criterion relationship (American Educational Research Association et al., 2014). For this purpose, we relied on a diverse set of theoretically relevant work-related constructs that showed meaningful associations with work engagement in prior studies, namely basic psychological need fulfillment at work, turnover intentions, work addiction, and work satisfaction.

Self-determination theory (SDT; Ryan and Deci, 2017), a macro-theory of human motivation, posits that there exist three *basic psychological needs* whose fulfillment is essential for optimal functioning, growth, and health (Deci and Ryan, 2000). The three needs are the need for autonomy (i.e., the experience of personal volition), the need for competence (i.e., the experience of mastery and efficacy), and the need for relatedness (i.e., the experience of having meaningful relationships with others). These needs are also thought to be universal, a proposition that is supported by studies conducted in the field of, for instance, education (Cox and Williams, 2008), health (Tóth-Király et al., 2019c) or sports (Adie et al., 2008). Not surprisingly, the importance of need fulfillment has also been highlighted in the domain of work (for a review, see Van den Broeck et al., 2016). There have been some studies which focused on the associations between work engagement and need fulfillment at work with most studies reporting moderate-to-strong associations between them regardless of relying on global levels of work engagement or its specific components (Shuck et al., 2015; Trépanier et al., 2015; Wang et al., 2018). The same associations remained present when reported between work engagement and basic psychological need fulfillment specific factors (Gillet et al., 2015; Goodboy et al., 2017). However, to the best of our knowledge, there are no prior studies that assessed the relationship between work engagement and need fulfillment while, at the same time, taking into account both their global and specific components.

Turnover intentions have long been regarded as a key variable of interest in organizations given that frequent turnovers imply substantial organizational costs both directly (e.g., constant recruitment and replacement of staff) and indirectly (e.g., the loss of organizational knowledge and the decrease in productivity; Fernet et al., 2017). Studies so far (Mills et al., 2012; Wefald et al., 2012; Lovakov et al., 2017) have reported moderate and negative associations between global levels of work engagement and turnover intentions, typically varying between −0.43 and −0.48. Albeit slightly weaker, the same associations have also been reported when studies focused on the three components of vigor (varying between −0.38 to −0.46), dedication (varying between −0.38 and −0.51), and absorption (varying between −0.31 and −0.36).

As a downside of work engagement, *work addiction* has been described as an extreme and unhealthy form of work involvement (Porter, 1996) that is associated with, for instance, psychiatric difficulties (Andreassen et al., 2016) and poorer work performance (Falco et al., 2013). From an organizational perspective (e.g., Schaufeli et al., 2009), work addiction is typically defined as an uncontrollable and compulsive need for excessive work; from a clinical perspective (Griffiths, 2005), work addiction is best understood as a constellation of components of behavioral addictions. However, recent theoretical works (Andreassen et al., 2018) acknowledge that both perspectives refer to the same underlying phenomenon. The relationship between work engagement and work addiction has been extensively investigated. Most prior studies generally showed weak, positive association between work addiction and global levels of work engagement (e.g., van Beek et al., 2012; Clark et al., 2014; Schaufeli et al., 2019) with only a few exceptions which reported either weak negative or non-significant associations (Zeijen et al., 2018; Schaufeli et al., 2019). Results become more nuanced when the specific components of work engagement are investigated. More specifically, studies typically reported work addiction having meaningful associations with the absorption component of work engagement, but not with vigor and dedication (Schaufeli et al., 2008; van Beek et al., 2012; Clark et al., 2016). The association between workaholism and absorption might be attributed to the fact that both engaged workers and workaholics are immersed in their work and might find it difficult to disengage from it.

Finally, the present study also included *work satisfaction* as it is considered to be a positive component of employee's wellbeing at work (Ryan and Deci, 2001) that is informative of employees' functioning (e.g., Faragher et al., 2005). Research focusing on the associations between work satisfaction and global levels of work engagement has generally shown positive relations between them as well as between work satisfaction and vigor (varying between 0.41 and 0.65), dedication (varying between 0.42 and 0.73), and absorption (varying between 0.36 and 0.58) (e.g., Schaufeli et al., 2008; Simbula et al., 2013; Littman-Ovadia et al., 2014).

Overall, these previous studies allow us to propose the following hypotheses:

> **Hypothesis 2.** Global levels of work engagement will be positively related to (2a) basic psychological need fulfillment at work, (2b) work addiction, (2c) work satisfaction, and (2d) negatively to turnover intentions.

Research Question

Given the lack of prior studies with regards to the validity evidence of work engagement based on its test-criterion relationship of the bifactor representation of work engagement, as well as the distinctness of first-order and bifactor S-factors, we leave it as an open research question whether the S-factors in the bifactor representation will demonstrate any additional associations with the correlates over and above of the G-factor.

METHODS

Procedure and Participants

The present study was conducted in accordance with the Declaration of Helsinki and with the approval of the Institutional Review Board of Eötvös Loránd University Faculty of Education and Psychology. Participants for this study were recruited through company mailing lists as well as through social media groups. Potential participants were informed about the content of the online survey and they had to explicitly indicate their intention for participation. Sample 1 was collected in January-September 2018 and Sample 2 was collected in January-April 2019, allowing us to minimize their overlap. Although the online survey did not collect any specific information that would make the identification of the participants possible, a duplicate check was conducted based on the combinations of the collected demographic and job-related information. This procedure showed no duplicates in either of the final databases, suggesting the presence of distinct participants in both samples. In addition, only participants working at the time of the data collection were included in the study (which was ensured by asking participants explicitly to indicate whether they worked at the time they responded to the survey).

Two samples were used in the current study. Participants in both samples were employees in a wide variety of organizations and job roles across Hungary. These samples were not representative of the population of Hungarian working adults. Sample 1, recruited between January-September 2018, consisted of 242 working adults (184 females, 76%) who were aged between 18 and 73 years ($M_{\text{Sample1}} = 35.81$, $SD_{\text{Sample1}} = 13.46$) and worked in different organizational levels (48 blue collars: 20%, 136 white collars: 56%, 58 managers: 24%). Sample 2, recruited between February-April 2019, consisted of 505 working adults (359 female, 71%) who were aged between 20 and 71 years ($M_{\text{Sample2}} = 37$, $SD_{\text{Sample2}} = 11.27$), and worked in different organizational levels (75 blue collars: 15%, 287 white collars: 57%, 143 managers: 28%).

Measures

Work Engagement (Both Sample 1 and 2)

The short version of the Utrecht Work Engagement Scale (UWES-9, Schaufeli et al., 2006) was used that measures the three underlying dimensions of work engagement: vigor (three items; e.g., "At my work, I feel bursting with energy"), dedication (three items; e.g., "I am enthusiastic about my job"), and absorption (three items; e.g., "I get carried away when I'm working"). See **Supplementary Appendix 1** in the online supplements for the Hungarian version. Responses were provided on a seven-point Likert-scale ranging from 1 (never) to 7 (always). The UWES-9 was adapted with a standardized translation-back translation protocol proposed by Beaton et al. (2000). Cronbach alpha values for all the factors indicated good internal consistency in both samples, ranging from 0.88 (absorption) to 0.90 (dedication) in Sample 1 and from 0.85 (vigor) to 0.90 (dedication) in Sample 2.

Turnover Intention (Sample 1)

A three-item scale adapted from the questionnaire developed to measure high school dropout intention (Vallerand et al., 1997; Hardre and Reeve, 2003) was used to measure workers' turnover intentions. Items were translated following the standardized translation-back translation protocol proposed by Beaton et al. (2000) and slightly modified to reflect turnover intention in the work context (e.g., "I will likely be looking for a new job soon."). Each item was scored on a five-point Likert-scale ranging from 1 (very uncharacteristic) to 5 (very characteristic). Cronbach's alpha in the present study was 0.93.

Basic Psychological Need Fulfillment (Sample 1)

The Hungarian version (Tóth-Király et al., 2018) of the 24-item Basic Psychological Need Satisfaction and Frustration Scale (BPNSFS, Chen et al., 2015) was used to measure individuals' work-related need satisfaction and frustration. Instructions were slightly adapted to the work context (all items started with the clause "At the workplace where I work...") while the items themselves were used without any modification. The scale measures six factors: autonomy satisfaction (four items; e.g., "I feel that my decisions reflect what I really want."; α = 0.78), relatedness satisfaction (four items; e.g., "I feel close and connected with other people who are important to me."; α = 0.78), competence satisfaction (four items; e.g., "I feel I can successfully complete difficult tasks."; α = 0.70), autonomy frustration (four items; e.g., "My daily activities feel like a chain of obligations."; α = 0.64), relatedness frustration (four items; e.g., "I feel the relationships I have are just superficial."; α = 0.78), and competence frustration (four items; e.g., "I have serious doubts about whether I can do things well."; α = 0.77). Respondents indicated their level of agreement using a seven-point Likert-scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Work Addiction (Sample 2)

The seven-item Hungarian version (Orosz et al., 2016) of the Bergen Work Addiction Scale (BWAS-H, Andreassen et al., 2012) was administered to measure work addiction based on the components model of addiction (Griffiths, 2005), including salience, tolerance, withdrawal, mood modification, tolerance, and relapse (e.g., "How often during the last year have you deprioritized hobbies, leisure activities, and exercise because of your work?"). Cronbach's alpha for this scale was satisfactory (α = 0.78). Items were rated on a five-point scale (1 = never, 5 = always).

Work Satisfaction (Sample 2)

A five-item scale adapted from the Satisfaction with Life Scale (Diener et al., 1985; Martos et al., 2014) was used to measure respondents' satisfaction with their works. Following prior applications (Fouquereau and Rioux, 2002; Tóth-Király et al., 2020), items were modified to refer to work instead of life in general (e.g., "The conditions of my work are excellent"). l. This modified scale indicated good internal consistency (α = 0.87). Respondents indicated their level of agreement using a seven-point Likert-scale ranging from 1 (strongly disagree) to 7 (strongly agree).

Statistical Analyses

Statistical analyses were performed with SPSS 22 and M*plus* 8 (Muthén and Muthén, 1998-2017). For factor analyses, the robust maximum likelihood estimator (MLR) was used as this estimator robust to non-normality and is more preferable when the response scale has more than five categories (Morin et al., 2020). The first step of the analyses comprised of the estimation of four alternative CFA solutions (see **Figure 1** for a graphical depiction of these models): (1) a one-factor solution; (2) a first-order (including the 3 specific factors); (3) a second-order (including the 3 specific factors and a higher-order work engagement factor); and a (4) bifactor solution (including the 3 specific factors and a co-existing work engagement factor). All these models were estimated separately for the two samples. In the three-factor CFA solution, items were set to load only on their *a priori* specific factors, cross-loadings were set to be zero, and factors were allowed to correlate with one another. In the second-order model, specifications were the same as in the first-order model, but the correlations between the factors were replaced by a second-order global work engagement factor. In bifactor-CFA solution, items were set to load on their respective S-factors as well as on the work engagement G-factor, and following typical bifactor specifications (Reise, 2012) factors were specified as orthogonal (i.e., not allowed to correlate with one another). In the comparison of first-order and bifactor models, we followed the guidelines of Morin et al. (2016a) and apart from goodness-of-fit, we also carefully examined the standardized parameter estimates with an emphasis on the size of the correlations between the factors.

In the second stage, using the most optimal measurement model, tests of measurement invariance were conducted (Meredith, 1993; Millsap, 2011) across samples (Sample 1 vs. Sample 2) to ascertain that we relied on identical sets of indicators when investigating validity evidence based on test-criterion relationship and to test the replicability of the measurement structure. In addition, to assess the generalizability of the most optimal model to subgroups of people, we conducted the same tests of measurement invariance across groups based on gender (male vs. female), age (young adult vs. middle-old adult), and organizational level (blue collar employee vs. white collar employee vs. managers). Following typical specifications, tests of measurement invariance were conducted in a sequence where equality constraints are gradually added to the various parameters, ranging from the least restrictive model to the most restrictive one (Millsap, 2011): configural invariance (i.e., factor structure), weak invariance (i.e., factor structure and factor loadings), strong invariance (i.e., factor structure, factor loadings and intercepts), strict invariance (factor structure, factor loadings, intercepts, and uniquenesses), latent variance-covariance invariance (factor structure, factor loadings, intercepts, uniquenesses, factor variances and factor covariances), and latent mean invariance (factor structure, factor loadings, intercepts, uniquenesses, factor variances, factor covariances, and latent means).

Models were evaluated on the basis of common goodness of fit indices and interpreted along their commonly used cut-off values (Hu and Bentler, 1999; Marsh et al., 2005): the Comparative Fit Index (CFI; $\geq$ 0.95 good, $\geq$ 0.90 acceptable), the Tucker–Lewis Index (TLI; $\geq$ 0.95 good, $\geq$ 0.90 acceptable), the Root-Mean-Square Error of Approximation (RMSEA; $\leq$ 0.06 good, $\leq$ 0.08 acceptable) with its 90% confidence interval. It has to be noted the RMSEA has been shown to tends to be overinflated under conditions of low degrees of freedom (Kenny et al., 2015); therefore, this indicator is reported for the sake of transparency and comparability with previous studies, but less emphasis will

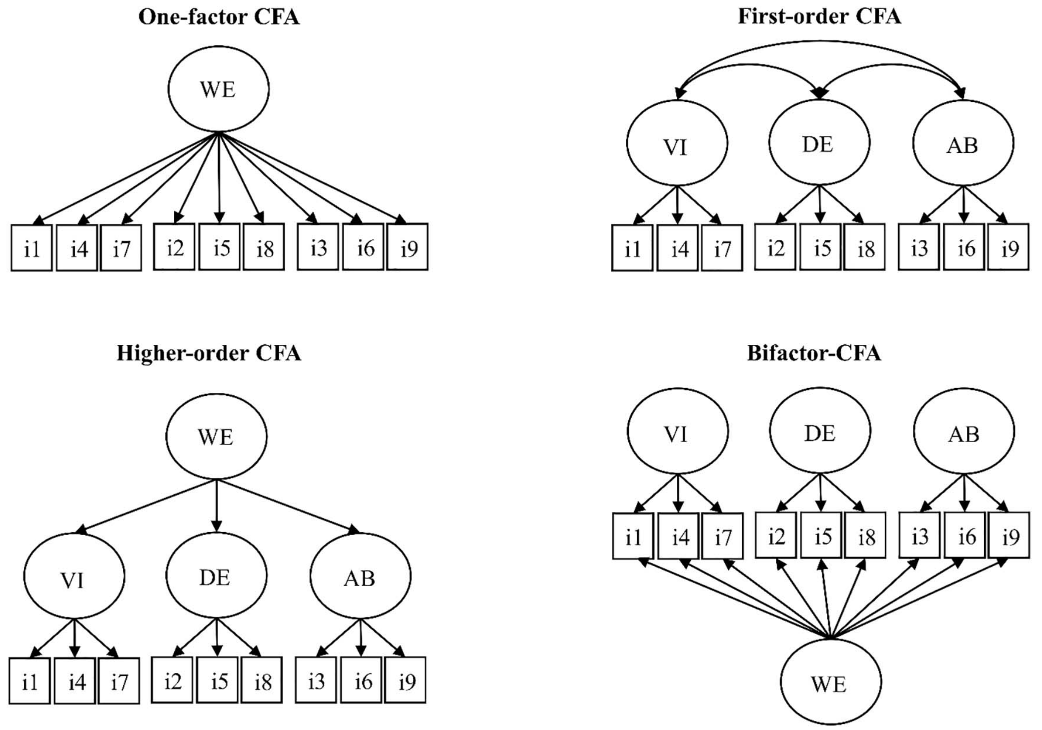

FIGURE 1 | Schematic representation of the estimated model for work engagement. *Note.* CFA, confirmatory factor analysis; i1-i9, item 1-9; VI, vigor; DE, dedication; AB, absorption; WE, work engagement. Unidirectional arrows represent factor loadings, bidirectional arrows represent correlations.

be put on its interpretation. As for measurement invariance, relative changes (Δ) in the fit indices were examined (Cheung and Rensvold, 2002; Chen, 2007) where a decrease of at least 0.010 for CFI and TLI and an increase of at least 0.015 for RMSEA indicate lack of invariance. We also calculated the root deterioration per restriction (RDR; Browne and Du Toit, 1992) index which rescales the chi-square difference to approximate an RMSEA metric. Following suggestions by Raykov and Penev (1998); see also Pekrun et al., 2019), RDR was interpreted in relation to RMSEA (i.e., RDR $<$ 0.05 indicates strong equivalence, RDR $<$ 0.08 indicates acceptable equivalence). Spearman correlations were calculated between the factors to assess the validity evidence of the bifactor-CFA solution based on its test-criterion relationship. Reliability was assessed with the model-based omega composite reliability coefficient (McDonald, 1970; Morin et al., 2020) and values above 0.500 are considered adequate (Perreira et al., 2018). All questions were mandatory; therefore, the sample sizes were the same for all analyses. The data can be found on the following link: https://osf.io/upn9c/?view_only=8fd4125ad1654e32b7219ba29aaa0ecf.

RESULTS

Structural Analysis and Measurement Invariance

Goodness-of-fit statistics of the UWES-9 can be seen in **Table 1**. The one-factor solution (S1M1 and S2M1) had poor fit in both samples. The three-factor CFA model (S1M2 S2M2) had marginally acceptable fit in Sample 1 (although RMSEA did not reach the minimum 0.080), and acceptable fit in Sample 2 (CFI and TLI $>$ 0.90, RMSEA = 0.08). Correlations between the three engagement factors were high in both Sample 1 (between 0.778 and 0.887, M = 0.827) and Sample 2 (between 0.773 and 0.907, M = 0.850), suggesting conceptual redundancies between the three factors. However, the magnitude of these correlations might be inflated by an unmodeled G-factor. To test this assumption, we contrasted second-order and bifactor models (incorporating one work engagement G-factor and the three S-factors). The fit of the second-order model (S1M3 and S2M3) was identical to that of the first-order model. However, fit for the bifactor models (S1M4 and S2M4) was good (CFI and TLI $>$ 0.95, RMSEA $\leq$ 0.08) and it was superior to the first-order models (Sample 1: ΔCFI = + 0.036, ΔTLI = + 0.043, ΔRMSEA = −0.036; Sample 2: ΔCFI = + 0.018; ΔTLI = + 0.021; ΔRMSEA = −0.018). The work engagement G-factor was well-defined in both samples (Sample 1: λ = 0.729 to 0.883; Sample 2: λ = 0.702 to 0.921) as were the vigor (Sample 1: λ = 0.160 to 0.602; Sample 2: λ = 0.142 to 0.513) and absorption (Sample 1: λ = 0.119 to 0.632; Sample 2: λ = 0.215 to 0.484) S-factors. In contrast, the dedication S-factor (Sample 1: λ = 0.187 to 0.399; Sample 2: λ = −0.500 to 0.042) had a comparatively weaker definition.

In the next step, measurement invariance was tested across the two samples (Models MS in **Table 1**) to verify the replicability of the final bifactor-CFA model (see **Table 1**). The configural model with no equality constraints provided a reasonably

TABLE 1 | Goodness-of-fit statistics of the alternative measurement models on the Hungarian version of Utrecht work engagement scale.

Model	χ^2 (df)	CFI	TLI	RMSEA	Comparison	$\Delta\chi^2$ (df)	ΔCFI	ΔTLI	ΔRMSEA	RDR
Sample 1										
S1M1. One-factor CFA	215.595* (27)	0.866	0.822	0.170 [0.149,0.191]	–	–	–	–	–	–
S1M2. Three-factor CFA	102.366* (24)	0.944	0.917	0.116 [0.094,0.140]	S1M1	74.048 (3)*	+ 0.078	+0.095	−0.054	Na
S1M3. Second-order CFA	102.370* (24)	0.944	0.917	0.116 [0.094,0.140]	S1M1	74.048 (3)*	+ 0.078	+0.095	−0.054	Na
S1M4. Bifactor CFA	46.016* (18)	0.980	0.960	0.080 [0.052,0.109]	S1M2	59.795 (6)*	+ 0.036	+0.043	−0.036	Na
Sample 2										
S2M1. One-factor CFA	242.039* (27)	0.905	0.873	0.126 [0.111,0.140]	–	–	–	–	–	–
S2M2. Three-factor CFA	101.819* (24)	0.966	0.948	0.080 [0.064,0.096]	S2M1	111.372 (3)*	+ 0.061	+0.075	−0.046	Na
S2M3. Second-order CFA	102.537* (24)	0.965	0.948	0.080 [0.065,0.097]	S2M1	132.544 (3)*	+ 0.060	+0.075	−0.046	Na
S2M4. Bifactor CFA	53.315* (18)	0.984	0.969	0.062 [0.043,0.082]	S2M2	48.279 (6)*	+ 0.018	+0.021	−0.018	Na
Measurement Invariance Across Gender										
MG1. Configural invariance	84.162* (36)	0.987	0.974	0.060 [0.043,0.077]	–	–	–	–	–	–
MG2. Weak invariance	105.197* (50)	0.985	0.978	0.054 [0.040,0.069]	MG1	20.511 (14)	−0.002	+ 0.004	−0.006	0.025
MG3. Strong invariance	111.108* (55)	0.985	0.980	0.052 [0.038,0.066]	MG2	4.151 (5)	0.000	+ 0.002	−0.002	NPC
MG4. Strict invariance	117.824* (64)	0.985	0.983	0.047 [0.034,0.061]	MG3	8.382 (9)	0.000	+ 0.003	−0.005	NPC
MG5. Latent variance-covariance invariance	124.139* (68)	0.985	0.984	0.047 [0.034,0.060]	MG4	6.337 (4)	0.000	+ 0.001	0.000	0.028
MG6. Latent means invariance	131.724* (72)	0.984	0.984	0.047 [0.034,0.060]	MG5	7.675 (4)	−0.001	0.000	0.000	0.035
Measurement Invariance Across Age										
MA1. Configural invariance	91.675* (36)	0.985	0.969	0.064 [0.048,0.081]	–	–	–	–	–	–
MA2. Weak invariance	110.681* (50)	0.983	0.976	0.057 [0.043,0.071]	MA1	16.046 (14)	−0.002	+ 0.007	−0.007	0.014
MA3. Strong invariance	132.854* (55)	0.978	0.972	0.062 [0.048,0.075]	MA2	27.379 (5)*	−0.005	−0.004	+ 0.005	0.077
MA4. Strict invariance	155.031* (64)	0.975	0.972	0.062 [0.049,0.074]	MA3	22.213 (9)*	−0.003	0.000	0.000	0.044
MA5. Latent variance-covariance invariance	185.608* (68)	0.967	0.965	0.068 [0.056,0.080]	MA4	22.446 (4)*	−0.008	−0.007	+ 0.006	0.079
MA6. Latent means invariance	206.883* (72)	0.963	0.963	0.071 [0.060,0.082]	MA5	24.914 (4)*	−0.004	−0.002	+ 0.003	0.084
Measurement Invariance Across Organizational Levels										
MO1. Configural invariance[a]	116.603* (56)	0.984	0.969	0.066 [0.049,0.083]	–	–	–	–	–	–
MO2. Weak invariance[b]	144.931* (82)	0.983	0.978	0.056 [0.040,0.070]	MO1	26.965 (26)	−0.001	+ 0.009	−0.010	0.007
MO3. Strong invariance	158.536* (92)	0.982	0.979	0.054 [0.039,0.068]	MO2	12.085 (10)	−0.001	+ 0.001	−0.002	0.017
MO4. Strict invariance	184.654* (110)	0.980	0.980	0.052 [0.039,0.065]	MO3	26.692 (18)	−0.002	+ 0.001	−0.002	0.025
MO5. Latent variance-covariance invariance	232.741* (118)	0.969	0.972	0.062 [0.051,0.074]	MO4	43.116 (8)*	−0.011	−0.008	+ 0.010	0.077
MO6. Latent means invariance	269.562* (126)	0.961	0.967	0.068 [0.056,0.079]	MO5	40.437 (8)*	-0.008	−0.005	+ 0.006	0.074
Measurement Invariance Across Samples										
MS1. Configural invariance	154.568* (36)	0.968	0.937	0.094 [0.079,0.109]	–	–	–	–	–	–
MS2. Weak invariance	102.508* (50)	0.986	0.980	0.053 [0.038,0.068]	MS1	52.533 (14)*	+ 0.018	+0.043	−0.041	0.061
MS3. Strong invariance	107.961* (55)	0.986	0.981	0.051 [0.036,0.065]	MS2	3.305 (5)	+ 0.000	+0.001	−0.002	NPC
MS4. Strict invariance	119.706* (64)	0.985	0.983	0.048 [0.035,0.062]	MS3	12.246 (9)	−0.001	+ 0.002	−0.003	0.022
MS5. Latent variance-covariance invariance	129.531* (68)	0.984	0.983	0.049 [0.036,0.062]	MS4	9.566 (4)	−0.001	0.000	+ 0.001	0.043
MS6. Latent means invariance	138.784* (72)	0.982	0.982	0.050 [0.037,0.062]	MS5	9.496 (4)	−0.002	−0.001	+ 0.001	0.028

$p < 0.01$; CFA, confirmatory factor analysis; χ^2, Chi-square; df, degrees of freedom; CFI, comparative fit index; TLI, Tucker-Lewis Index; RMSEA, root-mean-square error of approximation; 90% CI, 90% confidence interval of the RMSEA; $\Delta\chi^2$, Robust (Satorra-Bentler) chi-square difference test (calculated from loglikelihood for greater precision); ΔCFI, change in CFI value compared to the preceding model; ΔTLI, change in the TLI value compared to the preceding model; ΔRMSEA, change in the RMSEA value compared to the pre-ceding model; RDR, root deterioration per restriction index; Na, not applicable; NPC, not possible to calculate due to the fact that the chi-square difference value is smaller than the difference in the degrees of freedom. [a] The residual variance of item 3 was constrained to be higher than zero in all groups to achieve identification. [b] The residual variance of item 3 and the variance of the dedication S-factor were constrained to be higher than zero in group 2 and 3, respectively, to achieve identification.

good model fit based on CFI and TLI (0.968 and 0.937, respectively), but not RMSEA (0.094). Still, the confidence interval of the latter reached the level of acceptability (i.e., 0.080), suggesting that the factor structure is reasonably similar across samples. Next, we put equality constraints on the factor loadings, which led to substantial improvements in model fit (ΔCFI = + 0.018, ΔTLI = + 0.043, ΔRMSEA = −0.041; RDR = 0.061), providing good support for the weak invariance of the bifactor-CFA measurement model. The gradual inclusion of the equality constraints on the additional parameters (i.e., intercepts, uniquenesses, latent variances and covariances, and latent means) showed that (1) CFI, TLI, and RMSEA indicated good fit on all invariance levels; (2) decreases in CFI and TLI were never above 0.010 with the highest being −0.002; (3) increases in RMSEA were never above 0.015 with the highest change being + 0.001; and (4) all RDR values remained below 0.05. Highly similar results were obtained when the bifactor-CFA was contrasted along groups based on gender (Models MG in **Table 1**), age (Models MA in **Table 1**), and organizational level (Models MO in **Table 1**), all of which converged on the same conclusions and thus supporting the latent mean invariance and the replicability of the bifactor-CFA solution across samples, gender, age, and organizational level.

Parameter estimates from the latent mean invariant measurement model (derived from Model MS6) are reported in **Table 2**. These results showed a well-defined and highly reliable work engagement G-factor (λ = 0.712 to 0.905, M = 0.793, ω = 0.961). Once the effect of the G-factor was taken into account, the vigor (λ = 0.144 to 0.576, M = 0.395, ω = 0.655) and absorption (λ = 0.156 to 0.554, M = 0.343, ω = 0.573) S-factors retained a meaningful amount of specificity as opposed to the dedication S-factor (λ = 0.046 to 0.465, M = 0.193, ω = 0.379) which retained a smaller amount of specificity. The present results suggest that the dedication items mostly reflected participants' global levels of work engagement instead of the pure dedication associated with this S-factor over and above the G-factor. When examining a bifactor solution, it is important to keep in mind that not all S-factors should be strongly defined and that S-factors tend to be weaker in bifactor representations because the items are associated with two factors (G- and S-factors) instead of one (S-factor) as in the first-order solution. In a similar vein, it should also be kept in mind that the present model used fully latent variables (instead of manifest scale scores) which are naturally corrected for measurement error and thus the factors should be considered reliable.

Validity Evidence Based on Test-Criterion Relationship

In order to assess the validity evidence of the bifactor-CFA solution based on its test-criterion relationship, Spearman correlations were calculated between the factors. Factors were represented by factor scores (standardized with 0 mean and 1 standard deviation) derived from the latent mean invariant measurement model for work engagement and from preliminary measurement models estimated *a priori*. These preliminary measurement models also allowed us to ascertain that the correlates had adequate validity evidence and reliability (see **Supplementary Appendix 2** in the online supplements for more information).

TABLE 2 | Standardized parameter estimates from the latent mean invariant bifactor-CFA solution for the Hungarian version of Utrecht work engagement scale (Model MS6).

	ENG (λ)	VIG (λ)	DED (λ)	ABS (λ)	δ
Vigor					
Item 1	0.745**	0.576**			0.114
Item 2	0.761**	0.465**			0.205
Item 5	0.748**	0.144**			0.419
ω		0.655			
Dedication					
Item 3	0.905**		0.067*		0.176
Item 4	0.884**		0.465**		0.002
Item 7	0.793**		0.046		0.369
ω			0.379		
Absorption					
Item 6	0.769**			0.156**	0.384
Item 8	0.712**			0.554**	0.186
Item 9	0.824**			0.319**	0.219
ω	0.961			0.573	

*ENG, Work Engagement; VIG, Vigor; DED, Dedication; ABS, Absorption; CFA, Confirmatory factor analysis; λ, Factor loading; δ, Item uniqueness; ω, model-based omega composite reliability; *p < 0.05; **p < 0.01.*

Correlations between factors of work engagement, factors of need fulfillment and turnover intention can be seen in **Table 3**. Global levels of work engagement positively correlated with global levels of need fulfillment (r = 0.561, p < 0.001), as well as with specific levels of autonomy satisfaction (r = 0.440, p < 0.001) and relatedness satisfaction (r = 0.170, p = 0.008), while being negatively related to specific levels of autonomy frustration (r = –0.249, p < 0.001) and turnover intentions (r = −0.646, p < 0.001). Over and above the work engagement G-factor, some of the engagement S-factors also showed additional relations with the correlates, giving support for their added value. More specifically, there was a weak positive correlation between vigor and need fulfillment G-factor (r = 0.178, p = 0.006), between dedication and autonomy satisfaction (r = 0.158, p = 0.014), and between absorption and relatedness frustration S-factors (r = 0.160, p = 0.013). In addition, the dedication S-factor negatively correlated with turnover intention (r = −0.150, p = 0.020).

When taking a look on the correlations involving Sample 2 (see **Table 4**), there was a strong positive correlation (r = 0.713, p < 0.001) between work satisfaction and global levels of work engagement as well as a weak positive correlation between global levels of work engagement and work addiction (r = 0.134, p = 0.003). Once again, the added value of the S-factors is supported by the weak positive correlation between dedication S-factor and work satisfaction (r = 0.131, p = 0.003) and by the weak positive correlation between work addiction and absorption S-factor (r = 0.198, p < 0.001).

DISCUSSION

The aim of our study was to examine the representation of work engagement (as measured by the UWES-9) and to test whether the bifactor structure of work engagement would be a more adequate and improved representation compared to alternative first-order and the second-order solutions. This approach allowed us to bridge seemingly diverging perspectives by simultaneously considering both the global and specific components of work engagement. As an additional aim, the present study also documented the validity evidence of this representation based on its test-criterion relationship with basic psychological need fulfillment at work, turnover intentions, work addiction, and work satisfaction.

The Bifactor Representation of Work Engagement

Our results, in line with Hypothesis 1, supported the superiority of the bifactor representation of work engagement, thus also aligning with findings reported by de Bruin and Henn (2013) as well as Gillet et al. (2018, 2019). In addition, the bifactor representation was well-replicated across the two distinct samples. In this bifactor representation, the G-factor can be seen as a direct reflection of employees' global level of work engagement, while the S-factors are posited to reflect the presence

TABLE 3 | Spearman Bivariate correlations between the variables used in Sample 1 (N = 242).

	1	2	3	4	5	6	7	8	9	10	11
1. Work engagement G-factor	—										
2. Vigor S-factor	0	—									
3. Dedication S-factor	0	0	—								
4. Absorption S-factor	0	0	0	—							
5. Need fulfillment G-factor	0.561**	0.178**	0.052	0.095	—						
6. Autonomy satisfaction S-factor	0.440**	−0.044	0.158*	0.107	0.154*	—					
7. Relatedness satisfaction S-factor	0.170**	0.037	0.065	−0.086	0.067	0.014	—				
8. Competence satisfaction S-factor	−0.049	0.085	−0.006	0.061	0.118	−0.085	−0.042	—			
9. Autonomy frustration S-factor.	−0.249**	−0.114	0.020	0.031	−0.103	−0.009	0.095	0.127*	—		
10. Relatedness frustration S-factor	0.125	0.013	−0.008	0.160*	0.048	0.128*	0.032	0.008	−0.028	—	
11. Competence frustration S-factor	−0.091	0.030	−0.009	−0.067	−0.068	−0.024	0.056	−0.009	−0.031	−0.010	—
12. Turnover intention	−0.646**	−0.095	−0.150*	0.051	−0.569**	−0.415**	−0.219**	0.281**	0.210**	0.035	0.038

*G-factor, global factor from the bifactor model; S-factor, specific factor from the bifactor model; **$p < 0.01$, *$p < 0.05$.*

of employees' vigor, dedication, and absorption over and above, and independently from, their global levels of engagement. These specific dimensions also reflect the extent to which vigor, dedication and absorption deviate from the global levels of engagement. Previous studies using the UWES suggested that researchers should focus on using either the global or the specific components. However, our study shows that the two approaches are not mutually exclusive. Indeed, our study illustrates why it is important to carefully compare alternative measurement models in terms of model fit and standardized parameter estimates. The first-order CFA results demonstrated similar patterns to previous studies (e.g., Wefald et al., 2012; Littman-Ovadia and Balducci, 2013; Zeijen et al., 2018; Kulikowski, 2019) in that model fit was less than optimal across the two samples. Correlations between the three first-order factors were high, suggesting the potential presence of an unmodelled G-factor. By contrast, the fit indices for the bifactor solutions, which does incorporate a work engagement G-factor, were good in both samples.

Inspection of the parameter estimates associated with the bifactor model revealed a well-defined work engagement global factor, with a meaningful amount of specificity being retained in the vigor and absorption S-factors, and a smaller amount of specificity in the dedication S-factor. The weaker representation of the specific factors in the bifactor solutions can be attributed to scale items being associated with a specific and a global factor simultaneously. The small amount of specificity of the items of the dedication factor suggests that these items mostly reflected participants' global sense of work engagement. However, this particular result does not mean that the bifactor model is not optimal or that the dedication S-factor should be discarded. Indeed, as stated by Morin et al. (2016a), it is rare to observe that all S-factors are well-defined in bifactor solutions which typically include at least some well-defined S-factors apart from a strongly defined G-factor. A weaker S-factor shows that a subset of items only serves to reflect global levels of work engagement, and this weaker S-factor simply should be interpreted with caution. While it has been argued that partial bifactor solutions should be pursued in the case of weaker S-factors (de Bruin and Henn, 2013; Fong and Ho, 2015), we argue that the meaningfulness of the G- and S-factors should be tested in relation to theoretically-relevant correlates before removing any S-factors as these investigation might support the added value of the S-factors over and above the G-factor.

TABLE 4 | Spearman Bivariate correlations between variables used in Sample 2 (N = 505).

	1	2	3	4	5
1. Work engagement G-factor	—				
2. Vigor S-factor	0	—			
3. Dedication S-factor	0	0	—		
4. Absorption S-factor	0	0	0	—	
5. Work addiction	0.134**	−0.045	0.071	0.198**	—
6. Work satisfaction	0.713**	0.038	0.131**	0.055	−0.035

*G-factor, global factor from the bifactor model; S-factor, specific factor from the bifactor model; **$p < 0.01$, *$p < 0.05$.*

Test-Criterion Relationship Based Validity of the Bifactor Representation

Global Levels of Work Engagement

Our findings with respect to the validity evidence based on test-criterion relationship of the UWES-9 do not only highlight the importance of the global levels of work engagement, but also the added value of the specific levels of vigor, dedication, and absorption. More specifically, global levels of work engagement demonstrated a positive association with global levels of need fulfillment (e.g., Trépanier et al., 2015), providing support for Hypothesis 2a. These results suggest that experiencing high global levels of work engagement tend to be positively associated with experiencing high global levels of need fulfillment at work. When employees' basic psychological needs are fulfilled at their workplace, they are more likely to experience growth, wellness, and optimal functioning (Ryan and Deci, 2017) which can translate into functioning more effectively at work and experiencing higher levels of positive work-related states such as work engagement. Both cross-sectional (e.g., Trépanier et al., 2013) and longitudinal (e.g., Trépanier et al., 2015) studies have reported need fulfillment to be an important predictor of work engagement. Over and above the global levels of need fulfillment, global work engagement was also associated with high specific levels of autonomy satisfaction and relatedness satisfaction. Experiencing high levels of engagement at work thus might not only be related to global levels of need fulfillment, but also specific levels of autonomy and relatedness satisfaction, suggesting that engaged employees tend to experience high levels of autonomy and relatedness satisfaction over and above the global levels of work engagement.

In addition to these findings, global levels of work engagement were negatively related to specific levels of autonomy frustration and turnover intentions which is in line with previous empirical studies (e.g., Trépanier et al., 2013; Shuck et al., 2015; Wang et al., 2018) that relied on first-order representations of work engagement. These results highlight that the frustrated need for autonomy (i.e., feelings of pressure and conflict at work) might have a negative effect on employees' work engagement. Such need frustrated experiences might be attributed to need thwarting work conditions (Vansteenkiste and Ryan, 2013) in which employees are expected to behave in a certain way and have less control over what and how they need to do in their work, thus they cannot act in a volitional manner. Prior studies have already provided support for this explanation (e.g., Deci et al., 2001; Van den Berghe et al., 2016; see Deci et al., 2017 for an overview). Finally, the negative association between global levels of work engagement and turnover intentions is consistent with Hypothesis 2d, and is also in line with results of prior studies (e.g., Mills et al., 2012; Wefald et al., 2012; Lovakov et al., 2017). Thus, when employees do not feel engaged in their work, they might be more likely to detach themselves from the organization and potentially leave it.

Global levels of work engagement showed a positive and weak association with work addiction which is in line with Hypothesis 2b. This result is consistent with the results reported in most previous studies (e.g., van Beek et al., 2012; Clark et al., 2014; Littman-Ovadia et al., 2014; Di Stefano and Gaudiino, 2018). Even though this association was positive, its magnitude remained small which further supports the idea that global levels of work engagement and work addiction reflect two distinct construct that are relatively independent from one another. Additionally, global work engagement also showed a positive association with work satisfaction (i.e., engaged employees were more likely to be satisfied with their work), thus providing empirical support for Hypothesis 2c and further establishing the validity evidence of this representation. This result also corroborates findings reported in cross-sectional (e.g., Klassen et al., 2012; Littman-Ovadia and Balducci, 2013; Schaufeli et al., 2019) and meta-analytic (Christian et al., 2011) studies. While these constructs share conceptual similarities (i.e., the value of pleasure at work), they differ from one another in two main characteristics. First, they differ in their level of activation: work engagement is characterized by high level of energy as opposed to the low energy level in work satisfaction (Bakker and Oerlemans, 2011). Second, they have different sources of origin: work engagement is an affective outcome of work experience, while work satisfaction is an attitude toward work, which is based on the evaluation of conditions and characteristics of work (Christian et al., 2011; Salanova et al., 2014; Schaufeli et al., 2019).

Specific Levels of Work Engagement

Finally, our results also answered our Research Question by showing that some of the specific components of work engagement appeared to have an added value by demonstrating meaningful associations with the correlates. First, specific levels of *vigor* were positively related to global levels of need fulfillment at work. This result suggests that employees experiencing fulfilled basic psychological needs at work might have more work-related energy and mental resilience beyond the global levels of work engagement. Second, specific levels of *dedication* were positively related to specific levels of autonomy satisfaction and work satisfaction, but negatively to turnover intentions. These relationships suggest that by perceiving work as significant, inspiring, and meaningful (over and above the global levels of work engagement) might stem from having ample amount of choice and self-initiation at work, and it could also be protective of negative outcomes (i.e., lower levels of turnover intentions) and conductive of positive outcomes (i.e., higher levels of work satisfaction). Third, specific levels of *absorption* were positively related to specific levels of relatedness frustration. That is, when employees experience social rejection and exclusion at work by coworkers or supervisors, they might be more likely to become immersed in and obsessed with their work. This finding is consistent with prior studies (e.g., Tóth-Király et al., 2019b) documenting the potentially negative effects associated with relatedness frustration. This result is less surprising when we take into account that being isolated and lonely have already been related to decreased wellbeing and other maladaptive outcomes (e.g., Mellor et al., 2008; Kim et al., 2009). Becoming over-engaged with work (i.e., having high specific levels of absorption) might become a compensatory behavior for employees in order to counter the experiences of need frustration (Vansteenkiste and Ryan, 2013; Tóth-Király et al., 2019a; Bõthe et al., 2020). Specific levels of absorption, similar to prior findings relying on first-order factors (Líbano et al., 2012; Shimazu et al., 2015; Clark et al., 2016; Di Stefano and Gaudiino, 2018), were also positively related to work addiction. This positive relationship highlights the shared nature of absorption and work addiction as both are characterized with an immersion into the work-related activities from which it is difficult to disengage.

Overall, the present two-study investigation shows that work engagement might be best represented by a bifactor solution incorporating an overarching work engagement construct underlying all responses, as well as the three components of vigor, dedication, and absorption. Failure to taking into account this representation might lead to erroneous conclusions due to the high associations (i.e., multicollinearity) between the three work engagement components that appear to reflect a more global construct, while also masking the potential complementary effect of the S-factors beyond the G-factor. For these reasons, we would advise researchers to, in their pursuits, consider relying on fully latent measurement models that do not only make it possible to estimate the most optimal bifactor representation of work engagement, but they are also naturally corrected for measurement error. When the sample size is modest, similar to our approach, researchers could rely on factor scores derived from the bifactor measurement model in order to preserve its underlying nature (Morin et al., 2016b). In practical terms, this approach allows researchers to obtain a more precise and direct estimate of global work engagement as bifactor models weight items based on their contribution to the factor itself. To make this process seamless, as suggested by Perreira et al. (2018), automated scoring procedures could be developed, or the Mplus statistical package could be used, which has the advantage of providing standardized measurements interpretable as a function of the sample mean and standard deviation.

Strengths and Limitations

The current study provides an alternative solution to the debate about the appropriate representation of work engagement. While the bifactor-CFA solution was the most optimal in comparison to other alternative models, it also allows us to investigate the nature of work engagement both on the global and the specific level. An additional strength is the replication of our findings using an independent second sample. The current study also documented the validity evidence of bifactor-CFA representation of work engagement based on its test-criterion relationship which was an important step toward its better understanding.

Nevertheless, there are some limitations that should be considered. Both studies were cross-sectional, implying that causality cannot be inferred from our results. Given that self-reported measures were used, responses might have been biased (e.g., social desirability). Future longitudinal research would be necessary to give a deeper understanding of how the representation of work engagement changes over time. Alternatively, it would be important to complement the present results with longitudinal or intervention studies with enhanced methodological quality (Chacón-Moscoso et al., 2016). The generalization of the current results requires their replication on a larger, international sample. Moreover, the sample consisted of mostly female and white-collar/manager participants; therefore, the sample is not representative of the Hungarian population. Future studies should verify the findings on a representative and more diverse sample (e.g., a sample including health care professionals and respondents from other occupations). Further studies focusing on examining the bifactor-CFA representation should be conducted in other countries and languages as well. Future studies would also do well in re-assessing the validity evidence based on test-criterion relationship using different work-related measures. It would also be interesting to examine the representation of engagement towards other activities such as studies (Dierendonck et al., 2021) or job (Gillet et al., 2020). Given that the dedication S-factor had relatively low reliability, future studies should investigate whether this is a re-occurring phenomenon or whether it is a sample-specific result.

CONCLUSION

Taken together, the present research demonstrated the superiority of the bifactor solution, which not only provides an improved representation of work engagement, but also a clearer picture of the different relations of the global and specific components of work engagement to other, relevant work-related constructs. The importance of the specific factors of work engagement were illustrated by their diverse relations with these correlates. The results supported the discriminant validity evidence of vigor, dedication, and absorption as specific factors. The current findings support the simultaneous application of the global work engagement construct and its specific components.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in OSF at https://osf.io/upn9c/?view_only=2c0e7f703d2942438a630e7807f877b3.

ETHICS STATEMENT

The current study was reviewed and approved by the Institutional Review Board of Eötvös Loránd University Faculty of Education and Psychology. The patients/participants provided their written informed consent to participate in this study. Informed consent was obtained from all participants included in the study.

AUTHOR CONTRIBUTIONS

JS and IT-K contributed to the study design, literature review, data gathering, manuscript writing, and to the data analyses and interpretation. BB, GO, and TN contributed to the literature review, and to the manuscript writing. All authors commented on the draft and contributed to the final version, approved the publication of the manuscript, and agreed to be accountable for all aspects of the work.

FUNDING

JS was supported by the Hungarian National Research, Development and Innovation Office (Grant No.: FK124225). TN was supported by the Hungarian National Research, Development and Innovation Office (Grant Nos.: FK124225, PD131954). IT-K was supported in the preparation of this manuscript by a Horizon Postdoctoral Fellowship from Concordia University and by funding from the Social Sciences and Humanities Research Council of Canada (435-2018-0368). BB was supported by a postdoctoral fellowship from the SCOUP Team – Sexuality and Couples – Fonds de recherche du Québec, Société et Culture. GO was supported by the Young Researcher STARS grant from Conseil Régional Hauts de France.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.615581/full#supplementary-material

REFERENCES

Adie, J. W., Duda, J. L., and Ntoumanis, N. (2008). Autonomy support, basic need satisfaction and the optimal functioning of adult male and female sport participants: a test of basic needs theory. *Motiv. Emot.* 32, 189–199. doi: 10.1007/s11031-008-9095-z

Alessandri, G., Borgogni, L., Schaufeli, W. B., Caprara, G. V., and Consiglio, C. (2015). From positive orientation to job performance: the role of work engagement and self-efficacy beliefs. *J. Happiness Stud.* 16, 767–788. doi: 10.1007/s10902-014-9533-4

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, USA: American Educational Research Association.

Andreassen, C. S., Griffiths, M. D., Hetland, J., and Pallesen, S. (2012). Development of a work addiction scale: development of a work addiction scale. *Scand. J. Psychol.* 53, 265–272. doi: 10.1111/j.1467-9450.2012.00947.x

Andreassen, C. S., Griffiths, M. D., Sinha, R., Hetland, J., and Pallesen, S. (2016). The relationships between workaholism and symptoms of psychiatric disorders: a large-scale cross-sectional study. *PLoS One* 11:e0152978. doi: 10.1371/journal.pone.0152978

Andreassen, C. S., Schaufeli, W. B., and Pallesen, S. (2018). Myths about "The myths about work addiction": commentary on: ten myths about work addiction (Griffiths et al., 2018). *J. Behav. Addict.* 7, 858–862. doi: 10.1556/2006.7.2018.126

Bakker, A. B., and Oerlemans, W. G. M. (2011). "Subjective well-being in organizations," in *Handbook of Positive Organizational Scholarship*, eds K. Cameron and G. Spreitzer (Oxford: Oxford University Press).

Balducci, C., Fraccaroli, F., and Schaufeli, W. B. (2010). Psychometric properties of the Italian version of the Utrecht Work Engagement Scale (UWES-9): a cross-cultural analysis. *Eur. J. Psychol. Assess.* 26, 143–149. doi: 10.1027/1015-5759/a000020

Beaton, D. E., Bombardier, C., Guillemin, F., and Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 25, 3186–3191. doi: 10.1097/00007632-200012150-00014

Bõthe, B., Tóth-Király, I., Potenza, M. N., Orosz, G., and Demetrovics, Z. (2020). High-frequency pornography use may not always be problematic. *J. Sex. Med.* 17, 793–811. doi: 10.1016/j.jsxm.2020.01.007

Breevaart, K., Bakker, A. B., Demerouti, E., and Hetland, J. (2012). The measurement of state work engagement: a multilevel factor analytic study. *Eur. J. Psychol. Assess.* 28, 305–312. doi: 10.1027/1015-5759/a000111

Browne, M. W., and Du Toit, S. H. C. (1992). Automated fitting of nonstandard models. *Multivar. Behav. Res.* 27, 269–300. doi: 10.1207/s15327906mbr2702_13

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

Chaudhary, R., Rangnekar, S., and Barua, M. K. (2012). Psychometric evaluation of Utrecht work engagement scale in an Indian sample. *Asia-Pac. J. Manag. Res. Innov.* 8, 343–350. doi: 10.1177/2319510X1200800314

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., et al. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motiv. Emot.* 39, 216–236. doi: 10.1007/s11031-014-9450-1

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Modeling* 14, 464–504. doi: 10.1080/10705510701301834

Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Modeling* 9, 233–255.

Christian, M. S., Garza, A. S., and Slaughter, J. E. (2011). Work engagement: a quantitative review and test of its relations with task and contextual performance. *Pers. Psychol.* 64, 89–136. doi: 10.1111/j.1744-6570.2010.01203.x

Clark, M. A., Michel, J. S., Stevens, G. W., Howell, J. W., and Scruggs, R. S. (2014). Workaholism, work engagement and work-home outcomes: exploring the mediating role of positive and negative emotions: workaholism, work engagement and work-home. *Stress Health* 30, 287–300. doi: 10.1002/smi.2511

Clark, M. A., Michel, J. S., Zhdanova, L., Pui, S. Y., and Baltes, B. B. (2016). All work and no play? a meta-analytic examination of the correlates and outcomes of workaholism. *J. Manag.* 42, 1836–1873. doi: 10.1177/0149206314522301

Cox, A., and Williams, L. (2008). The roles of perceived teacher support, motivational climate, and psychological need satisfaction in students' physical education motivation. *J. Sport Exerc. Psychol.* 30, 222–239. doi: 10.1123/jsep.30.2.222

de Bruin, G. P., and Henn, C. M. (2013). Dimensionality of the 9-Item Utrecht Work Engagement Scale (UWES-9). *Psychol. Rep.* 112, 788–799. doi: 10.2466/01.03.PR0.112.3.788-799

Deci, E. L., Olafsen, A. H., and Ryan, R. M. (2017). Self-determination theory in work organizations: the state of a science. *Annu. Rev. Organ. Psychol. Organ. Behav.* 4, 19–43. doi: 10.1146/annurev-orgpsych-032516-113108

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., Ryan, R. M., Gagné, M., Leone, D. R., Usunov, J., and Kornazheva, B. P. (2001). Need satisfaction, motivation, and well-being in the work organizations of a former eastern bloc country: a cross-cultural study of self-determination. *Pers. Soc. Psychol. Bull.* 27, 930–942. doi: 10.1177/0146167201278002

Di Stefano, G., and Gaudiino, M. (2018). Differential effects of workaholism and work engagement on the interference between life and work domains. *Eur. J. Psychol.* 14, 863–879. doi: 10.5964/ejop.v14i4.1626

Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The satisfaction with life scale. *J. Pers. Assess.* 49, 71–75. doi: 10.1207/s15327752jpa4901_13

Dierendonck, C., Tóth-Király, I., Morin, A. J. S., Kerger, S., Milmesiter, P., and Poncelet, D. (2021). Testing associations between global and specific levels of student academic motivation and engagement in the classroom. *J. Exp. Educ.* 1–24. doi: 10.1080/00220973.2021.1913979

Falco, A., Girardi, D., Kravina, L., Trifiletti, E., Bartolucci, G. B., Capozza, D., et al. (2013). The mediating role of psychophysic strain in the relationship between workaholism, job performance, and sickness absence: a longitudinal study. *J. Occup. Environ. Med.* 55, 1255–1261. doi: 10.1097/JOM.0000000000000007

Faragher, E. B., Cass, M., and Cooper, C. L. (2005). The relationship between job satisfaction and health: a meta-analysis. *Occup. Environ. Med.* 62, 105–112. doi: 10.1136/oem.2002.006734

Fernet, C., Trépanier, S.-G., Demers, M., and Austin, S. (2017). Motivational pathways of occupational and organizational turnover intention among newly registered nurses in Canada. *Nurs. Outlook* 65, 444–454. doi: 10.1016/j.outlook.2017.05.008

Fong, T. C., and Ho, R. T. (2015). Dimensionality of the 9-item utrecht work engagement scale revisited: a bayesian structural equation modeling approach. *J. Occup. Health*, 15:0057. doi: 10.1539/joh.15-0057-OA

Fong, T. C., and Ng, S. (2012). Measuring engagement at work: validation of the chinese version of the utrecht work engagement scale. *Int. J. Behav. Med.* 19, 391–397. doi: 10.1007/s12529-011-9173-6

Fouquereau, E., and Rioux, L. (2002). L laboration de l'l chelle de satisfaction de vie professionnelle (I SVP) en langue fran aise une d6marche exploratoire. *Can. J. Behav. Sci.* 34, 210–215.

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: that is why the bifactor model tends to fit better. *Intelligence* 55, 57–68. doi: 10.1016/j.intell.2016.01.006

Gillet, N., Caesens, G., Morin, A. J. S., and Stinglhamber, F. (2019). Complementary variable- and person-centred approaches to the dimensionality of work engagement: a longitudinal investigation. *Eur. J. Work Organ. Psychol.* 28, 239–258. doi: 10.1080/1359432X.2019.1575364

Gillet, N., Fouquereau, E., Huyghebaert, T., and Colombat, P. (2015). The effects of job demands and organizational resources through psychological need satisfaction and thwarting. *Span. J. Psychol.* 18:E28. doi: 10.1017/sjp.2015.30

Gillet, N., Morin, A. J. S., Jeoffrion, C., and Fouquereau, E. (2020). A person-centered perspective on the combined effects of global and specific levels of job engagement. *Group Organ. Manag.* 45, 556–594.

Gillet, N., Morin, A. J. S., Sandrin, E., and Houle, S. A. (2018). Investigating the combined effects of workaholism and work engagement: a substantive-methodological synergy of variable-centered and person-centered methodologies. *J. Vocat. Behav.* 109, 54–77. doi: 10.1016/j.jvb.2018.09.006

Goodboy, A. K., Martin, M. M., and Bolkan, S. (2017). Workplace bullying and work engagement: a self-determination model. *J. Interpers. Violence* 35, 4686–4708. doi: 10.1177/0886260517717492

Gorgievski, M. J., Bakker, A. B., and Schaufeli, W. B. (2010). Work engagement and workaholism: comparing the self-employed and salaried employees. *J. Posit. Psychol.* 5, 83–96. doi: 10.1080/17439760903509606

Griffiths, M. D. (2005). Workaholism is still a useful construct. *Addict. Res. Theory* 13:100.

Hallberg, U. E., and Schaufeli, W. B. (2006). "Same Same" but Different?: can work engagement be discriminated from job involvement and organizational commitment? *Eur. Psychol.* 11, 119–127. doi: 10.1027/1016-9040.11.2.119

Hardre, P. L., and Reeve, J. (2003). A motivational model of rural students' intentions to persist in, versus drop out of, high school. *J. Educ. Psychol.* 95, 347–356. doi: 10.1037/0022-0663.95.2.347

Ho Kim, W., Park, J. G., and Kwon, B. (2017). Work engagement in South Korea: validation of the Korean version 9-item Utrecht work engagement scale. *Psychol. Rep.* 120, 561–578. doi: 10.1177/0033294117697085

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

Kenny, D. A., Kaniskan, B., and McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociol. Methods Res.* 44, 486–507. doi: 10.1177/0049124114543236

Kim, J., LaRose, R., and Peng, W. (2009). Loneliness as the cause and the effect of problematic internet use: the relationship between internet use and psychological well-being. *CyberPsychol. Behav.* 12, 451–455. doi: 10.1089/cpb.2008.0327

Klassen, R. M., Aldhafri, S., Mansfield, C. F., Purwanto, E., Siu, A. F. Y., Wong, M. W., et al. (2012). Teachers' engagement at work: an international validation study. *J. Exp. Educ.* 80, 317–337. doi: 10.1080/00220973.2012.678409

Kulikowski, K. (2019). One, two or three dimensions of work engagement? Testing the factorial validity of the Utrecht work engagement scale on a sample of polish employees. *Int. J. Occup. Saf. Ergon.* 25, 241–249. doi: 10.1080/10803548.2017.1371958

Kun, B., Urbán, R., Bõthe, B., Griffiths, M. D., Demetrovics, Z., and Kökönyei, G. (2020). Maladaptive rumination mediates the relationship between self-esteem, perfectionism, and work addiction: a largescale survey study. *Int. J. Environ. Res. Public Health* 17:7332. doi: 10.3390/ijerph17197332

Lathabhavan, R., Balasubramanian, S. A., and Natarajan, T. (2017). A psychometric analysis of the Utrecht work engagement scale in Indian banking sector. *Ind. Commer. Train.* 49, 296–302. doi: 10.1108/ICT-04-2017-0031

Líbano, M. D., Llorens, S., Salanova, M., and Schaufeli, W. B. (2012). About the dark and bright sides of self-efficacy: workaholism and work engagement. *Span. J. Psychol.* 15, 688–701. doi: 10.5209/rev_SJOP.2012.v15.n2.38883

Littman-Ovadia, H., and Balducci, C. (2013). Psychometric properties of the Hebrew version of the Utrecht Work Engagement Scale (UWES-9). *Eur. J. Psychol. Assess.* 29, 58–63. doi: 10.1027/1015-5759/a000121

Littman-Ovadia, H., Balducci, C., and Ben-Moshe, T. (2014). Psychometric properties of the Hebrew Version of the Dutch Work Addiction Scale (DUWAS-10). *J. Psychol.* 148, 327–346. doi: 10.1080/00223980.2013.801334

Lovakov, A. V., Agadullina, E. R., and Schaufeli, W. B. (2017). Psychometric properties of the Russian Version of the utrecht Work engagement scale (UWES-9). *Psychol. Russ.* 10, 145–162. doi: 10.11621/pir.2017.0111

Marsh, H. W. (2007). "Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology," in *Handbook of Sport Psychology*, 3rd Edn, eds G. Tenenbaum and R. C. Eklund (Hoboken, NJ: Wiley), 774–798.

Marsh, H. W., Hau, K.-T., and Grayson, D. (2005). "Goodness of fit in structural equation models," in *Multivariate Applications Book Series. Contemporary Psychometrics: A Festschrift for Roderick P. McDonald*, eds J. J. McArdle and A. Maydeu-Olivares (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 275–340.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., 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

Martos, T., Sallay, V., Désfalvi, J., Szabó, T., and Ittzés, A. (2014). Psychometric characteristics of the Hungarian version of the Satisfaction with Life Scale (SWLS-H). *Mentálhigiéné és Pszichoszomatika* 15, 289–303. doi: 10.1556/Mental.15.2014.3.9

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *Br. J. Math. Stat. Psychol.* 23, 1–21. doi: 10.1111/j.2044-8317.1970.tb00432.x

Mellor, D., Stokes, M., Firth, L., Hayashi, Y., and Cummins, R. (2008). Need for belonging, relationship satisfaction, loneliness, and life satisfaction. *Pers. Individ. Differ.* 45, 213–218. doi: 10.1016/j.paid.2008.03.020

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58, 525–543. doi: 10.1007/BF02294825

Mills, M. J., Culbertson, S. S., and Fullagar, C. J. (2012). Conceptualizing and measuring engagement: an analysis of the Utrecht work engagement scale. *J. Happiness Stud.* 13, 519–545. doi: 10.1007/s10902-011-9277-3

Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. Abingdon: Routledge.

Moreira-Fontán, E., García-Señorán, M., Conde-Rodríguez, Á., and González, A. (2019). Teachers' ICT-related self-efficacy, job resources, and positive emotions: their structural relations with autonomous motivation and work engagement. *Comput. Educ.* 134, 63–77. doi: 10.1016/j.compedu.2019.02.007

Morin, A. J. S., Arens, A. K., and Marsh, H. W. (2016a). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Struct. Equ. Modeling* 23, 116–139. doi: 10.1080/10705511.2014.961800

Morin, A. J. S., Boudrias, J.-S., Marsh, H. W., Madore, I., and Desrumaux, P. (2016b). Further reflections on disentangling shape and level effects in person-centered analyses: an illustration exploring the dimensionality of psychological health. *Struct. Equ. Modeling* 23, 438–454. doi: 10.1080/10705511.2015.1116077

Morin, A. J. S., 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*, 4th Edn, eds G. Tenenbaum and R. C. Eklund (Hoboken, NJ: Wiley).

Muthén, L. K., and Muthén, B. O. (1998). *Mplus User's Guide*, Eighth Edn. Los Angeles, CA: Muthén & Muthén.

Nerstad, C. G. L., Richardsen, A. M., and Martinussen, M. (2009). Factorial validity of the Utrecht Work Engagement Scale (UWES) across occupational groups in Norway: factorial validity of the UWES. *Scand. J. Psychol.* 51, 326–333. doi: 10.1111/j.1467-9450.2009.00770.x

Orosz, G., Dombi, E., Andreassen, C. S., Griffiths, M. D., and Demetrovics, Z. (2016). Analyzing models of work addiction: single factor and bi-factor models of the bergen work addiction scale. *Int. J. Ment. Health Addict.* 14, 662–671. doi: 10.1007/s11469-015-9613-7

Panthee, B., Shimazu, A., and Kawakami, N. (2014). Validation of Nepalese version of Utrecht work engagement scale. *J. Occup. Health* 56, 421–429. doi: 10.1539/joh.14-0041-OA

Pekrun, R., Murayama, K., Marsh, H. W., Goetz, T., and Frenzel, A. C. (2019). Happy fish in little ponds: testing a reference group model of achievement and emotion. *J. Pers. Soc. Psychol.* 117:166. doi: 10.1037/pspp0000230

Perreira, T. A., Morin, A. J. S., Hebert, M., Gillet, N., Houle, S. A., and Berta, W. (2018). The short form of the Workplace Affective Commitment Multidimensional Questionnaire (WACMQ-S): a bifactor-ESEM approach among healthcare professionals. *J. Vocat. Behav.* 106, 62–83. doi: 10.1016/j.jvb.2017.12.004

Petroviæ, I. B., Vukeliæ, M., and Èizmiæ, S. (2017). Work Engagement in Serbia: psychometric properties of the Serbian Version of the Utrecht Work Engagement Scale (UWES). *Front. Psychol.* 8:1799. doi: 10.3389/fpsyg.2017.01799

Porter, G. (1996). Organizational impact of workaholism: suggestions for researching the negative outcomes of excessive work. *J. Occup. Health Psychol.* 1, 70–84. doi: 10.1037/1076-8998.1.1.70

Raykov, T., and Penev, S. (1998). Nested structural equation models: noncentrality and power of restriction test. *Struct. Equ. Modeling* 5, 229–246. doi: 10.1080/10705519809540103

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivar. Behav. Res.* 47, 667–696. doi: 10.1080/00273171.2012.715555

Ryan, R. M., and Deci, E. L. (2001). On happiness and human potentials: a review of research on hedonic and eudaimonic well-being. *Annu. Rev. Psychol.* 52, 141–166. doi: 10.1146/annurev.psych.52.1.141

Ryan, R. M., and Deci, E. L. (2017). *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. New York, NY: Guilford Press.

Salanova, M., Del Líbano, M., Llorens, S., and Schaufeli, W. B. (2014). Engaged, workaholic, burned-out or just 9-to-5? toward a typology of employee well-being: employee well-being and work investment. *Stress Health* 30, 71–81. doi: 10.1002/smi.2499

Schaufeli, W. B. (2018). Work engagement in Europe: relations with national economy, governance, and culture. *Organ. Dyn.* 47, 99–106. doi: 10.1016/j.orgdyn.2018.01.003

Schaufeli, W. B., Bakker, A. B., and Salanova, M. (2006). The measurement of work engagement with a short questionnaire: a cross-national study. *Educ. Psychol. Meas.* 66, 701–716. doi: 10.1177/0013164405282471

Schaufeli, W. B., Bakker, A. B., van der Heijden, F. M. M. A., and Prins, J. T. (2009). Workaholism, burnout and well-being among junior doctors: the mediating role of role conflict. *Work Stress* 23, 155–172. doi: 10.1080/02678370902834021

Schaufeli, W. B., Salanova, M., González-Romá, V., and Bakker, A. B. (2002). The measurement of engagement and burnout: a two sample confirmatory factor analytic approach. *Journal Happiness Stud.* 3, 71–92.

Schaufeli, W. B., Shimazu, A., Hakanen, J., Salanova, M., and De Witte, H. (2019). An ultra-short measure for work engagement: the UWES-3 validation across five countries. *Eur. J. Psychol. Assess.* 35, 577–591. doi: 10.1027/1015-5759/a000430

Schaufeli, W. B., Taris, T. W., and van Rhenen, W. (2008). Workaholism, burnout, and work engagement: three of a kind or three different kinds of employee well-being? *Appl. Psychol.* 57, 173–203. doi: 10.1111/j.1464-0597.2007.00285.x

Seppälä, P., Mauno, S., Feldt, T., Hakanen, J., Kinnunen, U., Tolvanen, A., et al. (2009). The construct validity of the Utrecht work engagement scale: multisample and longitudinal evidence. *J. Happiness Stud.* 10, 459–481. doi: 10.1007/s10902-008-9100-y

Shimazu, A., Schaufeli, W. B., Kamiyama, K., and Kawakami, N. (2015). Workaholism vs. Work Engagement: the two different predictors of future well-being and performance. *Int. J. Behav. Med.* 22, 18–23. doi: 10.1007/s12529-014-9410-x

Shuck, B., Zigarmi, D., and Owen, J. (2015). Psychological needs, engagement, and work intentions: a Bayesian multi-measurement mediation approach and implications for HRD. *Eur. J. Train. Dev.* 39, 2–21. doi: 10.1108/EJTD-08-2014-0061

Simbula, S., Guglielmi, D., Schaufeli, W. B., and Depolo, M. (2013). An Italian validation of the Utrecht work engagement scale: characterization of engaged groups in a sample of schoolteachers. *Appl. Psychol. Bull.* 61, 43–54.

Sinval, J., Pasian, S., Queirós, C., and Marôco, J. (2018). Brazil-Portugal transcultural adaptation of the UWES-9: internal consistency, dimensionality, and measurement invariance. *Front. Psychol.* 9:353. doi: 10.3389/fpsyg.2018.00353

Tóth-Király, I., Bõthe, B., Márki, A. N., Rigó, A., and Orosz, G. (2019a). Two sides of the same coin: the differentiating role of need satisfaction and frustration in passion for screen-based activities. *Eur. J. Soc. Psychol.* 49, 1190–1205. doi: 10.1002/ejsp.2588

Tóth-Király, I., Bõthe, B., Orosz, G., and Rigó, A. (2019b). A new look on the representation and criterion validity of need fulfillment: application of the bifactor exploratory structural equation modeling framework. *J. Happiness Stud.* 20, 1609–1626. doi: 10.1007/s10902-018-0015-y

Tóth-Király, I., Gajdos, P., Román, N., Vass, N., and Rigó, A. (2019c). The associations between orthorexia nervosa and the sociocultural attitudes: the mediating role of basic psychological needs and health anxiety. *Eat. Weight Disord.* 26, 125–134. doi: 10.1007/s40519-019-00826-1

Tóth-Király, I., Morin, A. J., and Salmela-Aro, K. (2021). A longitudinal perspective on the associations between work engagement and workaholism. *Work Stress* 35, 27–56. doi: 10.1080/02678373.2020.1801888

Tóth-Király, I., Morin, A. J. S., Bõthe, B., Orosz, G., and Rigó, A. (2018). Investigating the multidimensionality of need fulfillment: a bifactor exploratory structural equation modeling representation. *Struct. Equ. Modeling* 25, 267–286. doi: 10.1080/10705511.2017.1374867

Tóth-Király, I., Morin, A. J., Bõthe, B., Rigó, A., and Orosz, G. (2020). Toward an improved understanding of work motivation profiles. *Appl. Psychol.* 70, 986–1017. doi: 10.1111/apps.12256

Trépanier, S.-G., Fernet, C., and Austin, S. (2013). Workplace bullying and psychological health at work: the mediating role of satisfaction of needs for autonomy, competence and relatedness. *Work Stress* 27, 123–140. doi: 10.1080/02678373.2013.782158

Trépanier, S.-G., Fernet, C., and Austin, S. (2015). A longitudinal investigation of workplace bullying, basic need satisfaction, and employee functioning. *J. Occup. Health Psychol.* 20, 105–116. doi: 10.1037/a0037726

Urbán, R., Kun, B., Mózes, T., Soltész, P., Paksi, B., Farkas, J., et al. (2019). A four-factor model of work addiction: the development of the work addiction risk test revised. *Eur. Addict. Res.* 25, 145–160. doi: 10.1159/000499672

Vallerand, R. J., Fortier, M. S., and Guay, F. (1997). Self-determination and persistence in a real-life setting: toward a motivational model of high school dropout. *J. Pers. Soc. Psychol.* 72, 1161–1176. doi: 10.1037//0022-3514.72.5.1161

Vallières, F., McAuliffe, E., Hyland, P., Galligan, M., and Ghee, A. (2017). Measuring work engagement among community health workers in Sierra Leone: validating the Utrecht Work Engagement Scale. *Rev. Psicol. Trab. y de Las Organ.* 33, 41–46. doi: 10.1016/j.rpto.2016.12.001

van Beek, I., Hu, Q., Schaufeli, W. B., Taris, T. W., and Schreurs, B. H. J. (2012). For fun, love, or money: what drives workaholic, engaged, and burned-out employees at work? *Appl. Psychol.* 61, 30–55. doi: 10.1111/j.1464-0597.2011.00454.x

Van den Berghe, L., Cardon, G., Tallir, I., Kirk, D., and Haerens, L. (2016). Dynamics of need-supportive and need-thwarting teaching behavior: the bidirectional relationship with student engagement and disengagement in the beginning of a lesson. *Phys. Educ. Sport Pedagogy* 21, 653–670. doi: 10.1080/17408989.2015.1115008

Van den Broeck, A., Ferris, D. L., Chang, C.-H., and Rosen, C. C. (2016). A review of self-determination theory's basic psychological needs at work. *J. Manag.* 42, 1195–1229. doi: 10.1177/0149206316632058

Vansteenkiste, M., and Ryan, R. M. (2013). On psychological growth and vulnerability: basic psychological need satisfaction and need frustration as a unifying principle. *J. Psychother. Integr.* 23, 263–280. doi: 10.1037/a0032359

Vazquez, A. C. S., Magnan, E., dos, S., Pacico, J. C., Hutz, C. S., and Schaufeli, W. B. (2015). Adaptation and validation of the Brazilian version of the Utrecht work engagement scale. *Psico-USF* 20, 207–217. doi: 10.1590/1413-82712015200202

Villotti, P., Balducci, C., Zaniboni, S., Corbière, M., and Fraccaroli, F. (2014). An analysis of work engagement among workers with mental disorders recently integrated to work. *J. Career Assess.* 22, 18–27. doi: 10.1177/1069072713487500

Wang, Z., Chen, L., Duan, Y., and Du, J. (2018). Supervisory mentoring and newcomers' work engagement: the mediating role of basic psychological need satisfaction. *Soc. Behav. Pers.* 46, 1745–1760. doi: 10.2224/sbp.7609

Wefald, A. J., Mills, M. J., Smith, M. R., and Downey, R. G. (2012). A comparison of three job engagement measures: examining their factorial and criterion-related validity: comparison of engagement measures. *Appl. Psychol.* 4, 67–90. doi: 10.1111/j.1758-0854.2011.01059.x

Yusoff, R. B., Ali, A., Khan, A., and Bakar, S. A. (2013). Psychometric evaluation of Utrecht work engagement scale among academic staff in universities of Pakistan. *World Appl. Sci. J.* 28, 1555–1560. doi: 10.5829/idosi.wasj.2013.28.11.2014

Zecca, G., Györkös, C., Becker, J., Massoudi, K., de Bruin, G. P., and Rossier, J. (2015). Validation of the French Utrecht work engagement scale and its relationship with personality traits and impulsivity. *Eur. Rev. Appl. Psychol.* 65, 19–28. doi: 10.1016/j.erap.2014.10.003

Zeijen, M. E. L., Peeters, M. C. W., and Hakanen, J. J. (2018). Workaholism versus work engagement and job crafting: what is the role of self-management strategies? *Hum. Resour. Manag. J.* 28, 357–373. doi: 10.1111/1748-8583.12187

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 in Psychology

ORIGINAL RESEARCH
published: 03 August 2021
doi: 10.3389/fpsyg.2021.705228

Check for updates

Spanish Adaptation of the Inventory Brief Child Abuse Potential and the Protective Factors Survey

Arturo Sahagún-Morales[1], Amada Ampudia Rueda[1], Salvador Chacón-Moscoso[2,3], Susana Sanduvete-Chaves[2], Ennio Héctor Carro Pérez[4] and Patricia Andrade Palos[1]*

[1] Facultad de Psicología, Universidad Nacional Autónoma de Mexico, Ciudad de Mexico, Mexico, [2] Departamento de Psicología Experimental, Universidad de Sevilla, Sevilla, Spain, [3] Departamento de Psicología, Universidad Autónoma de Chile, Santiago de Chile, Chile, [4] Facultad de Derecho y Ciencias Sociales, Centro de Investigación y Desarrollo Tecnológico Aplicado al Comportamiento, Universidad Autónoma de Tamaulipas, Tampico, Mexico

Child maltreatment is a public health problem with different consequences depending on the form of abuse. Measuring risk and protective factors has been a fertile ground for research, without involving instruments with sufficient evidence of validity. The aim of the study was to gather evidence of validity and reliability of the Inventory Brief Child Abuse Potential (IBCAP) and Protective Factors Survey (PFS) in the Mexican population. The instruments were translated into Spanish. In a non-probabilistic sample of 200 participants, the 7-factor model for the IBCAP [comparative fit index (CFI) = 0.984; root mean square error of approximation (RMSEA) = 0.067] and the 4-factor model for the PFS (CFI = 0.974; RMSEA = 0.061) were confirmed, showing adequate fit indices. Reliability was estimated and evidence of convergent, divergent, and discriminant validity was collected, controlling for effects of social desirability. We also report interpretability statistics of the scores. We achieved solid progress in the development of instrumentation that allows determining the presence or absence of protective and risk factors for child abuse.

Keywords: validity evidences, reliability, norms and interpretation of tests scores, child abuse, protective and risk factors

OPEN ACCESS

Edited by:
Laura Galiana,
University of Valencia, Spain

Reviewed by:
África Borges,
University of La Laguna, Spain
Christophe Leys,
Université Libre de Bruxelles, Belgium

****Correspondence:***
Arturo Sahagún-Morales
asahagun@comunidad.unam.mx

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

Received: *04 May 2021*
Accepted: *30 June 2021*
Published: *03 August 2021*

Citation:
Sahagún-Morales A, Ampudia Rueda A, Chacón-Moscoso S, Sanduvete-Chaves S, Carro Pérez EH and Andrade Palos P (2021) Spanish Adaptation of the Inventory Brief Child Abuse Potential and the Protective Factors Survey. Front. Psychol. 12:705228. doi: 10.3389/fpsyg.2021.705228

INTRODUCTION

The World Health Organization defines child abuse as all forms of physical and/or emotional ill-treatment, sexual abuse, neglect, or negligent treatment or commercial or other exploitation, resulting in actual or potential harm to the health, survival, development or dignity of a child in the context of a relationship of responsibility, trust, or power [Organización Panamericana de la Salud (OPS) Oficina Regional para las Américas de la Organización Mundial de la Salud (OMS), 2003, p. 65], being the most widely used definition worldwide (Chahine, 2014; Weibela et al., 2017; Assink et al., 2018; Hayes and O'Neal, 2018; Cicchetti and Handley, 2019; Kaufman and Torbey, 2019; Marco et al., 2019; Sigad et al., 2019).

Studies point to physical abuse as a form of child abuse, which is prevalent in the world (Kessler et al., 2010). However, estimates vary according to the measurement methodologies used. Regarding its prevalence, self-reported physical abuse records 226 victims per 1,000 boys and girls, with no differences in prevalence by sex (Stoltenborgh et al., 2013). Sexual abuse is the most studied form of child abuse and its prevalence by sex worldwide records 180 victims per 1,000 girls and 76 per 1,000 boys (Stoltenborgh et al., 2011, p. 89). There is little information on the prevalence of emotional abuse compared to physical and sexual abuse [Organización Panamericana de la Salud (OPS) Oficina Regional para las Américas de la Organización Mundial de la Salud (OMS), 2003]; however, the self-reported prevalence of emotional abuse is found to be 363 victims per 1,000 boys and girls (Stoltenborgh et al., 2012a). In prevalence by sex, 363 victims of emotional abuse are reported for every 1,000 boys and 384 for every 1,000 girls (Stoltenborgh et al., 2012a). On the other hand, Stoltenborgh et al. (2012b) reported that only 16 scientific studies have recorded the self-reported prevalence. The worldwide prevalence of child abuse is found to be 163 self-reported victims per 1,000 children in physical neglect, and 184 victims per 1,000 children in emotional neglect (Stoltenborgh et al., 2012b).

In Mexico, the System for the Integral Development of the Family conducted in 2014, at the national and state level, an average of 152 children and adolescents for probable cases of child abuse, of which 35% correspond to abuse physical, 27% to neglect of care, 18% to emotional abuse, 15% to abandonment, and 4% to sexual abuse (COMPREVNNA, 2017). The same year, the National Institute of Statistics and Geography (INEGI) reported that 83% of the victims of child violence between the ages of 12 and 17 had as a perpetrator a person known as members of the household, partner, classmates and work, family, close friends, or acquaintances by sight (INEGI, 2016). Between 2010 and 2014, the main victims of child homicide were men aged from 15 to 17 years (INEGI, 2016).

The consequences of child abuse vary according to the form of abuse; in addition, there are consequences due to multiple forms of abuse. The OMS (2016) reports that child abuse is a cause of stress and is associated with early brain development disorders. In adults who have been abused in childhood, there is a greater risk of suffering and committing acts of violence, suffering depression and obesity, consuming snuff, showing sexual high-risk behavior, unwanted pregnancies, alcohol and excessive drugs, among others behavioral, physical, and mental problems. Therefore, child abuse indirectly contributes to heart disease, cancer, suicide, and sexually transmitted infections (OMS, 2016).

In general, abuse is a risk factor for a wide range of psychiatric disorders, substance abuse, behavioral problems, physical and emotional health problems, decreased well-being, propensity to commit child abuse, impaired cognitive and emotional development in children, feelings of hopelessness, low self-esteem, low self-esteem, low satisfaction with life, low sense of social support, and attachment style problems (Kessler et al., 2010; Stoltenborgh et al., 2011, 2012a, 2013, 2014; INEGI, 2016; Weibela et al., 2017; Kaufman and Torbey, 2019; Liel et al., 2019).

Taking into account the different existing definitions of child abuse that hinder the collection of verifiable information, it is considered that the official figures understate (between 50 and 80% of cases of child maltreatment are not recorded) the real prevalence of abuse (Schwab-Reese et al., 2018), so it is important to study the associated factors, both in terms of increased risk and protective factors.

Protective factors of child abuse are defined as "characteristics of a family or relationship that reduces the likelihood of child maltreatment" (Sprague-Jones et al., 2019, p. 122). In contrast, the potential factors for child abuse, or risk factors, are understood as the characteristics of a person, environment, or society that increase the probability of occurrence of child abuse (Aschengrau and Seage, 2019). Both protective and risk factors for child abuse include a wide range of environmental characteristics (physical and social), behaviors, thoughts, beliefs, and attitudes occurring in the context of a relationship, which regulate the behaviors of the members of this relationship, in such a way that they are more or less likely to commit, voluntarily or involuntarily, acts that mistreat a minor.

Studies have identified recurrent risk and protective factors for child maltreatment (McCoy and Keen, 2014). Family functioning (Thornock et al., 2019), parental relationship (McCoy and Keen, 2014), preparation of parents in parenting strategies and parental knowledge (Albertos et al., 2016; Morrongiello et al., 2019), parental values (McCoy and Keen, 2014), the participation of the child in family activities (McCoy and Keen, 2014), social support (Cutrona et al., 1994; Piko, 2000), and even community environments and characteristics of the physical properties of the home (Labella and Masten, 2018) are some of the most important protective factors (McCoy and Keen, 2014). In terms of risk factors, poverty (Delgado, 2016), family stress (Musitu and Callejas, 2017), family and intimate partner violence (Henry, 2018; Lawson, 2019), among others have been reported (McCoy and Keen, 2014).

Measuring risk and protective factors have been fertile grounds for research, without implying these instruments with sufficient validity evidence. In this case, we worked with the second edition of the Protective Factors Survey (PFS; Sprague-Jones et al., 2019) and the Inventory Brief Child Abuse Potential (IBCAP; Ellonen et al., 2019).

The IBCAP is a self-report instrument developed by Ondersma et al. (2005) from the Inventory Child Abuse Potential (ICAI; Milner, 1986). It is answered using dichotomous items of agreement/disagreement. It is a brief inventory that includes 24 items for the risk factor, scales, plus nine items for the ICAI validity scales. Stability has been reported in the factors that make up the IBCAP, showing, in the US population (Ondersma et al., 2005), a structure of seven factors, which include: Distress, Family Conflict, Rigidity, Happiness, Feelings or persecution, Loneliness and Financial insecurity. Likewise, the version by Ondersma et al. (2005) maintains the scale of lies and random response (validity scales) of the ICAI. Although the IBCAP shows acceptable validity evidences in its different versions (Ondersma et al., 2005; Ellonen et al., 2019; Liel et al., 2019), more validity evidences are required that we will seek to collect in this study.

For its part, the PFS was developed in 2005 by the FRIENDS National Center in collaboration with the Institute for Educational Research and Public Service at the University of Kansas (FRIENDS National Center for Community Based Child Abuse Prevention, 2021). The creation of PFS responded to the need for a reliable and valid instrument for the evaluation of child abuse prevention programs, given that at that time, there was no adequate instrument for measuring changes in multiple protective factors for child abuse and neglect (Sprague-Jones et al., 2019). The PFS has 20 items in 7-point Likert scale and is designed for caregivers of minors, users of prevention of child abuse services. It has a traditional version (non-retrospective self-report) and a retrospective response version and measures the factors: (a) Family Functioning and Resilience, (b) Social Supports, (c) Concrete Supports, and (d) Nurturing and Attachment; in addition to items that indicate knowledge of the development of parenting and child together without enough features to speak of a latent factor. All factors have a good reliability (FRIENDS National Center for Community Based Child Abuse Prevention, 2020).

Starting with the first PFS, a Spanish short version has been developed (for the Latino population residing in the United States, Conrad-Hiebner et al., 2015), and the second edition was also in retrospective and non-retrospective self-report format (Sprague-Jones et al., 2019). Likewise, the relationship of PFS with instruments like the Perceived Stress Scale (PSS), the PRIME-MD Patient Health Questionnaire and the same IBCAP (Counts et al., 2010) has been tested. The second edition of the PFS has 29 items in 5-point Likert scale and measures the following factors: (a) Family Functioning and Resilience, (b) Social Supports, (c) Concrete Supports, (d) Nurturing and Attachment and (e) Caregiver/Practitioner Relationship, this last factor being the only one with poor internal consistency (FRIENDS National Center for Community Based Child Abuse Prevention, 2018); although there is a more recent version and with better levels of internal consistency (Sprague-Jones et al., 2019), this remains precisely as the one used in this study.

In both the IBCAP and the PFS, the psychometric analyzes are limited to the internal consistency determined with the Cronbach's Alpha coefficient and the Exploratory Factor Analysis (EFA). This aspect is remarkable because they are insufficient and inadequate to determine the reliability and validity of an instrument (Batista-Foguet et al., 2004; Agbo, 2010). Cronbach's Alpha coefficient adequately estimates only the true internal consistency when the items are at least tau-equivalents, assuming that it is not tested and that it is practically impossible to fulfill, in addition to the fact that unidimensionality is required, which is not fulfilled in multidimensional scales (Contreras-Espinoza and Novoa-Muñoz, 2018). In the EFA, the euphemism for rotation (Batista-Foguet et al., 2004) is an arbitrary element in the decision about matching the items to the latent factor, leading to different interpretations of the same analysis according to the rotation method factor chosen. Another methodological flaw lies in assuming continuity in items that are inherently ordinal (Hoffmann et al., 2013), leading to an indiscriminate use of statistical methods involving measurement levels above the ordinal as Pearson's correlation.

Either the validation studies do not present evidence or they only present correlation matrices between variables of a nomological network without controlling for social desirability effects (Mikulic et al., 2016) or reliability attenuation effects (Domínguez-Lara, 2017) while that with regard to discrimination by item and discriminant validity, there are no indicators that demonstrate them. Finally, although both instruments have versions in different languages, there is no version that presents validity or reliability indices in the Mexican population, a crucial aspect considering that its use is common in child abuse prevention programs (Chacón-Moscoso et al., 2016, 2019).

Therefore, this paper aims to gather evidence of validity and reliability of the IBCAP and PFS in the Mexican population, resolving faults present in the previous psychometric studies.

MATERIALS AND METHODS

IBCAP and PFS Spanish Translation Study

Participants

A non-probabilistic intentional sample was used. We worked with three translators whose native language is Spanish. The first translator is an expert translator, the second is a licensed psychologist with experience in working with children and parents, and the third is a Doctor of Psychology with experience in measuring the psychological evaluation. Everyone worked independently, without knowing the research objectives to maintain masked the process. Additionally, there was an evaluator of the translations who has experience in the development of psychological measurement instruments.

Instruments

Inventory Brief Child Abuse Potential

The IBCAP (Ellonen et al., 2019) consists of 21 items divided into five factors: *Loneliness and distress* (LD, nine items), *Impact of others* (IO, four items), *Family conflict* (FC, three items), *Rigidity* (R, three items), and *Financial insecurity* (FI, two items). Here the Finnish version which responds by dichotomous items of agreement/disagreement and which has a total Cronbach's Alpha of 0.781 was used for its adaptation.

Protective Factors Survey

The PFS (Sprague-Jones et al., 2019) consists of 29 items distributed into five factors: *Family Functioning and Resilience* (FFR, four items), *Nurturing and Attachment* (NA, seven items), *Social Supports* (SS, seven items), *Concrete Supports* (CS, eight items), and *Caregiver/Practitioner Relationship* (CPR, three items). It is a self-report instrument that is answered through 5-point Likert-type items with labels of 1 = not at all like my life, 2 = not much like my life, 3 = somewhat like my life, 4 = quite a lot like my life, and 5 = just like my life, for the FFR, NA, and SS factors respectively; of 1 = never, 2 = rarely, 3 = sometimes, 4 = often, and 5 = almost always for the CS factor; and 1 = strongly agree, 2 = agree, 3 = neither agree nor disagree, 4 = disagree, and 5 = strongly disagree for the CPR factor. Here, the American version of Sprague-Jones et al. (2019) which explains 54.1% of variance and has Cronbach's aalpha >0.750, was used for its adaptation. It was decided not to use the Spanish short version by Conrad-Hiebner et al. (2015) because, despite having been validated in the Spanish-speaking population, it only has 15 items, an aspect that limits the use of the tool in the evaluation at the individual level due to the high impact of the standard error of measurement (SEM) on short instruments (Sijtsma, 2011). Added to the above is the fact that the validation study was developed in the residents of the United States, a fact that implies important cultural differences within the population living in Mexico.

Format for Translation

Translation format was developed with 21 items of the IBCAP (Ellonen et al., 2019) and 29 of the PFS (Sprague-Jones et al., 2019). This instrument is the one that was presented to the translators for the translation of all items. It consists of three columns, one where the original English version, one for the translators to place their version translated into Spanish and another column is placed where the translators make observations about each item if they deem it necessary.

Procedure

Translation Process

Although the use of backward translation design is common, it has been documented that this design frequently generates translations in the target language (Spanish, in this case) that facilitate a reverse translation but do not maximize the suitability of the translation to the target population (International Test Commission, 2017). Considering this disadvantage, a forward translation design with multiple translators and subsequent revision was chosen (Muñiz et al., 2013; Hambleton and Patsula, 2014) because it allows for identifying and eliminating discrepancies between the different direct translations and creating a single version in the target language (International Test Commission, 2017). Translators were contacted *via* e-mail and the translation form was sent. Translations were performed over a period of 17–33 calendar days. The translators were asked to translate each item from English into Spanish, prioritizing meaning over literality. It was specified to all that the Spanish version should have the colloquial language.

Translation Evaluation and Selection Process

Concluded translations were compared with the original English version to evaluate and select the best translations. This task was performed by a psychologist with expertise in the subject of child abuse (author of this work) without prior knowledge of the identity of the persons who carried out the translation. He ruled out, one by one, each translation of the 50 items (150 translations in total) choosing the one he considered the best. The reviewers could choose one of the following options: Translation 1 is better, Translation 2 is better, Translation 3 is better, Translations 1 and 2 are better, Translations 1 and 3 are better, Translations 2 and 3 are better, All three translations are just as good.

Item Writing Process From Translations

With selected translations, drafts of the items of the IBCAP and PFS were developed. The writing consisted of using the terms of the selected translations to write a version that kept the meaning of the original item. At this stage, adaptations of the items to be applicable to people were performed with and without children, and to be answered using the same scale of responses (e.g., 7-point Likert scale). Also, sometimes several items were drawn from a single item because the original version contained more than an idea, something that could generate confusion among respondents.

Study Results of Spanish Translation

In the translation process, the IBCAP proceeded from 21 to 30 items. After translating the Finnish version of Ellonen et al. (2019), one of the translators recommended using the German version of Liel et al. (2019) as well. It was decided to comply with the recommendation because both the versions have the most recent validation studies up to the moment of doing this research, in addition to sharing 76.19% (16) of the items (the five items that were exclusively part of the German version were translated by the first author of this study focusing on the functional rather than on the literal equivalence and avoiding cultural references, idiosyncratic items, and inadequate response formats as recommended by the International Test Commission, 2017). Therefore, to the 21 items of the Finnish version of Ellonen et al. (2019), translated by the panel of translators (Muñiz et al., 2013; Hambleton and Patsula, 2014), the 5 items of the German version of Liel et al. (2019), translated by the first author of this paper, were added. The integration of both the versions resulted in a 7-factor theoretical structure in which the Impact of Others, Family Conflict, and Rigidity factors of the Finnish version remained intact, but the Loneliness and Distress factor (LD, nine items) was separated into Loneliness (L, four items) and Distress (D, four items) factors, in addition to the Unhappiness factor (U, three items) which was only found in the German version of Liel et al. (2019). Furthermore, when integrating both versions, the Financial Insecurity (FI) factor was made up of a single item, which is why three items were created directly in Spanish that complemented the factor; these items were developed by the first author of this paper. The resulting seven factors are consistent with the original version of Milner (1986). Translations and changes of the two original English versions of the IBCAP and preliminary Spanish version are detailed in **Appendix A**.

In the case of PFS, it proceeded from 29 to 49 items, but the original 5-factor structure of Sprague-Jones et al. (2019) was maintained. It is also possible to find all the translation details and modifications made in **Appendix A**.

Study of Evidence of Validity and Reliability of the IBCAP and PFS

Participants

An accidental non-probabilistic sample was used (Kerlinger and Lee, 2002). The sample size was determined in 200 participants because it is an amount necessary to obtain classic statistical items as well as a stable correlation matrix for the development of factor analysis (Downing and Haladyna, 2006). Because it was sought to work with a general population, the only inclusion criteria were that the participants were between 18 and 65 years and were residing in Mexico at the time of research. There were no misses in the sample during the development of the research. The sociodemographic characteristics of the sample are presented in **Table 1** and the structural characteristics of the families are presented in **Appendix B**.

Instruments

Inventory Brief Child Abuse Potential Translated

The translated version of the IBCAP developed in the previous phase was used. It is made up of 30 items distributed in

TABLE 1 | Sociodemographic characteristics of the participants ($N = 200$).

Characteristic	f/M	%/SD	Characteristic	f	%	Characteristic	f	%
People in the same home	3.84	1.83	Maximum degree of study			Total Monthly Income		
Age	31.79	13.12	Incomplete or in-process high school	3	1.5	Between $0 and 2,699	17	8.5
Sex			Complete high school	22	11	Between $2,700 and 6,799	47	23.5
Men	44	22	Incomplete or in-process bachelor's degree	75	37.5	Between $6,800 and 11,599	60	30
Women	156	78	Completed bachelor's degree	59	29.5	Between $11,600 and 34,999	64	32
Children			Incomplete or in-process specialty	2	1	Between $35,000 and 84,999	11	5.5
Yes	68	34	Completed specialty	4	2	$85,000 or more	1	0.5
Do not	132	66	Incomplete or in-process mastery	11	5.5			
Marital status			Complete mastery	14	7			
Married	39	19.5	Incomplete or in the process PhD	7	3.5			
Divorced	7	3.5	Complete PhD	3	1.5			
Single	129	64.5	History of alcohol / drug abuse					
Free Union	22	11	Do not	186	93			
Widower	3	1.5	Yes	14	7			

f, absolute frequency; %, relative frequency; M, mean; SD, standard deviation.

seven factors: *Loneliness* (L, six items), *Distress* (D, four items), *Impact of Others* (IO, four items), *Family Conflict* (FC, four items), *Rigidity* (R, four items), *Financial Insecurity* (FI, five items), and *Unhappiness* (U, three items). The response options were adjusted to seven points from 1 (*Total disagreement*) to 7 (*Total agreement*).

Protective Factors Survey Translated

The translated version of the PFS developed in the previous phase was used. It is made up of 49 items divided into five factors, which include: FFR, four items; NA, seven items; SS, 15 items; CS, 20 items; CPR, three items. The response options for the different factors were standardized on a 7-point scale from 1 (*Total disagreement*) to 7 (*Total agreement*), although in 13 items of the CS factor, the option, not applicable was also added.

Balanced Inventory of Desirable Responding

To control the effects of social desirability, the BIDR (Mikulic et al., 2016) was used. The BIDR consists of 18 items that make up a single factor, *Social Desirability* (SDes). It is a self-report instrument that is answered by Likert-type items with seven points from 1 (*False*) to 7 (*True*). In this study, the Spanish version of Mikulic et al. (2016) was validated using Confirmatory Factor Analysis (CFA) with polychoric correlations (Holgado-Tello et al., 2008; Brown, 2015; Desjardins and Bulut, 2018) and estimation of unweighted least squares with robust standard errors and test statistic adjusted to the mean (ULSM; Shi et al., 2018). The results of the validation of the BIDR are presented in this section because they are not part of the central objective of the research, but correspond to a secondary analysis, that is necessary for the fulfillment of the objectives. It was obtained a reduced version (nine items) with good fit [χ^2 (26) = 38.605, $p = 0.053$; $\chi 2/\text{df} = 1.485$; CFI = 0.987; TLI = 0.982; RMSEA = 0.049, 95% CI (0.000, 0.090), $p = 0.466$; SRMR = 0.049] in a two-factor model (*Self-deception* and *Printing Handling* factors), such as that found in the Mexican population by Moral de la Rubia et al. (2012). In this study, evidence of convergent validity was obtained through the average variance extracted (AVE) of the Factors ≥ 0.500 (Fornell and Larcker, 1981; Cheung and Wang, 2017) as well as the factor loadings (λ) ≥ 0.500 (Cheung and Wang, 2017); evidence of discriminant validity using the $r_{\text{between}-\text{factors}} \leq 0.700$ (Cheung and Wang, 2017) and the $r^2_{\text{betweenfactors}} <$ AVE (Fornell and Larcker, 1981); evidence of discrimination by item with the corrected total-element correlation, (r_{tec}) > 0.200 (Abad et al., 2011); and evidence of total internal consistency and by factors with the coefficients, α_{Ordinal}, ω_{Ordinal}, and $\text{GLB}_{\text{Ordinal}} > 0.700$ (Trizano-Hermosilla and Alvarado, 2016; George and Mallery, 2017, see full psychometric properties of Spanish version of BIDR-9 in **Appendix C**).

Procedure

For reasons of the quarantine due to the Covid-19 pandemic, the instruments were applied *via* Google Forms. Digital forms were distributed in 19 states of Mexico using Facebook Ads service (https://www.facebook.com/permalink.php?story_fbid=104765114762260&id=104716831433755). This system allows sampling by establishing diffusion points in the states of the Mexican Republic with high population density or that are physically very distant from each other, such as Nuevo León and Yucatán, for example. Responses were collected over a period of 31 calendar days. The form included an informed consent and confidentiality statement. The study design was non-experimental, single-group, and cross-sectional.

Data Analysis

Validity Evidence Concerning the Internal Structure of the Instrument

Confirmatory factor analysis taking the matrix, polychoric correlations (Holgado-Tello et al., 2008 Brown, 2015; Desjardins and Bulut, 2018) was used. The estimation method used unweighted least squares with robust standard errors and test statistic adjusted to the mean (ULSM, Shi et al., 2018) due to the lack of multivariate normality (negative Mardia test, Porras, 2016). For the IBCAP-T a structure of seven correlated latent variables was tested, while in the PFS-T a structure of five correlated latent variables was tested. Correlated factor structures were tested in both the IBCAP-T and the PFS-T because the theoretical background suggests that the structures of both constructs are not independent (Ellonen et al., 2019; Liel et al., 2019; Sprague-Jones et al., 2019). Structures with the independent factors were also tested as rival models. The fit was evaluated using the following fit indices and interpretation criteria (Abad et al., 2011; Kline, 2011): Chi square/degrees of freedom (χ^2/df) $\leq$ 3 (good fit); CFI $\geq$ 0.950 (good fit); Tucker–Lewis Index (TLI) $\geq$ 0.960 (good fit); RMSEA $\leq$ 0.060 (good fit) with 90% CI and $p \geq 0.050$, Standardized Root Mean Residual (SRMR) $\leq$ .080 (good fit).

Item Analysis

The discrimination capacity of the items was determined using the corrected total-element correlation, (r_{tec}) > 0.200 (Abad et al., 2011) calculated on totals by factor. Furthermore, to know the contribution of each item to reliability, the reliability coefficient per item (r_{i}) was calculated, expecting values $\geq$ 0.500 (Fornell and Larcker, 1981).

Evidence of Validity Regarding the Relationship With Other Variables

Evidence of convergent, divergent, and discriminant validity was collected. For convergent validity, the AVE of all factors was calculated, with values $\geq$ 0.500 indicative of convergent validity (Fornell and Larcker, 1981; Cheung and Wang, 2017). Also, convergent validity criterion was considered the factor loadings (λ) $\geq$ 0.500 (Cheung and Wang, 2017). Finally, the pattern of correlations between the IBCAP-T and PFS-T factors was evaluated, expecting positive or negative correlations as theoretically expected (calculating the attenuation by reliability and controlling the effect of the SDes using partial correlations); Spearman's Rho coefficient was used in this analysis due to the lack of normality (negative Shapiro–Wilk test). For discriminant validity, the $r_{\text{between}-\text{factors}}$ of each pair of factors of the same scale was calculated, where the values $\leq$ 0.700 being indicative of discriminant validity (Cheung and Wang, 2017). Also, the $r^2_{\text{between}-\text{factors}}$ were compared, indicating discriminant validity as < AVE (Fornell and Larcker, 1981).

Reliability Evidence

McDonald's Omega (ω) and greatest lower bound (GLB) coefficients were used because they have been shown to be better estimators of internal consistency than Cronbach's Alpha coefficient (α, Trizano-Hermosilla and Alvarado, 2016). The latter was also calculated because the coefficients, ω and GLB are not yet widely used; therefore, the coefficient, α allows comparison with other works; However, to reduce the impact of non-compliance with the α coefficient assumptions (Batista-Foguet et al., 2004), the 95% confidence interval (CI) is reported. All internal consistency coefficients were calculated from polychoric correlation matrices (Holgado-Tello et al., 2008; Brown, 2015; Desjardins and Bulut, 2018), and the values >0.700 were considered good (George and Mallery, 2017). Finally, in a complementary way, the maximum and minimum split-half reliability was estimated (Abad et al., 2011) interpreting the scores with the same criteria.

Norms and Interpretation of Test Scores

As criteria for the interpretability of scores, the following statistics by factor were calculated: mean, standard deviation, skewness and kurtosis coefficients, Shapiro–Wilk test, and SEM.

The programming language, R version 4.0.3 was used with lavaan package (R Core Team, 2020) and the software, SPSS v.24 (IBM Corporation, 2016) and Microsoft Excel Professional Plus 2016 (Microsoft Corporation, 2016) were used for the statistical treatment of the data.

RESULTS

Validity Evidence Concerning the Internal Structure of the Instrument

Mardia test indicated no symmetry and kurtosis multivariate indicated both IBCAP-T (symmetry multivariate = 4,106.741, $p < 0.001$; kurtosis multivariate = 22.255, $p < 0.001$) and PFS-T (symmetry multivariate = 2,668.980, $p < 0.001$; kurtosis multivariate = 21.461, $p < 0.001$), for which the ULSM estimation was used. Confirmatory models of each are presented in **Table 2**.

Table 2 shows that the IBCAP-T 7-correlated factor model was confirmed by eliminating five items, fitting better than the original model with 30 items and the modified independent model. In the PFS-T, the NA factor was eliminated, achieving a good fit with a model of 4 correlated factors and 25 items.

The item deletion was performed by the modification indices. These allow decisions for re-specification of the models and reduce the size of the chi-square statistic by removing parameters (Hair et al., 1999; Escobedo-Portillo et al., 2016). Also, an additional criterion to remove items was to present correlated error variances and have a factor loading <0.40. These criteria were considered important because, together, they allow for identifying those items that may not have a relationship with the construct to which they theoretically belong and those items that have an exogenous source of variance (non-random variance unexplained by the construct). This model of re-specification procedure was chosen because it allows for a more parsimonious model to be generated (Brown, 2015). Therefore, the items with high modification indices and factor loadings <0.40 were eliminated one by one until an acceptable fit was reached in the different fit indices.

As can be seen, the contrast of rival models (original vs. modified and correlated vs. modified independent) allows us to safely conclude that the data better fit the theoretical models which include both the elimination of parameters with residuals that covariate with each other (modified models eliminating variables) as a degree of covariation between the factors of the same scale (correlated models). This was true both for IBCAP-T and PFS-T; however, the elimination of NA factor in the PFS-T may indicate a differential functioning of the items in the Mexican culture, in such a way that Nurturing and Attachment

TABLE 2 | Goodness-of-fit indicators of the IBCAP-T and PFS-T confirmatory models with ULSM estimation and polychoric correlation matrix ($N = 200$).

	χ^2	df	p (χ^2)	χ^2/df	CFI	TLI	RMSEA (CI 90%)	p (RMSEA)	SRMR
IBCAP-T									
M1 (30 items)	1,107.976	384	<0.001	2.885	0.968	0.964	0.097 (0.086, 0.109)	<0.001	0.062
M2 (25 items)	9,555.120	275	<0.001	34.746	0.350	0.291	0.412 (0.404, 0.420)	<0.001	0.338
M3 (25 items)	479.541	254	<0.001	1.888	0.984	0.981	0.067 (0.051, 0.083)	0.045	0.049
PFS-T									
M4 (49 items)	4,718.315	1,117	<0.001	4.224	0.759	0.747	0.127 (0.123, 0.131)	<0.001	0.127
M5 (25 items)	1,138.266	275	<0.001	4.139	0.888	0.878	0.126 (0.118, 0.134)	<0.001	0.128
M6 (25 items)	469.795	269	<0.001	1.747	0.974	0.971	0.061 (0.049, 0.073)	0.061	0.066

IBCAP-T, Inventory Brief Child Abuse Potential Translated; PFS-T, Protective Factors Survey Translated; M1, Original 7-factor model; M2, Modified Independent 7-factor model; M3, Modified Correlated 7-factor model; M4, Original 5-factor model; M5, Modified Independent 4-Factor Model; M6, Modified Correlated 4-Factor Model; CFI, Comparative Fit Index; TLI, Tucker–Lewis Index; RMSEA, Root Mean Square Error of Approximation; SRMR, Standardized Root Mean Residual; CI, Confidence Interval; p, p-value.

are manifested differently from what is found in the context of the United States. It is worth mentioning that the variance explained by the factor should be taken with caution because they are correlated structures in which there may be an overestimation of the variance explained; However, the theoretical background of the IBCAP and the PFS suggests that a structure of correlated factors is the most expected one (Liel et al., 2019; Sprague-Jones et al., 2019). The factorial structures of the models with the best fit of the IBCAP-T and the PFS-T are presented in **Figures 1**, **2**, respectively.

Item Analysis

In the item analysis, the results for the IBCAP-T and PFS-T are shown in **Tables 3**, **4**.

Table 3 shows that the IBCAP-T items had discrimination levels that ranged between 0.258 and 0.943, and reliability levels between 0.326 and 0.944. In the PFS-T items, discrimination ranged between 0.473 and 0.848 and reliability ranged between 0.382 and 892 (**Table 4**). In both instruments, the levels of discrimination and reliability were good or excellent. For a list of items of both psychometric instruments in English and Spanish, see **Appendix D**.

Evidence of Validity Regarding the Relationship With Other Variables

Evidence of convergent validity (λ and AVE of the M3 and M6 models) and discriminants (r_{bf} and r^2_{bf}) are presented in **Figures 1**, **2**, and in **Table 5** for the IBCAP-T and PFS-T, respectively. Also, the correlations between the IBCAP-T and PFS-T factors (convergent and divergent validity) are presented in **Table 6**.

It can be seen in **Figures 1**, **2** that the λ meet the criteria ($\lambda > 0.50$) to assume convergent validity for both instruments (except 1 item from the IBCAP-T and 2 items from the PFS-T). Since the factor loadings are the correlation of the item with its latent factor, it is expected that higher values in λ items indicate convergent validity. Meanwhile, the AVE indicates the amount of variance explained by the construct such that the higher the AVE, the more it is argued that the items contribute to the measurement, i.e., high AVE values indicate the convergence of the items of a construct. In this regard, the AVE show that both for the IBCAP-T and the PFS-T, all factors showed an explained variance <0.50 (see **Table 5**).

In terms of discriminant validity, the correlations between factors (r_{bf}), of the same scale indicates the absence of collinearity, that is, the items of one factor measure the same as the items of a different factor. For this reason, although it is expected that there is a low or medium correlation between the factors that make up a scale, it is expected that these correlations do not reach a value high enough to cause confusion in the dimensions of the construct. In the same sense, the Squared correlation between factors (r^2_{bf}) can be understood as the shared variance between the factors of the same scale, that is, between the dimensions of a construct. Thereupon, it is expected that the items of the same factor shared more variance with each other (AVE) than that they share with another factor (r^2_{bf}), so values of r^2_{bf} must be less than the values of AVE to assert discriminant validity. It can be seen in **Table 5** for the IBCAP-T, that only three of the 21 r_{bf} are slightly above 0.700 (see values below the diagonal marked with -); However, when comparing the r^2_{bf} (observe the values above the diagonal marked with -) and the AVE, in each comparison, the AVE values are greater than the r^2_{bf}, which indicates that the variance shared by the items of the same factor is greater than the shared variance between factors. In the PFS-T, all the discriminant validity indicators met the expected criteria.

Regarding the correlations between the IBCAP-T factors and the PFS-T factors, **Table 6** shows that the crude correlations adjusted for reliability increased in a range that goes from 2.70 to 17.89%, which can be interpreted as the percentage of the true correlation that is not registered due to the measurement error. On the other hand, the bias by SD showed, in most of the correlations, lower values than the crude correlations, which represents a high impact of the SD. In terms of convergent and divergent validity, median correlations were found with p-values < 0.05 and 0.01 even after removing the effect of social desirability, although factor 4 of the PFS-T only moderately correlated with factor 7 of the IBCAP- T. In the same sense, factor 6 of the IBCAP-T only moderately correlated with factor 3 of the PFS-T.

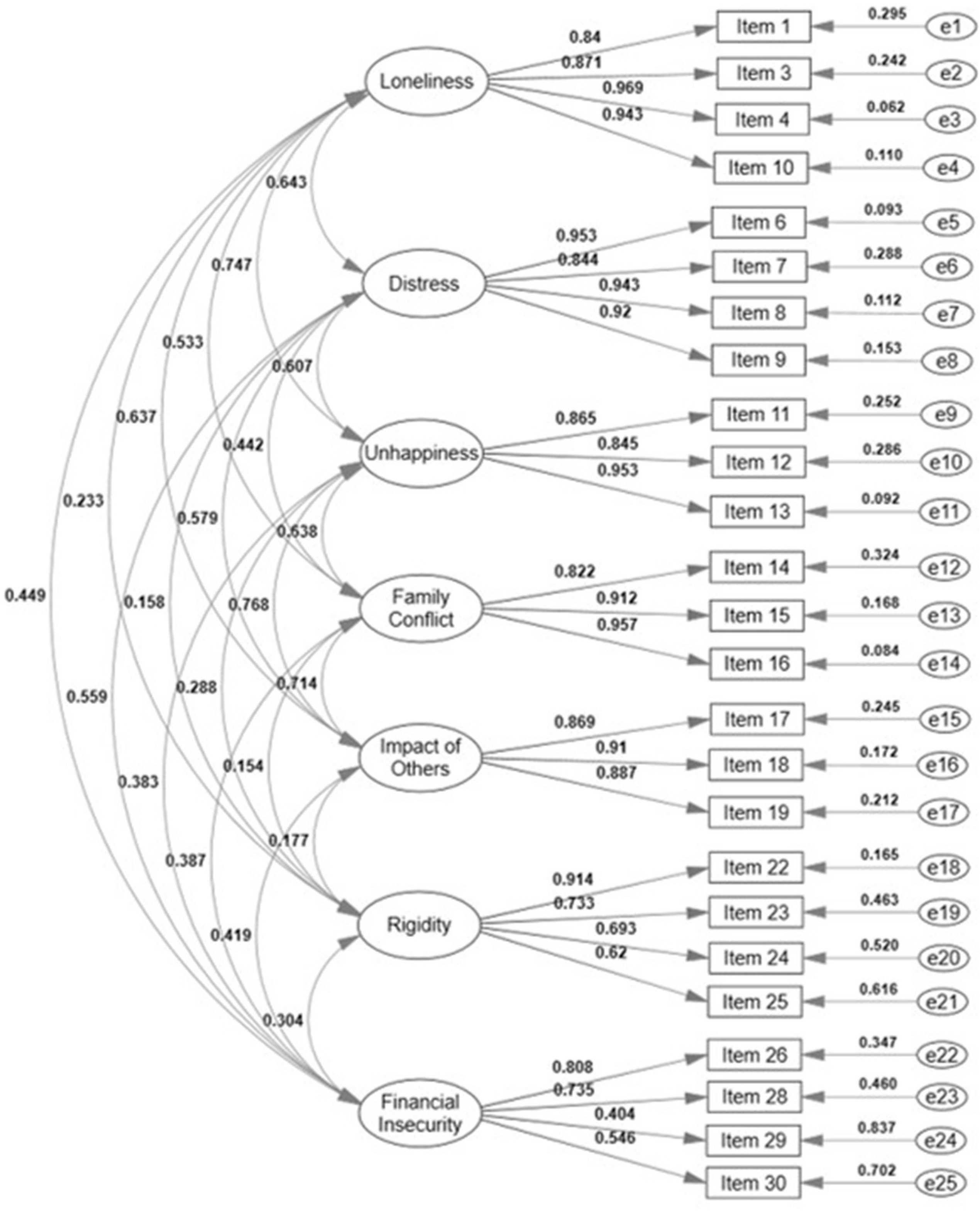

FIGURE 1 | Modified correlated 7-factor model of IBCAP-T. The estimates of the presented factor loadings, variances, and covariances are standardized.

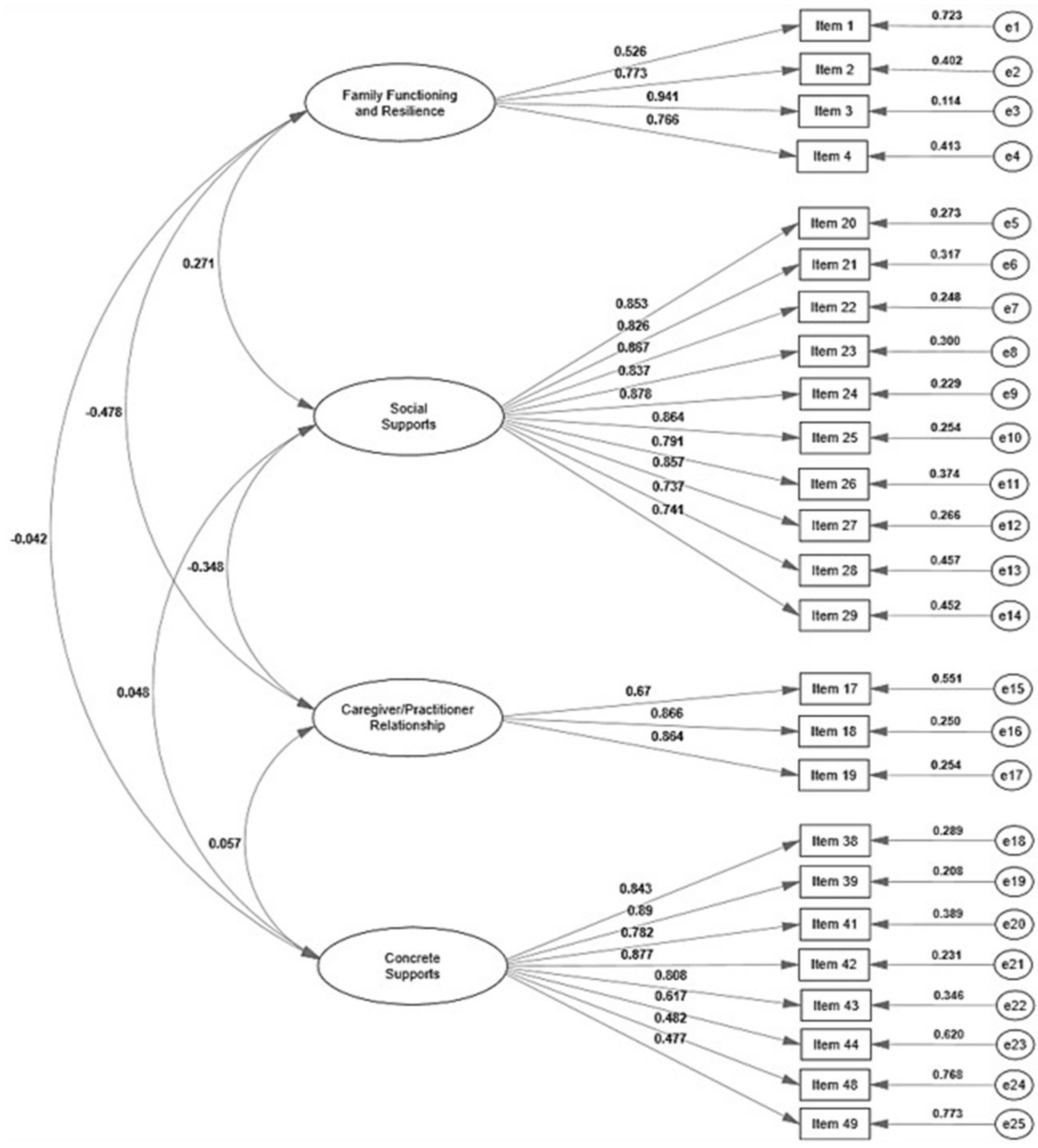

FIGURE 2 | Modified correlated 4-factor model of PFS-T. The estimates of the presented factor loadings, variances, and covariances are standardized.

Reliability Evidence

Tables 3, **4** show the reliability coefficients by factor. It is notable that the only coefficient that did not obtain a value ≥ 0.700 was the α coefficient in factor 7 of the IBCAP-T. On the other hand, both in the IBCAP-T and PFS-T, the relationship α ≤ ω ≤ GLB was maintained.

Norms and Interpretation of Tests Scores

Tables 3, **4** also show that no factor had measures normally distributed. In the IBCAP-T, all the averages were <4 with SD close to 1, while the PFS-T showed means >4 in the FFR and SS factors, and <4 in the CPR and CS factor, the latter having the lower mean (M) and SD (M = 1,458, SD = 1,240). Finally, it is

TABLE 3 | Item analysis, reliability evidence, and statistics for the interpretability of the IBCAP-T ($N = 200$).

Factor	Item	Item analysis		Reliability					Interpretability					
		r_{tec}	r_i	α_{Ord} (CI 95%)	ω_{Ord}	GLB_{Ord}	r_{min}	r_{max}	M	SD	Skew	Kurt	S–W (2-tailed p)	SEM
F1	ITEM1	0.859	0.740	0.949 (0.939, 0.959)	0.950	0.971	0.941	0.954	3.318	1.823	0.460	−0.874	0.928 (< 0.001)	0.310
	ITEM3	0.883	0.783											
	ITEM4	0.943	0.940											
	ITEM10	0.818	0.896											
F2	ITEM6	0.899	0.911	0.954 (0.945, 0.963)	0.954	0.970	0.944	0.970	3.304	1.741	0.406	−0.883	0.941 (< 0.001)	0.302
	ITEM7	0.854	0.746											
	ITEM8	0.874	0.894											
	ITEM9	0.922	0.857											
F3	ITEM11	0.797	0.774	0.917 (0.900, 0.933)	0.920	0.951	0.789	0.791	2.305	1.415	0.767	−0.434	0.844 (< 0.001)	0.313
	ITEM12	0.794	0.747											
	ITEM13	0.906	0.912											
F4	ITEM14	0.763	0.717	0.924 (0.908, 0.938)	0.928	0.941	0.764	0.850	3.067	1.783	0.562	−0.766	0.913 (< 0.001)	0.433
	ITEM15	0.889	0.844											
	ITEM16	0.887	0.919											
F5	ITEM17	0.829	0.780	0.919 (0.902, 0.934)	0.919	0.916	0.813	0.814	3.077	1.782	0.487	−0.855	0.915 (< 0.001)	0.516
	ITEM18	0.831	0.841											
	ITEM19	0.845	0.807											
F6	ITEM22	0.699	0.847	0.832 (0.797, 0.863)	0.835	0.867	0.790	0.867	3.209	1.383	0.185	−0.532	0.969 (< 0.001)	0.504
	ITEM23	0.729	0.613											
	ITEM24	0.567	0.571											
	ITEM25	0.648	0.502											
F7	ITEM26	0.605	0.700	0.699 (0.637, 0.755)	0.726	0.764	0.673	0.733	3.969	1.354	−0.188	−0.481	0.983 (0.019)	0.658
	ITEM28	0.619	0.615											
	ITEM29	0.258	0.326											
	ITEM30	0.485	0.438											

Values that did not meet the defined criteria are highlighted. F1, Loneliness; F2, Distress; F3, Unhappiness; F4, Family Conflict; F5, Impact of Others; F6, Rigidity; F7, Financial Insecurity; r_{tec}, Corrected Total-Element Correlation correlating each item with the total of the factor to which it belongs.; r_i, Reliability by Item following the formula $r_i = \frac{\lambda_i^2}{\lambda_i^2+Var(\varepsilon_i)}$, where λ_i^2 is the factor loading raised to the square of the i-th item, and $Var(\varepsilon_i)$ is the error variance of the i-th item (Fornell and Larcker, 1981); α_{Ord}, Ordinal Cronbach's Alpha Coefficient; CI, Confidence interval; ω_{Ord}, Ordinal McDonald's Omega Coefficient; GLB_{Ord}, Ordinal Greatest Lower Bound Coefficient; r_{min}, minimum split-half reliability; r_{max}, maximum split-half reliability; M, Mean; SD, Standard Deviation; Skew, Skewness coefficient; Kurt, Kurtosis coefficient; S–W, Shapiro–Wilk test; p, p-value; SEM, Standard Error of Measurement.

observed that the factor with the highest SEM was the FI factor of the IBCAP-T, in contrast to the SS factor of the PFS-T that showed the lowest SEM.

DISCUSSION

The main purpose of the present study was to collect evidence of validity and reliability of the IBCAP and PFS in versions translated into Spanish. The results showed that both instruments have adequate psychometric properties.

By doing factor analysis and estimating reliability from polychoric correlation matrices, more refined and robust results were achieved that better reflect the psychometric characteristics of the instruments (Holgado-Tello et al., 2006, 2007, 2008; Brown, 2015; Desjardins and Bulut, 2018). In addition, the collection of different validity indicators and their consistency is a better approximation to reality than those approaches focused on a single indicator, because each of the analysis, estimation method, and psychometric indicator has limitations or even biases that make a complementary approach necessary which allows for a triangulation of results (Kimchi et al., 1991; Shadish, 1993; Letourneau and Allen, 1999; Heale and Forbes, 2013).

The IBCAP-T was the instrument that required the least adjustments to achieve a satisfactory model, since only five items were eliminated but the structure of seven factors was maintained, which are congruent with the factors of the original extended version of Milner (1986) as well as with the short versions of Ondersma et al. (2005), Ellonen et al. (2019) and Liel et al. (2019). It is noteworthy that the CFI and TLI were adequate with the initial 30 items; however, the RMSEA showed values outside the acceptable in the original model, probably because this indicator is sensitive to the number of estimated parameters and sample size (Kline, 2011).

At the item level, in the IBCAP-T, the levels of discrimination and reliability evidenced the potential for a classificatory use of the instrument, given that most of the items adequately differentiate between subjects with high and low true scores, and all of the Items contribute significantly to reliability.

TABLE 4 | Item analysis, reliability evidence, and statistics for the interpretability of the PFS-T ($N = 200$).

Factor	Item	Item analysis		Reliability					Interpretability					
		r_{tec}	r_i	α_{Ord} (CI 95%)	ω_{Ord}	GLB_{Ord}	r_{min}	r_{max}	M	SD	Skew	Kurt	S–W (2-tailed p)	SEM
F1	ITEM1	0.503	0.421	0. 838 (0.805, 0.868)	0.847	0.883	0.790	0.876	5.508	1.265	−1.011	0.557	0.907 (<0.001)	0.433
	ITEM2	0.791	0.658											
	ITEM3	0.709	0.892											
	ITEM4	0.693	0.650											
F2	ITEM20	0.829	0.758	0.955 (0.946, 0.963)	0.956	0.975	0.905	0.972	4.924	1.633	−0.784	−0.293	0.919 (<0.001)	0.258
	ITEM21	0.821	0.723											
	ITEM22	0.839	0.778											
	ITEM23	0.816	0.736											
	ITEM24	0.824	0.793											
	ITEM25	0.848	0.773											
	ITEM26	0.773	0.679											
	ITEM27	0.844	0.763											
	ITEM28	0.730	0.617											
	ITEM29	0.722	0.621											
F3	ITEM17	0.673	0.549	0.843 (0.811, 0.872)	0.850	0.886	0.722	0.813	3.205	1.616	0.324	−0.865	0.949 (<0.001)	0.546
	ITEM18	0.795	0.776											
	ITEM19	0.662	0.773											
F4	ITEM38	0.784	0.745	0.896 (0.875, 0.915)	0.902	0.922	0.818	0.931	1.458	1.240	1.399	2.134	0.878 (<0.001)	0.346
	ITEM39	0.825	0.811											
	ITEM41	0.734	0.668											
	ITEM42	0.815	0.792											
	ITEM43	0.744	0.700											
	ITEM44	0.583	0.499											
	ITEM48	0.482	0.386											
	ITEM49	0.473	0.382											

Values that did not meet the defined criteria are highlighted. F1, Family Functioning and Resilience; F2, Social Supports; F3, Caregiver/Practitioner Relationship; F4, Concrete Supports; r_{tec}, Corrected Total-Element Correlation correlating each item with the total of the factor to which it belongs.; r_i, Reliability by Item following the formula $r_i = \frac{\lambda_i^2}{\lambda_i^2+Var(\varepsilon_i)}$ where λ_i^2 is the factor loading raised to the square of the i-th item, and $Var(\varepsilon_i)$ is the error variance of the i-th item (Fornell and Larcker, 1981); α_{Ord}, Ordinal Cronbach's Alpha Coefficient; CI, Confidence interval; ω_{Ord}, Ordinal McDonald's Omega Coefficient; GLB_{Ord}, Ordinal Greatest Lower Bound Coefficient; r_{min}, minimum split-half reliability; r_{max}, maximum split-half reliability; M, Mean; SD, Standard Deviation; Skew, Skewness coefficient; Kurt, Kurtosis coefficient; S–W, Shapiro–Wilk test; p, p value; SEM, Standard Error of Measurement.

In terms of validity, the IBCAP-T successfully met all indicators, being a measure that provides valid test scores even after considering the effect of social desirability. Likewise, the correlations between the factors of the IBCAP-T and the PFS-T were congruent with what was expected, since negative (divergent) correlations were found with the factors of SS and FFR, and positive (convergent) with CPR and CS, although the latter only had a medium relationship with the FI factor. This lack of relation of the IBCAP-T factors with the may be due, in part, to the effect of social desirability on the responses of the subjects; that is, the respondents have a way of answering which tend to be self-positive descriptions, such that their responses are consistently different from their true values (Mikulic et al., 2016).

In the PFS-T, the 5-factor model of Sprague-Jones et al. (2019) was not confirmed and the 4-factor model that was confirmed does not coincide in content with that reported by the FRIENDS National Center (FRIENDS National Center for Community Based Child Abuse Prevention, 2018) since the CPR factor that was confirmed in this study is only found in the version by Sprague-Jones et al. (2019). It should be noted that this factor is named for its use in the United States in abuse prevention programs; however, in this study, the factor can be better interpreted if it is considered a measure of relationship with others in general. The fact that the original 5-factor model was not confirmed can be partially explained by the different changes that were made in the scale, both in the response options (all items were unified on a scale from 1 to 7) and in the disaggregation of some items (see **Appendix A**), aspect that can also explain the elimination of 24 items.

Despite the modifications made to the PFS-T, the items of the adjusted 4-factor model showed adequate levels of discrimination and reliability. The same is true at the level of factors for internal consistency and validity of the different indicators. However, in the correlations with the IBCAP-T factors, the CPR factor was the only one that correlated with the Rigidity factor. It is noteworthy that the CPR factor showed the highest correlations with the IBCAP-T factor but also showed the greatest effects on social desirability, since it consistently showed the highest levels of difference between the estimated true correlation and the correlation controlled by SD.

TABLE 5 | Evidence of convergent and discriminant validity of the IBCAP-T and PFS-T (N = 200).

	IBCAP-T						
	1	2	3	4	5	6	7
1 L	–	0.413	0.558	0.284	0.406	0.054	0.202
2 D	0.643	–	0.368	0.195	0.335	0.025	0.312
3 U	0.747	0.607	–	0.407	0.590	0.083	0.147
4 FC	0.533	0.442	0.638	–	0.510	0.024	0.150
5 IO	0.637	0.579	0.768	0.714	–	0.031	0.176
6 R	0.233	0.158	0.288	0.154	0.177	–	0.092
7 FI	0.449	0.559	0.383	0.387	0.419	0.304	–
AVE	0.897	0.908	0.876	0.887	0.876	0.683	0.523

	PFS-T			
	1	2	3	4
1 FFR	–	0.073	0.228	0.002
2 SS	0.271	–	0.121	0.002
3 CPR	−0.478	−0.348	–	0.003
4 CS	−0.042	0.048	0.057	–
AVE	0.587	0.683	0.648	0.547

Values that did not mee the defined criteria are highlighted. The correlation between factors (r_{bf}) is below the diagonal. The coefficient of determination between factors (r^2_{bf}) is above the diagonal. AVE, Average variance extracted.

On the interpretability, the report of the SEM is important for estimating intervalar true scores of an individual (Gempp, 2006), which is an important aspect for the use of the instrument in individual diagnoses. In this sense, the extension of the instruments ratifies the practical potential of the instruments in individual evaluation, since being long instruments (more than 20 questions), the effects of the SEM in decisions at the individual level are mitigated (e.g., correctly conclude the presence of risk factors in an individual; being able to detect medium effects in before–after comparisons and not just large effects) (Sijtsma, 2011).

It is important to note that the IBCAP-T showed a more robust behavior with respect to previous validation studies since the modifications made to the Spanish version were minor, achieving comparability with other existing versions. This does not happen with the PFS-T because substantial changes were introduced to the adaptation to the Mexican population. A direct consequence of the lack of robustness of PFS-T is the inability to make comparisons with other versions of the instrument. However, the numerous validity and reliability evidence obtained, as well as the statistics for the interpretability obtained in this study indicate that the use of the PFS-T in the Mexican population is extremely promising in terms of being able to have indices of validity, reliability, and feasibility, which are unprecedented in Mexico and which will allow investigations into child abuse area.

This study has some limitations. It is important to explore semantic aspects that may affect the quality of the items and that could have been omitted due to the lack of a back-translation process. However, this aspect is cushioned by the contribution of three experts, one on linguistic issues, another on expertise on the subject, and a third in the development of psychometric instruments. In terms of the heterogeneity and sample size, it is necessary to carry out subsequent studies that analyze, in larger samples, the differential functioning of the instruments mainly in variables, such as sex, family structure, and the preference of children. However, the intrinsic complexity of child abuse, the territorial extent, and the cultural diversity of the country always demand a careful use of these instruments in the Mexican territory, contemplating variables, such as the region (north, center, or south) and socioeconomic conditions, as well as the inclusion of indigenous communities. The analysis of all these variables is beyond the scope of this study; however, valuable information is provided on the usability of the two tools. On the other hand, when making adjustments to the factorial structure after checking the modification indices (and therefore make apparently exploratory use of the CFA), there is a risk of biases due to "chance capitalization" (Batista-Foguet et al., 2004); However, given the severe defects of the EFA (Batista-Foguet et al., 2004), and the strengths of the Confirmatory Factor Analysis (CFA) (Brown, 2015), the process of modifying the models by eliminating items using the CFA is highly preferable to the use of the EFA, despite the probable chance capitalization. Furthermore, the re-specification of the models by eliminating the items with correlated error variances is a process that generates alternative models that have a legitimate psychometric interpretation, in contrast to the process of "correlating the error variance of the parameters" which lacks psychometric interpretation, despite its relatively extended use. Although, in general terms, re-specification can be considered a form of exploration, the conditions of its development are considerably different because the re-specification that we carry out in this work started from a pre-existing theoretical model that was gradually simplified (more parcimonious models) and that it is interpretable within the framework of general theories that contain it (child abuse theories); There were no cross-loads (greater restriction in the specification compared to the EFA) and it was constantly possible to contrast the fit of re-specified models, which allowed to achieve solidly integrated and psychometrically interpretable factorial structures. Despite all of the above, for further development, we intend to strengthen the inferences made from the results obtained in this work, checking the model in different and larger samples. A fourth limitation lies in the size of the sample and distribution by sex and children; we did not conduct an analysis of invariance measurement, so that comparisons between subgroups are inadvisable until we have sufficient information (Chen, 2007). The fifth limitation lies in that,

TABLE 6 | Evidence of convergent and divergent validity between the IBCAP-T and PFS-T factors adjusted for reliability and bias for Social Desirability ($N = 200$)[a,b].

	1 IBCAP-T	2 IBCAP-T	3 IBCAP-T	4 IBCAP-T	5 IBCAP-T	6 IBCAP-T	7 IBCAP-T
Crude correlations[c]							
1 PFS-T	−0.442**	−0.235**	−0.351**	−0.537**	−0.365**	−0.042	−0.170*
2 PFS-T	−0.349**	−0.216**	−0.261**	−0.261**	−0.228**	−0.035	−0.085
3 PFS-T	0.595**	0.385**	0.501**	0.381**	0.422**	0.298**	0.266**
4 PFS-T	0.075	0.174*	0.156*	0.153*	0.109	0.026	0.377**
Reliability-adjusted correlations[d]							
1 PFS-T	−0.477**	−0.254**	−0.383**	−0.589**	−0.406**	−0.048	−0.207**
2 PFS-T	−0.359**	−0.222**	−0.271**	−0.272**	−0.241**	−0.038	−0.098
3 PFS-T	0.641**	0.415**	0.546**	0.417**	0.468**	0.340**	0.323**
4 PFS-T	0.079	0.184**	0.167*	0.164*	0.119	0.029	0.449**
Correlations bias by Social Desirability[e,f]							
1 PFS-T	−0.446**	−0.154*	−0.329**	**−0.561****	−0.345**	−0.015	−0.111
2 PFS-T	−0.308**	−0.204**	−0.207**	−0.255**	−0.213**	−0.010	−0.096
3 PFS-T	0.546**	0.309**	0.434**	0.364**	0.383**	**0.309****	0.246**
4 PFS-T	−0.014	0.128	0.100	0.129	0.062	0.004	**0.449****
Attenuation index[g]							
1 PFS-T	7.34	7.48	8.36	8.83	10.10	12.50	17.87
2 PFS-T	2.79	2.70	3.69	4.04	5.39	7.89	13.27
3 PFS-T	7.18	7.23	8.24	8.63	9.83	12.35	17.65
4 PFS-T	5.06	5.43	6.59	6.71	8.40	10.34	16.04
Difference by bias[h]							
1 PFS-T	−0.031	−0.100	−0.054	−0.028	−0.061	−0.033	−0.096
2 PFS-T	−0.051	−0.018	−0.064	−0.017	−0.028	−0.028	−0.002
3 PFS-T	0.095	0.106	0.112	0.053	0.085	0.031	0.077
4 PFS-T	0.093	0.056	0.067	0.035	0.057	0.025	0.000

[a]The values of the correlations after being corrected for reliability and bias of social desirability and passed from $p < 0.05$ to $p > 0.05$ are underlined. [b]Correlations that increased after removing the effect of social desirability were marked in bold. [c]Spearman's Rho coefficients were calculated due to the absence of univariate normality in the total scores by factor. [d]The correlations were adjusted with the formula, $r_{true} = \frac{r_{observed}}{\sqrt[2]{r_{xx} \cdot r_{yy}}}$ where r_{xx} y r_{yy} are the GLB_{Ord} coefficients by factor, in such a way that the adjusted correlation is an estimate of the true correlation. [e]The partial correlations were worked with the factors of Self-deception and Printing Handling in such a way that the reported partial correlations are of the second order. [f]The bias was conducted on the correlations previously corrected by reliability, using in all cases the coefficient GLB_{Ord} to make the adjustment. [g]Attenuation index $= [(r_{true} - r_{observed})/r_{true}]$ (100). [h]Difference by bias $= r_{true} - r_{biased}$.
$^{*}p < 0.05$. $^{**}p < 0.01$.

although interpretability indicators were reported, it would be interesting to conduct a study to establish non-arbitrary cut points (Abad et al., 2011).

CONCLUSION

This study sought to gather solid evidence on the validity and reliability of the Spanish translated versions of the IBCAP-T and PFS-T. One solid starting point was provided for the development of tools to determine a valid and reliable way, the presence or absence of factors that may increase the likelihood of child abuse as well as those factors that can reduce its incidence.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Ethical approval was not provided for this study on human participants because Informed consent was provided by all participants. This study passed the approval of the Master's and Doctorate Program in Psychology Committee at Universidad Nacional Autónoma de México (UNAM). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AS-M and AA supported, planned, and developed the research. EC and PA supervised and advised data collection. AS-M performed the ordering, statistical analysis of the data, and wrote the first draft. SC-M and SS-C supervised the statistical analysis of data, revised, and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The present study has been supported by the Chilean national projects, Fondecyt Regular 2019 [ref. Number 1190945] and by the Operational Program ERDF Andalusia 2014–2020, the Government of Andalusia, Spain (ref. US-1263096); the VI Own Research and Transfer Plan (VIPPITUS), the Universidad de Sevilla, Spain (ref. VIPP PRECOMPETI 2020/1333); and the CONACyT National Quality Postgraduate Program (ref. Number 848825).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.705228/full#supplementary-material

REFERENCES

Abad, F., Olea, J., Ponsoda, V., and García. C. (2011). *Medición en Ciencias Sociales y de la Salud*. Madrid: Síntesis.

Agbo, A. A. (2010). Cronbach's alpha: review of limitations and associated recommendations. *J. Psychol. Afr.* 20, 233–239. doi: 10.1080/14330237.2010.10820371

Albertos, A., Osorio, A., Lopez-del Burgo, C., Carlos, S., Beltramo, C., and Trullols, F. (2016). Parental knowledge and adolescents' risk behaviors. *J. Adolesc.* 53, 231–236. doi: 10.1016/j.adolescence.2016.10.010

Aschengrau, A., and Seage, G. R. (2019). *Essentials of Epidemiology in Public Health*. Burlington, MA: Jones y Bartlett Learning.

Assink, M., Spruit, A., Schuts, M., Lindauer, R., Put, C., and Stams, G. -J. (2018). The intergenerational transmission of child maltreatment: a three-level meta-analysis. *Child Abuse Neglect* 84, 131–145. doi: 10.1016/j.chiabu.2018.07.037

Batista-Foguet, J. M., Coenders, G., and Alonso, J. (2004). Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud. *Med Clín.* 122(Suppl. 1), 21–27. doi: 10.1157/13057542

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

Chacón-Moscoso, S., Anguera, M. T., Sanduvete-Chaves, S. 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

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

Chahine, E. F. (2014). Child abuse and its relation to quality of life of male and female children. *Proc. Soc. Behav. Sci.* 159, 161–168. doi: 10.1016/j.sbspro.2014.12.350

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Eq. Model. Multidisciplin. J.* 14, 464–504. doi: 10.1080/10705510701301834

Cheung, G., and Wang, C. (2017). Current approaches for assessing convergent and discriminant validity with SEM: issues and solutions. *Acad. Manag. Proc.* 2017:12706. doi: 10.5465/AMBPP.2017.12706abstract

Cicchetti, D., and Handley, E. D. (2019). Child maltreatment and the development of substance use and disorder. *Neurobiol. Stress* 10, 1-19. doi: 10.1016/j.ynstr.2018.100144

COMPREVNNA (2017). *Alianza Global Para Poner fin a la Violencia Contra Niñas, Niños y Adolescentes, Plan de Acción de México.* Comisión para Poner Fin a Toda Forma de Violencia contra Niñas, Niños y Adolescentes. Available online at: https://www.gob.mx/cms/uploads/attachment/file/643743/PLAN_DE_ACCI_N_DE_M_XICO_2017-2018.pdf (accessed November 02, 2020).

Conrad-Hiebner, A., Schoemann, A. M., Counts, J. M., and Chang, K. (2015). The development and validation of the Spanish adaptation of the protective factors survey. *Child. Youth Serv. Rev.* 52, 45–53. doi: 10.1016/j.childyouth.2015.03.006

Contreras-Espinoza, S., and Novoa-Muñoz, F. (2018). Ventajas del alfa ordinal respecto al alfa de Cronbach ilustradas con la encuesta AUDIT-OMS. *Rev. Panam Salud Publ.* 42, 1–6. doi: 10.26633/RPSP.2018.65

Counts, J. M., Buffington, E. S., Chang-Rios, K., Rasmussen, H. N., and Preacher, K. J. (2010). The development and validation of the protective factors survey: a self-report measure of protective factors against child maltreatment. *Child Abuse Negl.* 34, 762–772. doi: 10.1016/j.chiabu.2010.03.003

Cutrona, C. E., Cole, V., Colangelo, N., Assouline, S. G., and Russell, D. W. (1994). Perceived parental social support and academic achievement: an attachment theory perspective. *J. Pers. Soc. Psychol.* 66, 369–378. doi: 10.1037/0022-3514.66.2.369

Delgado, J. (2016). El maltrato infantil por negligencia: conceptos y visión general sobre su evaluación. *I+D Rev. Investig.* 7, 14–23. doi: 10.33304/revinv.v07n1-2016002

Desjardins, C. D., and Bulut, O. (2018). *Handbook of Educational Measurement and Psychometrics Using R*. Florida, FL: CRC Press.

Domínguez-Lara, S. (2017). Atenuación de correlaciones y baja fiabilidad: 'realmente importa' *Nutr. Hosp.* 34, 1261–1262. doi: 10.20960/nh.1391

Downing, S. M., and Haladyna, T. M. (2006). *Handbook of Test Development.* Mahwah, NJ: Lawrence Erlbaum Associates.

Ellonen, N., Rantanen, H., Lepistö, S., Helminen, M., and Paavilainen, E. (2019). The use of the brief child abuse potential inventory in the general population in Finland. *Scand. J. Prim. Health Care* 37, 249–255. doi: 10.1080/02813432.2019.1571002

Escobedo-Portillo, M. T., Hernández-Gómez, J. A., Estebané-Ortega, V., and Martínez-Moreno, G. (2016). Modelos de ecuaciones estructurales: características, fases, construcción, aplicación y resultados. *Ciencia Trabajo* 18, 16–22. doi: 10.4067/S0718-24492016000100004

Fornell, C., and Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J. Market. Res.* 18, 39–50. doi: 10.1177/002224378101800104

FRIENDS National Center for Community Based Child Abuse Prevention (2018). *Protective Factors Survey, User Manual, 2nd Edn*. Retrieved from https://static1.squarespace.com/static/586370cec534a5dbdedba846/t/5c4f755b4d7a9c3614d360dd/1548711265804/PFS-2+User+Manual+10.22.18.pdf (accessed April 15, 2021).

FRIENDS National Center for Community Based Child Abuse Prevention (2020). *The Protective Factors Survey User's Manual*. Retrieved from: https://friendsnrc.org/evaluation/protective-factors-survey/ (accessed April 15, 2021).

FRIENDS National Center for Community Based Child Abuse Prevention (2021). *Protective Factors Surveys*. Retrieved from: https://friendsnrc.org/evaluation/protective-factors-survey/ (accessed April 15, 2021).

Gempp, R. (2006). El error estándar de medida y la puntuación verdadera de los tests psicológicos: Algunas recomendaciones prácticas. *Terapia Psicol.* 24,117–129.

George, D., and Mallery, P. (2017). *IBM SPSS Statistics 23 Step by Step. A Simple Guide and Reference*. New York, NY: Allyn and Bacon. doi: 10.4324/9781315545899

Hair, J., Anderson R., Tatham R., and Black W. (1999). *Análisis Multivariante*, *5a Edn*. Madrid: Prentice Hall.

Hambleton, R., and Patsula, L. (2014). Increasing the validity of adapted tests: myths to be avoided and guidelines for improving test adaptation practices. *J. Appl. Test. Technol.* 1, 1–13.

Hayes, B. E., and O'Neal, E. N. (2018). The effects of individual- and national-level factors on attitudes toward child maltreatment. *Child Abuse Neglect* 83, 83–93. doi: 10.1016/j.chiabu.2018.07.012

Heale, R., and Forbes, D. (2013). Understanding triangulation in research. *Evid. Based Nurs.* 16, 98–98. doi: 10.1136/eb-2013-101494

Henry, C. (2018). Exposure to domestic violence as abuse and neglect: constructions of child maltreatment in daily practice. *Child Abuse Neglect* 86, 79–88. doi: 10.1016/j.chiabu.2018.08.018

Hoffmann, F. A., Stover, J. B., de la Iglesia, G., and Fernández L. M. (2013). Correlaciones policóricas y tetracóricas en estudios factoriales exploratorios y confirmatorios. *Ciencias Psicol.* 7, 151–164.

Holgado-Tello, F. P., Carrasco-Ortiz, M. Á., del Barrio-Gándara, M. V., and Chacón-Moscoso, S. (2007). Factor analysis of the big five questionnaire using polychoric correlations in children. *Qual. Quant.* 43, 75–85. doi: 10.1007/s11135-007-9085-3

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., and Sanduvete-Chaves, S. (2006). Training satisfaction rating scale: development of a measurement model using polychoric correlations. *Eur. J. Psychol. Assess.* 22, 268–279. doi: 10.1027/1015-5759.22.4.268

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., and Vila-Abad, E. (2008). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Qual. Quant.* 44, 153–166. doi: 10.1007/s11135-008-9190-y

IBM Corporation (2016). *IBM SPSS Statistics (Version 24) [Software]*. Armonk, NY: International Business Machines Corporation (IBM).

INEGI (2016). *Violencia Contra Niñas, Niños y Adolescentes: Consideraciones Conceptuales, Metodológicas y Empíricas Para el Caso de México.* Instituto Nacional de Estadística y Geografía. Retrieved from: http://internet.contenidos.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825088927.pdf (accessed August 03, 2020).

International Test Commission (2017). *The ITC Guidelines for Translating and Adapting Tests, 2nd Edn*. Retrieved from: https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf (accessed June 14, 2021).

Kaufman, J., and Torbey, S. (2019). Child maltreatment and psychosis. *Neurobiol. Dis.* 131:104378. doi: 10.1016/j.nbd.2019.01.015

Kerlinger, F. N., and Lee, H. B. (2002). Investigación del Comportamiento. México, DF: McGraw-Gill.

Kessler, R. C., McLaughlin, K. A., Green, J. G., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., et al. (2010). Childhood adversities and adult psychopathology in the WHO World Mental Health Surveys. *Br. J. Psychiatry* 197, 378–385. doi: 10.1192/bjp.bp.110.080499

Kimchi, J., Polivka, B., and Stevenson, J. S. (1991). Triangulation. *Nurs. Res.* 40, 364–366. doi: 10.1097/00006199-199111000-00009

Kline, R. (2011). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Press.

Labella, M. H., and Masten, A. S. (2018). Family influences on the development of aggression and violence. *Curr. Opin. Psychol.* 19, 11–16. doi: 10.1016/j.copsyc.2017.03.028

Lawson, J. (2019). Domestic violence as child maltreatment: differential risks and outcomes among cases referred to child welfare agencies for domestic violence exposure. *Child Youth Serv. Rev.* 98, 32–41. doi: 10.1016/j.childyouth.2018.12.017

Letourneau, N., and Allen, M. (1999). Post-positivistic critical multiplism: a beginning dialogue. *J. Adv. Nurs.* 30, 623–630. doi: 10.1046/j.1365-2648.1999.01133.x

Liel, C., Meinck, F., Steinert, J. I., Kindler, H., Lang, K., and Eickhorst, A. (2019). Is the Brief Child Abuse Potential Inventory (BCAPI) a valid measure of child abuse potential among mothers and fathers of young children in Germany? *Child Abuse Neglect* 88, 432–444. doi: 10.1016/j.chiabu.2018.11.008

Marco, M., Gracia, E., López-Quílez, A., and Freisthler, B. (2019). Child maltreatment and alcohol outlets in Spain: does the country drinking culture matters?. *Child Abuse Neglect* 91, 23–30. doi: 10.1016/j.chiabu.2019.02.010

McCoy, M. L., and Keen, S. M. (2014). *Child Abuse and Neglect.* New York, NY: Psychology Press.

Microsoft Corporation (2016). *Microsoft Excel Profesional Plus (Versión 16.0) [Software]*. Redmond, Washington: Microsoft Corporation.

Mikulic, I. M., Crespi, M., and Caballero, R. (2016). Estudio psicométrico del inventario balanceado de respuesta deseable. *Anuar Psicol.* 46, 58–66. doi: 10.1016/j.anpsic.2016.07.002

Milner, J. (1986). *The Child Abuse Potential Inventory Manual, 2nd Edn.* Webster, NC: Psytec.

Moral de la Rubia, J., García-Cadena, C. H., and Antona-Casas, C. J. (2012). Traducción y validación del Inventario Balanceado de Deseabilidad Social al responder en una muestra probabilística de estudiantes universitarios mexicanos. *Rev. Psicol. GEPU* 3, 54–72. doi: 10.22199/S07187475.2012.0003.00006

Morrongiello, B. A., Hou, S., and Cox, A. (2019). Keeping pre-adolescents safe: parental knowledge and safety promotion strategies. *J. Appl. Dev. Psychol.* 62, 59–65. doi: 10.1016/j.appdev.2019.01.002

Muñiz, J., Elosua, P., and Hambleton, R. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema* 25, 151–157. doi: 10.7334/psicothema2013.24

Musitu, G., and Callejas, J. (2017). El modelo de estrés familiar en la adolescencia: MEFAD. Revista INFAD de Psicología. *Int. J. Dev. Educ. Psychol.* 1, 11-20. doi: 10.17060/ijodaep.2017.n1.v1.894

OMS, (2016). *Página Informativa Sobre Maltrato Infantil*. Organización Mundial de la Salud. Retrieved from: https://www.who.int/es/news-room/fact-sheets/detail/child-maltreatment (accessed August 10, 2020).

Ondersma, S. J., Chaffin, M. J., Mullins, S. M., and LeBreton, J. M. (2005). A brief form of the child abuse potential inventory: development and validation. *J. Clin. Child Adolesc. Psychol.* 34, 301–311. doi: 10.1207/s15374424jccp3402_9

Organización Panamericana de la Salud (OPS) Oficina Regional para las Américas de la Organización Mundial de la Salud (OMS) (2003). *Informe Mundial Sobre la Violencia y la Salud.* Retrieved from: https://iris.paho.org/bitstream/handle/10665.2/725/9275315884.pdf (accessed August 10, 2020).

Piko, B. (2000). Perceived social support from parents and peers: Which is the stronger predictor of adolescent substance use? *Substance Use y Misuse* 35, 617–630. doi: 10.3109/10826080009147475

Porras, C. J. C. (2016). Comparación de pruebas de normalidad multivariada. *Anal. Cient.* 77, 141–146. doi: 10.21704/ac.v77i2.483

R Core Team (2020). *R: A Language and Environment for Statistical Computing [Software]*. R Foundation for Statistical Computing. Retrieved from: http://www.R-project.org/ (accessed April 01, 2021).

Schwab-Reese, L. M., Hovdestad, W., Tonmyr, L., and Fluke, J. (2018). The potential use of social media and other internet-related data and communications for child maltreatment surveillance and epidemiological research: Scoping review and recommendations. *Child Abuse Neglect*. 85, 187–201. doi: 10.1016/j.chiabu.2018.01.014

Shadish, W. R. (1993). Critical multiplism: a research strategy and its attendant tactics. *New Direct. Progr. Eval.* 60, 13–57. doi: 10.1002/ev.1660

Shi, D., DiStefano, C., McDaniel, H. L., and Jiang, Z. (2018). Examining chi-square test statistics under conditions of large model size and ordinal data. *Struct. Eq. Model. Multidisciplin. J.* 25, 924–945. doi: 10.1080/10705511.2018.1449653

Sigad, L. I., Beker, G., Lev-Wiesel, R., and Eisikovits, Z. (2019). "Alone with our interpretations": uncertainty in child abuse intervention among healthcare professionals. *Child. Youth Serv. Rev*. 100, 206–213. doi: 10.1016/j.childyouth.2019.02.027

Sijtsma, K. (2011). Future of psychometrics: ask what psychometrics can do for psychology. *Psychometrika* 77, 4–20. doi: 10.1007/s11336-011-9242-4

Sprague-Jones, J., Counts, J., Rousseau, M., and Firman, C. (2019). The development of the protective factors survey, 2nd edition: a self-report measure of protective factors against child maltreatment. *Child Abuse Negl.* 89, 122–134. doi: 10.1016/j.chiabu.2019.01.008

Stoltenborgh, M., Bakermans-Kranenburg, M. J., Alink, L. R. A., and van IJzendoorn, M. H. (2012a). The universality of childhood emotional abuse: a meta-analysis of worldwide prevalence. *J. Aggress. Maltreat. Trauma* 21, 870–890. doi: 10.1080/10926771.2012.708014

Stoltenborgh, M., Bakermans-Kranenburg, M. J., Alink, L. R. A., and van IJzendoorn, M. H. (2014). The prevalence of child maltreatment across the globe: review of a series of meta-analyses. *Child Abuse Rev.* 24, 37–50. doi: 10.1002/car.2353

Stoltenborgh, M., Bakermans-Kranenburg, M. J., and van IJzendoorn, M. H. (2012b). The neglect of child neglect: a meta-analytic review of the prevalence of neglect. *Soc. Psychiatry Psychiatr. Epidemiol.* 48, 345–355. doi: 10.1007/s00127-012-0549-y

Stoltenborgh, M., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., and Alink, L. R. A. (2013). Cultural–geographical differences in the occurrence of child physical abuse? A meta-analysis of global prevalence. *Int. J. Psychol.* 48, 81–94. doi: 10.1080/00207594.2012.697165

Stoltenborgh, M., van IJzendoorn, M. H., Euser, E. M., and Bakermans-Kranenburg, M. J. (2011). A global perspective on child sexual abuse: meta-analysis of prevalence around the world. *Child Maltreat.* 16, 79–101. doi: 10.1177/1077559511403920

Thornock, C. M., Nelson, L. J., Porter, C. L., and Evans, C. A. (2019). There's no place like home: the associations between residential attributes and family functioning. *J. Environ. Psychol.* 64, 39–47. doi: 10.1016/j.jenvp.2019.04.011

Trizano-Hermosilla, I., and Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: congeneric and asymmetrical measurements. *Front. Psychol.* 7:769. doi: 10.3389/fpsyg.2016.00769

Weibela, S., Vidal, S., Olié, E., Hasler, R., Torriani, C., Prada, P., et al. (2017). Impact of child maltreatment on meaning in life in psychiatric patients. *Psychiatry Res.* 251, 204–211. doi: 10.1016/j.psychres.2017.02.026

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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 17 August 2021
doi: 10.3389/fpsyg.2021.703580

Gender-Based Performance in Mathematical Facts and Calculations in Two Elementary School Samples From Chile and Spain: An Exploratory Study

Violeta Pina[1]*, *Diana Martella*[2]*, *Salvador Chacón-Moscoso*[3,4], *Mahia Saracostti*[5,6] *and Javier Fenollar-Cortés*[7]

[1] *Departamento de Psicología Evolutiva y de la Educación, Facultad de Educación, Economía y Tecnología de Ceuta, Universidad de Granada, Ceuta, Spain,* [2] *Instituto de Estudios Sociales y Humanísticos, Facultad de Ciencias Sociales y Humanidades, Universidad Autónoma de Chile, Región Metropolitana, Chile,* [3] *Departamento de Psicología Experimental, Universidad de Sevilla, Sevilla, Spain,* [4] *Departamento de Psicología, Universidad Autónoma de Chile, Santiago, Chile,* [5] *Núcleo Científico y Tecnológico en Ciencias Sociales y Humanidades, Universidad de la Frontera, Temuco, Chile,* [6] *Escuela de Trabajo Social, Facultad de Ciencias Sociales, Universidad de Valparaíso, Valparaíso, Chile,* [7] *Departamento de Psicología, Universidad Loyola Andalucía, Sevilla, Spain*

OPEN ACCESS

Edited by:
Pietro Cipresso, University of Turin, Italy

Reviewed by:
Isabel Ramirez, National University of Distance Education (UNED), Spain
Begoña Delgado, National University of Distance Education (UNED), Spain

***Correspondence:**
Violeta Pina
violetapina@ugr.es
Diana Martella
diana.martella@uautonoma.cl

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

Received: *30 April 2021*
Accepted: *08 July 2021*
Published: *17 August 2021*

Citation:
Pina V, Martella D, Chacón-Moscoso S, Saracostti M and Fenollar-Cortés J (2021) Gender-Based Performance in Mathematical Facts and Calculations in Two Elementary School Samples From Chile and Spain: An Exploratory Study. Front. Psychol. 12:703580. doi: 10.3389/fpsyg.2021.703580

Gender differences in mathematical performance are not conclusive according to the scientific literature, although such differences are supported by international studies such as the Trends in International Mathematics and Science Study (TIMSS). According to TIMSS 2019, fourth-grade male students outperformed female students in Spanish-speaking countries, among others. This work approaches the study on gender difference by examining the basic calculation skills needed to handle more complex problems. Two international samples of second and third graders from Chile and Spain were selected for this exploratory study. Tests on basic mathematical knowledge (symbolic and non-symbolic magnitude comparisons, fluency, and calculation) were administered. The tests did not show significant difference or size effect between genders for mean performance, variance in the distribution of performance, or percentiles. As noted in the existing literature on this topic and reiterated by these findings, great care should be exercised when reporting on possible gender differences in mathematical performance, as these can contribute to low self-concept among female students.

Keywords: gender differences, primary education, mathematical fluency, calculation, children

INTRODUCTION

As technology grows by leaps and bounds, new fields of study and analysis tools are also expanding: Big Data, artificial intelligence, modeling engineering, software architecture, etc. This has increased the demand for professionals, both men and women, with the solid knowledge of mathematics needed for such roles (Belloum et al., 2019). However, women continue to be underrepresented in this kind of professional roles but overrepresented in lower paying jobs (Adams et al., 2019). Determining the origin of these differences is challenging as cognitive, social, and cultural factors over the course of one's life may all contribute, making it difficult to connect mathematical abilities during childhood to gender differences on the job market. However, if this gender gap were intrinsic, it could appear during childhood, making it essential to determine whether this is the case and, thus, intervene to avoid future underrepresentation of women.

The reduced presence of women in mathematical fields is nothing new. As indicated by Kane and Mertz (2012, p. 10), differences between males and females have been found in the mathematics participation rate, mean and high-end performance, and variance in the distribution of performance, making gender difference the subject of extensive debate. Havelock Ellis (1894) was the first to intend to empirically justify the differences between men and women with his variability hypothesis. According to this hypothesis, men could be expected to show greater variability than women in physical and intellectual abilities. This would explain the predominance of men in higher academic and scientific positions, but also their prevalence in crime. This hypothesis coincided with the expanded presence of women in higher education and their demands for recognition (Shields, 1982) and continued to be defended, with minor modifications, until the 21st century, when the focus turned to extreme scores and percentiles. In particular, some studies found more females in the lower percentiles and more males in the higher percentiles for different math assignments (Barbaresi et al., 2005; Reigosa-Crespo et al., 2012; Anaya et al., 2017).

Nevertheless, in a wide study carried out by Baker and Jones (1993) with 77,602 students from 19 countries, mathematical performance between genders was found to vary by country and the type of education students received; additionally, the differences were noted to be decreasing over time. The authors also largely debunked the hypothesis of variability when they pointed out that gender differences in performance decrease when women have more equitable access to higher education and qualified work. Indeed, several studies (for instance, Kane and Mertz, 2012) have moved away from the deterministic hypothesis of variability, providing empirical evidence for a hypothesis of sociocultural factors.

Coinciding with the argument that sociocultural factors have the greatest bearing on mathematical performance, differences between genders should gradually disappear as society becomes more progressive and inclusive. Lindberg et al. (2010) support this hypothesis in a broad meta-analysis. Based on 242 studies published between 1999 and 2007 (a total sample of 1,286,350 individuals both males and females), the authors concluded that mathematical performance is similar for both genders, noting that any gender difference can be attributed to unrelated factors. The authors conclude that it is crucial to disseminate these results to counteract stereotypes about [alleged] female math inferiority (Lindberg et al., 2010, p. 1134). More recent studies have replicated these results (for instance, Voyer and Voyer, 2014; Scheiber et al., 2015; Kersey and Cantlon, 2018), and others have limited the differences to a minority of tasks (Hutchison et al., 2019).

Given the empirical evidence reported in the literature, males and females would be expected to perform similarly on international studies such as the TIMSS (Instituto Nacional de Evaluación Educativa [INEE], 2016, 2020). Unfortunately, this is not the case. According to the TIMSS, the gender gap in mathematical performance is present as early as fourth grade and, according to the last edition (2019), has even increased in countries such as Spain. Gender differences at the fourth grade level can also be noted, although to a lesser extent, in other Spanish-speaking countries that participated in TIMSS 2019. Chile, for example, went from having no significant gender differences in mathematics in the 2015 report to having significant ones in the 2019 report (Instituto Nacional de Evaluación Educativa [INEE], 2016, 2020). In fact, samples from Spanish-speaking countries show a large gender gap in international reports, but few studies include them in their analysis (e.g., the meta-analysis by Voyer and Voyer, 2014).

This exploratory study considers two Spanish-speaking samples from Chile and Spain, both of which are underrepresented in the literature. First, it is important to analyze potential gender differences for these countries. Second, by looking into the samples from these two countries, common patterns may be found that could allow an early intervention to be designed.

Both Spain and Chile have similar math curricula at the elementary school level. In both Chile (MINEDUC, 2018) and Spain (e.g., Decreto n.°, 198/2014), students are doing single-digit addition and subtraction at the end of the first grade (at around age seven). In second grade, they learn multiplication and begin adding and subtracting beyond single digits. In third grade, they begin using multiplication and division to solve math problems. Although both samples are socially and culturally distinct, the objective of this study was to investigate whether gender differences exist and, if so, whether the pattern is the same for the two Spanish-speaking countries. Since both Chile and Spain have similar math curricula, if this was the case, the gender gap could be considered intrinsic.

International evaluations and even math tests administered at schools often do not consider the skills and basic knowledge for acquiring mathematical abilities, and are instead focused on complex calculation or problem solving (Nosworthy et al., 2013). Hence, when gender differences are observed in such tests, it is hard to determine whether such differences are also present for more basic math skills. A logical hypothesis is that the basic abilities needed for complex tasks would also show a gender gap. In recent years, different studies have explored what basic numerical skills necessary for mathematical achievement are responsible for gender differences (for instance, Kersey and Cantlon, 2018; Hutchison et al., 2019).

In this regard, the literature has widely reported on the basic skills required for mathematical achievement. On one hand, non-symbolic magnitude comparison (e.g., the ability to look at two groups of objects and determine which is greater in number), is considered a stepping stone to learning numbers (see Landerl, 2019). For non-symbolic comparison, visuospatial perception, not the ability to count, is required (Kersey and Cantlon, 2018). Once it is acquired, boys and girls can relate quantities with symbolic numbers (for instance, relating the number four with four objects) and learn the meaning of the Arabic numeral. On the other hand, symbolic magnitude comparison (the ability to choose which of two numbers is larger) requires participants to efficiently access the analog magnitude representations that correspond to the Arabic numerals in order to determine which number is greater (Landerl, 2019). Two tasks are widely used to evaluate these abilities: the symbolic magnitude comparison task and the non-symbolic magnitude comparison task. A recent meta-analysis has confirmed that both tasks correlate with subsequent mathematical performance, although the correlation is higher for the symbolic comparison task (Schneider et al., 2016).

Once children understand the meaning of the Arabic numerals, they start to count. The first operations are simple single-digit addition and subtraction with a sum of <10. Depending on one's knowledge of arithmetic facts, two strategies are mainly used. In the beginning, the child will have to count to solve the operation; and through repetition, the answers become lodged in their memory (De Smedt et al., 2019). The automation of these operations has been considered critical to developing robust complex calculations (Royer et al., 1999).

Calculation and mathematical problems are considered more complex areas of mathematical performance. Mathematical problems also require other abilities such as reading comprehension (Abedi and Lord, 2001; Donlan et al., 2007). Calculation usually includes the four basic mathematical operations (addition, subtraction, multiplication, and division) and generally cannot be solved from memory, since it usually involves double digits or more. Children can use a wide variety of strategies for approaching these operations, such as decomposition or sequencing (Hickendorff et al., 2019). Calculation also requires executive functions, such as working memory. The strategy children use depends not only on their individual skills but also on the instruction they have received. Gender differences have been found (e.g., Diamantopoulou et al., 2012) for these complex tasks. Winkelmann et al. (2008) proposed that disparities in different studies on gender differences can be attributed to what is considered "basic ability" in a particular study and on the cognitive resources required for each task. The breakdown of tasks by gender differences allows for the design of learning and support strategies in specific areas to gradually reduce the gender gap.

Gender differences tend to be more evident in upper elementary. Some studies do present evidence of a gender gap for lower elementary (Jordan et al., 2006; Scheiber et al., 2015), but others do not (Lachance and Mazzocco, 2006). Third grade seems to be a critical year for the detection of gender differences in mathematical performance, although the studies at this level are scarce. As an example, Germany presents gender differences starting in third grade (Winkelmann et al., 2008). In Spain, the TIMSS (2015 and 2019) reports a large gender gap in fourth grade, thus suggesting a possible gender gap in third grade as well. Another international study, TERCE 2015 (UNESCO, 2016), shows that the number of Latin American countries where males have an advantage over females increases between third and sixth grade. Therefore, second and third grade are considered crucial. The aim of this study is not to determine the age at which gender differences, if they exist, appear, but to observe whether these differences are detected in lower elementary in Spanish-speaking countries.

This study delves into these differences to determine if they are present in the first years of school for basic math skills. After controlling for reading proficiency, mathematical performance by Spanish-speaking second and third graders in Chile and Spain will be compared for non-symbolic comparison, symbolic comparison, fluency, and calculation. As part of this study, any differences in mean performance, variance in the distribution of performance, and percentiles will be assessed. According to the literature, differences are not expected in basic mathematical tasks such as symbolic and non-symbolic comparison but could appear for tasks that require mathematical fluency and calculation. We hypothesize that a greater number of males will be in the higher percentiles and a greater number of females will be in the lower percentiles.

METHODOLOGY

Participants

The participants were second and third graders at four schools in the O'Higgins region (Chile) and seven schools in Murcia (Spain). These children were organized into two separate samples, one for each country, with a mean age of 7.75 for the second graders ($M = 7.68$, $SD = 0.31$ for the Spanish children, and $M = 7.88$, $SD = 0.48$ for the Chilean children) and 8.77 for the third graders ($M = 8.7$, $SD = 0.33$ for the Spanish children, and $M = 8.89$, $SD = 0.39$ for the Chilean children). The sample was formed by 201 Spanish school children (107 second graders and 94 third graders; 55.2% male and 44.8% female) and 184 Chilean school children (90 second graders and 94 third graders; 48.9% male and 51.1% female) and was collected through incidental sampling. The following exclusion criteria were established: vision difficulties, diagnosis of autism spectrum disorder or cognitive impairment, insufficient knowledge of the Spanish language, and, in the case of the third graders, significant difficulties reading. **Table 1** details the number of males and females by grade and country. Possible differences regarding age (Mann–Whitney U) and gender proportion by grade and country (χ^2) were examined.

This study was carried out following the recommendations of the Chilean Commission for Scientific and Technological Investigation (CONICYT in its Spanish acronym). The protocol was approved by the Ethics Committee of Universidad de la Frontera (Act 066-2017, on Sheet 036-17). The study complied with the standards of the ethical committee of Universidad Autónoma de Chile and with the agreement between the Department of Education and Universities of the autonomous community of Region of Murcia and Universidad de Murcia. Informed consent was requested to participate in this study.

Instruments

The following paper-and-pencil tests were administered to measure mathematical performance and reading fluency.

Numerical magnitude comparisons (Nosworthy et al., 2013). This task includes two different parts:

TABLE 1 | Distribution in the Spanish language sample of second and third graders: scoring on symbolic test by gender.

	Spanish sample					Chilean sample				
M(DT)	**Male**	**Female**	***U***	***p***	***r*[a]**	**Male**	**Female**	***U***	***p***	***r***
Second grade										
N	107					90				
Gender *n (%)*	56(52.3)	51(47.7)				49(54.4)	41(45.6)			
Magnitude comparison										
Symbolic	29.3(9.3)	33.0(6.8)	1055.0	0.020	−0.26	29.9(10.1)	30.9(10.6)	961.0	0.727	
Num. errors	0.1(0.4)	0.1(0.9)	1572.2	0.060		0.6(1.3)	0.7(1.2)	931.5	0.850	
Non-symbolic	29.1(7.6)	31.6(5.7)	1103.5	0.042	−0.23	28.8(13.8)	33.9(12.0)	796.5	0.092	
Num. errors	0.3(1.0)	0.6(0.1)	1419.5	0.932		1.4(2.7)	2.2(2.9)	826.0	0.110	
Third grade										
N	94					94				
Gender *n (%)*	55(58.5)	39(41.5)				55(58.5)	39(41.5)			
Magnitude comparison										
Symbolic	37.7(7.0)	39.8(6.2)	837.0	0.069		39.3(10.2)	37.4(11.2)	1158.0	0.513	
Num. errors	0.2(0.6)	0.1(0.4)	1126.5	0.443		0.6(1.1)	0.6(1.1)	1051.5	0.847	
Non-symbolic	32.3(7.3)	31.7(4.8)	1196.0	0.344		37.0(11.3)	35.5(11.4)	1193.5	0.355	
Num. errors	0.5(1.0)	0.3(0.6)	1135.5	0.412		1.7(2.7)	1.1(1.6)	1133.5	0.618	
PreDisCal										
Sentences	13.6(14.7)	14.7(2.9)	905.5	0.199		13.2(6.9)	12.8(6.2)	1090.0	0.896	
Num. errors	2.2(1.9)	1.6(1.2)	1237.0	0.193		6.5(9.3)	3.2(2.4)	1252.5	0.163	
Mathematical fluency	17.7(6.8)	17.1(4.5)	1079.5	0.960		12.9(5.5)	10.7(5.1)	1369.5	0.023	0.27
Num. errors	1.5(2.5)	1.1(1.7)	1127.0	0.652		1.8(2.6)	1.3(1.4)	1050.5	0.864	
Calculation	11.6(4.2)	11.0(3.4)	1143.0	0.589		9.0(4.3)	9.0(4.3)	1126.0	0.683	
Num. errors	0.2(0.6)	0.3(0.6)	1202.0	0.313		4.1(8.0)	1.9(8.0)	1294.5	0.083	

Scores are shown in terms of mean and standard deviation (not mean scores) to facilitate understanding.
[a]*Effect size corresponding to the biserial correlation between ranges. Only values corresponding to significant differences between the groups are provided.*

- Symbolic task (56-digit pairs), where the participants were asked to compare numerical pairs (numbers one through nine) and indicate which number was higher.
- Non-symbolic task (56 dot arrays), where dot arrays of varying quantities are present and the participants were asked to indicate which side had more dots.

Each test had a time limit of 1 min, and the dependent variable was the number of correct answers.

PreDisCal (Pina et al., 2020). This is a set of three tests that measure reading fluency, mathematical fluency, and calculation, in that order. Although the PreDisCal scale was developed in Spain, a cross-validation process was carried out in several Latin American countries. Some of the items were modified because of interpretation issues (related to sociocultural factors) in Latin American countries. The test included the following tasks:

- Sentences. This task consists of 47 sentences that assess reading fluency. Each sentence has a missing word, for instance: "The strawberries are …" followed by five answer options (one is correct). The incorrect alternatives are close grammatically or semantically. All the sentences are easy to understand. The test had a time limit of 3 min. The dependent variable is the number of correct answers. It is a validated test with adequate test-retest reliability [$r(169) = 0.8, p < 0.001$].
- Mathematical fluency. The arithmetic facts are tested through 63 single-digit addition and subtraction operations. All sums are <10. The maximum time is 1 min in this test, and the dependent variable is the number of correct answers. The reliability of the test is adequate [$r(169) = 0.85, p < 0.001$].
- Calculation. This test comprises 45 items of increasing difficulty and assesses complex calculation. Participants have to determine what number or symbol, or if both are missing in a comparison of two operations (e.g., "$3 + \_ = 5 + 1$"). The test had a time limit of 3 min, and the dependent variable was the number of correct answers. The test-retest reliability of the validated scale was acceptable [$r(169) = 0.75, p < 0.001$].

Procedure

The tasks were administered in February for Spain and in September of the same year for Chile in order to reflect the same moment of the academic calendar in both countries. Trained evaluators collectively administered the test during the school day. Given that PreDisCal required a certain degree of reading proficiency and complex calculation ability, it was administered only to the third graders, while the magnitudes comparison tests were administered to both the second and third grade samples. In the case of the magnitude comparisons, the symbolic test was administered first, followed by the non-symbolic test.

The PreDisCal was administered in keeping with the order established for that test, e.g., sentences, mathematical fluency, and calculation.

Data Analysis

First, the comparability of the groups was verified by assessing age and gender differences by grade and country (χ^2), and mean performance in the sentences test between countries (Student t). Second, basic parametric assumptions for data analysis were verified (normality using the Kolmogorov–Smirnov test, homoscedasticity using the Levene test). Non-parametric tests were carried out using the Mann–Whitney U and statistical biserial correlations ($r = 0.1$ as low; $r = 0.3$ as medium, and $r = 0.5$ as high effect; Cohen, 1988); and χ^2 to compare gender and percentile performance distribution in third grade.

RESULTS

Both samples showed a similar gender distribution of participants ($\chi^2 = 0.09$, $p = 0.768$ for second grade; $\chi^2 = 0$, $p = 1$ for second grade). There was no significant evidence between countries in relation to the mean performance in the sentences test ($t = 1.35$; $p = 0.178$). However, significant differences were detected between countries regarding the mean age of students ($t = -5.04$, $p < 0.001$, $d = -0.52$ for the third grade; $t = -4.18$, $p < 0.001$, $d = -0.86$ for the second grade), with moderate to high effect sizes (Cohen, 1988). Nevertheless, these differences in age are not relevant to this study, since the analysis between countries is carried out comparing gender, a variable for which there are no significant differences in the samples of both countries.

The distribution of data was examined for each dependent variable according to country and course. Except for the mathematical fluency and calculation tests in the Chilean third grade sample, the remaining variables followed a non-normal distribution (K-S < 0.05); thus, non-parametric analyses were applied.

The Levene test showed significant difference in variance between genders in the symbolic test for the Spanish sample of second graders ($F = 3.97$, $p = 0.049$). The scores of boys showed higher variability ($\sigma^2 = 86.5$) with skewness of 0.18, ($SE = 0.32$) and kurtosis of -0.19 ($SE = 0.63$), compared with those of the girls ($\sigma^2 = 46.3$) with skewness of 0.1 ($SE = 0.33$), and kurtosis of 0.27 ($SE = 0.66$), as shown in **Figure 1**. Differences in variance were not found for the other tests for either sample.

Spearman's correlation was carried out between tests for each country (**Table 2**). The correlation between scores on the symbolic and non-symbolic tests was crucially positive and high in the second grade sample. As for the third graders, mathematical fluency positively correlated to calculation and with sentences; these last two subscales correlated positively. The symbolic test also correlated positively with the non-symbolic one ($r_s = 0.46$, p < 0.001).

Mean Performance by Gender and Grade

If differentiated by gender, the Mann–Whitney tests indicated that the girls scored significantly higher (See **Table 2**) in the symbolic test of the Spanish second graders ($U = 1{,}055$, $p = 0.02$, $r = -0.26$), but this was not the case for the Chilean sample ($p = 0.727$). However, the effect size of the significant differences ranged from low to moderate. Regarding the non-symbolic test, the results were similar, with girls outperforming the boys in the Spanish sample ($U = 1{,}103$, $p = 0.42$, $r = -0.23$), but not in the Chilean sample ($p = 0.092$). In this case, the effect size was lower than in the symbolic test. The differences detected in the sample disappear in the third grade. In terms of the number of mistakes made on both tests, no significant differences were found between genders in either country.

In the third-grade sample, significant differences were only found between genders for the mathematical fluency test in the Chilean sample ($U = 1{,}369$, $p = 0.023$, $r = 0.27$). Here, the girls had lower scores ($Md = 9$) than the boys ($Md = 13$). In terms of the number of mistakes made on both tests, no significant differences were found between genders in either country at the third grade level.

Ratio of Boys to Girls at Various Percentiles

Four groups for each sample were formed according to the scores they obtained and their corresponding percentiles. Group 1 scored in the 1–25th percentile; group 2 in the 26–50th percentile; group 3 in the 51–75th percentile; and group 4 in the 76th percentile or higher. χ^2 and its significance were calculated to examine significant differences in the ratio of males and females in the different percentile groups. As no significant differences were found in terms of performance of boys and girls in the third-grade tests (except for mathematical fluency, which had a low effect size), the groups were formed by students from both countries. In the case of the second graders, the calculations varied by country.

For the third group, there were no significant gender differences in terms of the ratio of boys to girls, the different percentile groups, or for any of the tests (see **Table 3**). However, significant differences were found in the symbolic test for the Spanish sample of second graders ($\chi^2 = 9.65$, $p = 0.022$). The percentages were different for the first group (scorings between the 1–25th percentile; $\chi^2 = 6.54$, $p = 0.011$), with a larger quantity of boys (30.4%) than girls (9.8%). Significant differences were not found in the Chilean sample of second graders.

DISCUSSION

This study set out to compare the basic mathematical abilities of boys and girls, which prove essential for the acquisition of complex math skills. The sample was composed of boys and girls from two Spanish-speaking yet culturally different countries in the second and third grade, ages at which the gender gap starts to appear. The results allow us to conclude, in keeping with the scientific literature on the topic, that in terms of basic mathematical knowledge, boys and girls show no significant differences in mathematical performance. Previous investigations, including various meta-analyses (Lachance and Mazzocco, 2006; Hyde et al., 2008; Lindberg et al., 2010; Voyer and Voyer, 2014; Scheiber et al., 2015), have noted as much. Similarly, Hutchison et al. (2019) found no differences in numerical and magnitude comparison, numerical ordering,

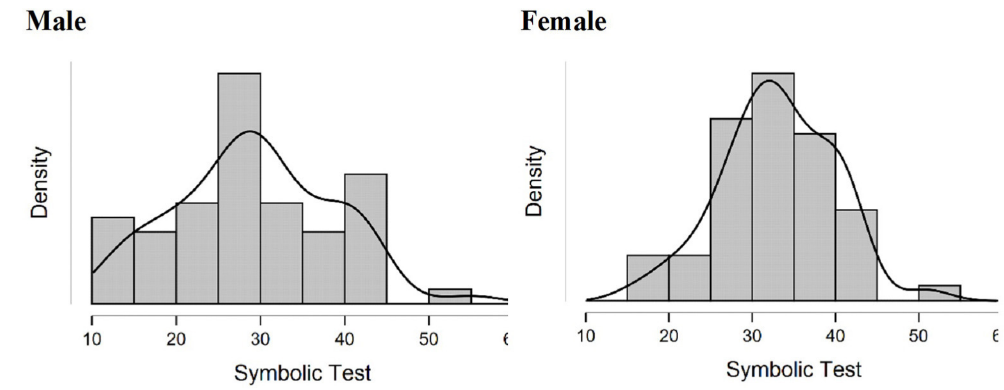

FIGURE 1 | Distribution in the Spanish language sample of third grader scores on the symbolic test by gender.

TABLE 2 | Spearman's correlations between the PreDisCal and magnitude comparison (third grade tests).

	PreDisCal			Magnitude comparison
	Sentences	**Mathematical fluency**	**Calculation**	**Symbolic**
PreDisCal				
Mathematical fluency	0.194			
Calculation	0.253*	0.451***		
Comparison of magnitude				
Symbolic	0.054	0.389***	0.065	
Non-symbolic	0.025	0.271**	0.314**	0.318**

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

magnitude estimation, multiplication, and division, although the population analyzed for their study hailed from the Netherlands, meaning that results could be different in countries with a higher gender gap. The results of this study, however, reveal no significant differences in most of the tasks considered for two Spanish-speaking samples from Chile and Spain, despite the high gender gap detected in international tests for the two countries (such as TIMSS 2019).

The main differences are found in the symbolic test (comparison of two numbers to determine which one is higher) in the Spanish sample of second graders. In particular, the girls performed better and the boys showed a greater variability. Additionally, females had a higher mean performance, and a greater ratio of males fell below the 25th percentile. However, the effect size is small in second grade, and no such results were found in the third grade sample. For the non-symbolic comparison test, the girls performed better once again, although this difference disappeared in third grade. The direction of these gender differences goes against the hypothesis of higher performance in males. In line with these results, previous studies have shown gender differences in favor of females in basic abilities (for instance, Halpern and Wright, 1996).

The only difference in favor of males was found in the Chilean sample in average performance for mathematical fluency in the third grade. Differences in addition and subtraction have been found previously, although there has not always been evidence of these differences (Hutchison et al., 2019). In this case, the effect size was small and no differences were found for calculation, meaning that caution is critical when interpreting the results. In the case of the Chilean third graders, test abilities were recently acquired, and the differences would likely disappear with practice. Importantly, mathematical fluency is decisive for the robust development of complex calculation and is an ability that can be practiced until it is learned (Royer et al., 1999).

Prior studies have indicated that differences in mathematical performance rely on the kind of task students are asked to complete. These studies have found that females perform better in arithmetic and calculus, while males perform better in mathematical problem solving (Byrnes and Takahira, 1993). However, there have also been conflicting results in the research. For instance, Royer et al. (1999) identified that males perform better in math-fact retrieval than females. In another study, Winkelmann et al. (2008) showed that females had poorer basic skills, while males had a slight advantage in mathematical problems, such as equations containing a missing number. The authors explained that the results depended on basic elementary skills and on how math problems are defined, i.e., on the cognitive resource that each task demanded.

One hypothesis is that gender differences can be found in complex areas of mathematics that demand more cognitive resources. In this sense, for example, the literature suggests a relationship between spatial abilities and gender differences

TABLE 3 | Percentage and comparison of third grade boys and girls by groups corresponding to the 25th, 50th, 75th, and 100th percentile.

	Group 1(P25)	Group 2 (P50)	Group 3 (P75)	Group 4 (P100)	Differences between groups	
					χ^2	*P*
PreDisCal						
Sentences						
Male	30,0	20,9	29,1	20,0	3.80	0.284
Female	21,8	29,5	23,1	25,6		
N	61	44	39	44		
Calculation						
Male	32,7	20,0	23,6	23,6	2.39	0.495
Female	32,1	28,2	16,7	23,1		
N	61	44	39	44		
Mathematical fluency						
Male	24,5	29,1	22,7	23,6	1.60	0.659
Female	32,1	29,5	19,2	19,2		
N	52	55	40	41		
Magnitude comparison						
Symbolic						
Male	27,3	32,7	17,3	22,7	7.10	0.069
Female	24,4	21,8	33,3	20,5		
N	61	44	39	44		
Non-symbolic						
Male	27,3	18,2	30,9	23,6	5.65	0.130
Female	26,9	32,1	25,6	15,4		
N	61	44	39	44		

Groups are organized by percentiles: 1–25th percentile (Group 1), 26–50th percentile (Group 2), 51–75th percentile (Group 3), and 76–100th percentile (Group 4).

in mathematical performance. However, a recent study did not confirm these results and instead concluded that gender differences may depend on the test and the strategy used to solve each item (Ramírez-Uclés and Ramírez-Uclés, 2020). The results support this hypothesis. In particular, the calculation is the task that involves the highest cognitive demand, but no gender differences were found. Another possibility is that gender differences in mathematical performance emerge later in more complex mathematical tasks or are influenced by cultural or social stereotypes, although the results show that at least lower elementary school children do not exhibit such differences.

Another hypothesis for gender differences in mathematical performance is related to the way in which the answer options are presented on tests. For instance, in a study with PISA 2000 data, Lafontaine and Monseur (2009) found that in all the countries analyzed, open-ended (as opposed to multiple choice) answers for reading tasks favored females. For mathematics, Routitsky and Turner supported these conclusions using PISA 2003 data but pointed out that this difference disappeared as the complexity of items increased (Routitsky and Turner, 2003). Future investigations should assess how the questions and type of answer options influence the gender gap.

Previous studies have revealed that the learning routine shows no differences between genders, although boys and girls do express different attitudes toward mathematics (Barbero-García et al., 2007). This could explain why differences can be found in higher grades despite being absent when basic math skills are evaluated. Importantly, gender differences may well be related to the wording of tasks or instructions, or social behaviors that influence how parents, teachers, and even female students view their own mathematical ability, not by any intrinsic math difficulties (see Lindberg et al., 2010).

This study presents several limitations. All the tasks considered had a time limit, which can influence the results. Processing speed is essential for school performance (Dodonova and Dodonov, 2012), and it is important for the diagnosis of learning disabilities, since children with a diagnosis are identical to those with mean performance if we provide them sufficient time (Jordan and Montani, 1997). The second constraint of this study is that no information was obtained on the socioeconomic level of the children. However, both samples attended public schools, and none were disadvantaged. Lastly, this study is exploratory, and despite having a large sample, it cannot be deemed representative of the total population of both countries.

Gender differences in mathematical performance have been less studied in Spanish-speaking countries. This study was laid out as an initial approach to researching this population. Based on existing literature and the results, gender differences vary by task, strategies used to solve problems, cohort, age, and instructions provided. For future research, the sample should be expanded geographically for higher representativeness in relation to social, educational, and cultural differences. Furthermore, it would be interesting to expand the number of courses evaluated to observe the pattern between genders and to include more complex tasks; this should allow specific mathematical areas to be identified in which gender differences can easily be detected (Chacón-Moscoso et al., 2019). These new developments would potentially reveal differences related to the country, factorial invariance, differential item functioning (DIF), and the influence of open-ended (as opposed to multiple choice) questions (Chacón-Moscoso et al., 2016).

Practical Value, Implications for Educational Intervention

It should be noted that the tests applied in this study could be used for an early diagnosis of gender difference, and with respect to children with low mathematical performance.

CONCLUSIONS

In conclusion, results emphasizing that males outperform females in math do not hold true in the key mathematical areas analyzed with the exception of mathematical fluency in the Chilean sample, where the effect size was small. Therefore, the acquisition of complex abilities should be the same for males and females. At a scientific level and in the news media, care should be exercised when reporting on gender differences because it can influence own opinion of the girls of their mathematical ability. To the best knowledge of the authors, this study is the first to examine gender differences in basic math skills in Spanish-speaking populations that reveals considerable gaps in international studies on the topic.

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 Chilean Commission for Scientific and Technological Investigation (CONICYT, in its Spanish acronym). The protocol was approved by the Ethics Committee of Universidad de la Frontera (Act 066-2017, on Sheet 036-17). The study was in accordance with the standards of the ethical committee of Universidad Autónoma de Chile and with the agreement between the Region of Murcia's Comunidad Autónoma through the Department of Education and Universities, and Universidad de Murcia. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

VP, JF-C, and DM conceptualized and designed the study. MS acquired the data. VP, JF-C, and SC-M analyzed and interpreted the data. VP, JF-C, and DM wrote the manuscript. JF-C prepared the figures and tables. SC-M discussed results and reviewed writing. All the authors read and approved the final version of the manuscript.

FUNDING

This study has been supported by the Chilean national projects Fondecyt Regular 2018 ID1181472, Anid/Fondecyt Regular ID1210172 and Fondecyt Regular 2019 (ref. number 1190945) and by a fellowship for residencies abroad of young researchers from Mare Nostrum (2017). Also, this research has been supported by the ERDF Andalusia Operational Program 2014–2020, government of Andalusia, Spain (ref. US-1263096); the VI Own Research and Transfer Plan (VIPPITUS), Universidad de Sevilla, Spain (ref. VIPP PRECOMPETI 2020/1333).

REFERENCES

Abedi, J., and Lord, C. (2001). The language factor in mathematics tests. *Appl. Measure. Edu.* 14, 219–234. doi: 10.1207/S15324818AME1403_2

Adams, R. B., Barber, B. M., and andOdean, T. (2019). *The Math Gender Gap and Women's Career Outcomes.* Available online at SSRN: https://ssrn.com/abstract=2933241 (accessed December 17, 2019).

Anaya, L., Stafford, F. P., and andZamarro, G. (2017). Gender Gaps in Math Performance, Perceived Mathematical Ability and College STEM Education: The Role of Parental Occupation. EDRE Working Paper No. 2017-21. Available online at SSRN: https://ssrn.com/abstract=3068971 (accessed November 10, 2017).

Baker, D. P., and Jones, D. P. (1993). Creating gender equality: cross-national gender stratification and mathematical performance. *Sociol. Edu.* 66, 91–103. doi: 10.2307/2112795

Barbaresi, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., and Jacobs, S. J. (2005). Math learning disorder: incidence in a population-based birth cohort, 1976 – 82, Rochester, Minn. *Ambul. Pediatr.* 5, 281–289B. doi: 10.1367/A04-209R.1

Barbero-García, I., Holgado-Tello, F. P., Vila-Abad, E., and Chacón-Moscoso, S. (2007). Actitudes, hábitos de estudio y rendimientoenMatemáticas: diferencias por género. *Psicothema* 19, 413-421. Available online at: http://www.psicothema.com/pdf/3379.pdf

Belloum, A. S. Z., Koulouzis, S., Wiktorski, T., and andManieri, A. (2019). Brinding the demand and the offer in data science. *Concurrency Comput. Pract. Exp.* 31:E5200. doi: 10.1002/cpe.5200

Byrnes, J. P., and Takahira, S. (1993). Explaining gender differences on SAT-math items. *Dev. Psychol.* 29, 805–812. doi: 10.1037/0012-1649.29.5.805

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

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

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Lawrence Erlbaum Associates.

De Smedt, B., Peters, L., and Ghesquière, P. (2019). "Neurobiological origins of mathematical learning disabilities or dyscalculia: a review of brain imaging data," in *International Handbook of Mathematical Learning Difficulties*, eds A.Fritz, V. G. Haase, & P. Räsänen (Cham, Switzerland: Springer), 367–384. doi: 10.1007/978-3-319-97148-3_23

Decreto n.° (198/2014). BORM de 5 de septiembre, por el que se establece el currículo de la Educación Primaria en la Comunidad Autónoma de la Región de Murcia (Región de Murcia: Boletin Oficial de la Region de Murcia (BORM), n° 206, 2014).

Diamantopoulou, S., Pina, V., Valero-Garcia, A., González-Salinas, C., and Fuentes, L. J. (2012). Validation of the Spanish version of the woodcock–johnson mathematics achievement tests for children aged 6 to 13. *J. Psychoedu. Assess.* 30, 466–477. doi: 10.1177/0734282912437531

Dodonova, Y. A., and Dodonov, Y. S. (2012). Processing speed and intelligence as predictors of school achievement: Mediation or unique contribution? *Intelligence* 40, 163–171. doi: 10.1016/j.intell.2012.01.003

Donlan, C., Cowan, R., Newton, E. J., and Lloyd, D. (2007). The role of language in mathematical development: Evidence from children with specific language impairment. *Cognition* 103, 23–33. doi: 10.1016/j.cognition.2006.02.007

Halpern, D. F., and Wright, T. M. (1996). A process-oriented model of cognitive sex differences. *Learn. Indiv. Dif.* 8, 3–24. doi: 10.1016/S1041-6080(96)90003-5

Havelock Ellis (1894). *Man and Woman: A Study of Human Secondary Sexual Characters*. New york, NY: Scribner's.

Hickendorff, M.,Torbeyns, J., and Verschaffel, L. (2019). "Multi-digit addition, subtraction, multiplication, and division strategies," in *International Handbook of Mathematical Learning Difficulties*, eds A. Fritz, V. Haase& P. Räsänen (Cham, Switzerland: Springer), 543–560. doi: 10.1007/978-3-319-97148-3_32

Hutchison, J., Lyons, I., and Ansari, D. (2019). More similar than different: gender differences in basic numeracy are the exception, not the rule. *Child Dev.* 90, 66000–79000. doi: 10.1111/cdev.13044

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., and Williams, C. C. (2008). Gender similarities characterize math performance. *Science* 321, 494–495. doi: 10.1126/science.1160364

Instituto Nacional de Evaluación Educativa [INEE] (2016). TIMSS 2015. Estudio internacional de tendencias en Ciencias y Matemáticas. IEA. Informe español: Resultados y contexto. Madrid: Ministerio de Educación, Cultura y Deporte.

Instituto Nacional de Evaluación Educativa [INEE] (2020). TIMSS 2019. Estudio internacional de tendencias en Ciencias y Matemáticas. IEA. Informe español: Resultados y contexto. Madrid: Ministerio de Educación, Cultura y Deporte.

Jordan, N. C., Kaplan, D., Olah, L., and Locuniak, M. N. (2006). Number sense growth in kindergarten: a longitudinal investigation of children at risk for mathematics difficulties. *Child Dev.* 77, 153–175. doi: 10.1111/j.1467-8624.2006.00862.x

Jordan, N. C., and Montani, T. O. (1997). Cognitive arithmetic and problem solving: a comparison of children with specific and general mathematics difficulties. *J. Learn. Disabil.* 30, 624–634. doi: 10.1177/002221949703000606

Kane, J. M., and Mertz, J. E. (2012). Debunking myths about gender and mathematics performance. *Not. AMS* 59, 10–21. doi: 10.1090/noti790

Kersey, A. J.;, Braham, E. J.;, Csumitta, K. D.;, Libertus, M. E., and Cantlon, J. F. (2018). No intrinsic gender differences in children's earliest numerical abilities. *NPJ Sci. Learn.* 3, 1–10. doi: 10.1038/s41539-018-0028-7

Lachance, J. A., and Mazzocco, M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learn. Indiv. Dif.* 16, 195–216. doi: 10.1016/j.lindif.2005.12.001

Lafontaine, D., and Monseur, C. (2009). Gender gap in comparative studies of reading comprehension: to what extent do the test characteristics make a difference? *Eur. Edu. Res. J.* 8, 69–79. doi: 10.2304/eerj.2009.8.1.69

Landerl, K. (2019). "Neurocognitive Perspective on Numerical Development," in *International Handbook of Mathematical Learning Difficulties*, eds A. Fritz, V. G. Haase, and P. Räsänen (Cham, Switzerland: Springer), 9–25. doi: 10.1007/978-3-319-97148-3_2

Lindberg, S. M., Hyde, J. S., and Petersen, J. L.,and Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychol. Bull.* 136, 1123–1135. doi: 10.1037/a0021276

MINEDUC (2018). *Bases Curriculares Primero a Sexto Básico*. Santiago: Ministerio de Educacion.

Nosworthy, N., Bugden, S., Archibald, L. A., Evans, B., and andAnsari, D. (2013). A two-minute paper and pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS ONE* 8:e67918. doi: 10.1371/journal.pone.0067918

Pina, V., Hernández-Pérez, E., Rabadán-Rubio, J. A., Hernández-Pallarés, L., and andFenollar-Cortés, J. (2020). *PreDisCal: Prueba de cribado para la detección de dificultades lectoras o matemáticas de segundo a sexto de primaria*. Madrid: TEA Ediciones.

Ramírez-Uclés, I. M., and Ramírez-Uclés, R. (2020). Gender differences in visuospatial abilities and complex mathematical problem solving. *Front. Psychol.* 11:191. doi: 10.3389/fpsyg.2020.00191

Reigosa-Crespo, V., Valdes-Sosa, M., Butterworth, B., Estevez, N., Rodriguez, M., Santos, E., et al. (2012). Basic numerical capacities and prevalence of developmental dyscalculia: the havana survey. *Dev. Psychol.* 48, 123–135. doi: 10.1037/a0025356

Routitsky, A., and Turner, R. (2003). "Item format types and their influences on cross-national comparisons of student performance," *in Paper Presented at the Annual Meeting of the American Educational Research Association* (Chicago).

Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. G., and Marchant, H. G. (1999). Math fact retrieval as the cognitive mechanism underlying gender differences in math achievement test performance. *Contemp. Edu. Psychol.* 24, 181–266. doi: 10.1006/ceps.1999.1004

Scheiber,C., Reynolds, M. R., Hajovsky, D. B., and Kaufman, A.S. (2015). Gender differences in achievement in a large, nationally representative sample of children and adolescents. *Psychol. Schools* 52, 335–348. doi: 10.1002/pits.21827

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., and DeSmedt (2016). Association of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372

Shields, S. A. (1982). The variability hypothesis: the history of a biological model of sex differences in intelligence. *Signs. J. Women Cult. Soc.* 7, 769–797. doi: 10.1086/493921

UNESCO, Gelber et al. (2016). *Inequidad de género en los logros de aprendizaje en educación primaria? 'Qué nos puede decir TERCE?* Santiago: Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación.

Voyer, D., and Voyer, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychol. Bull.* 140, 1174–1204. doi: 10.1037/a0036620

Winkelmann, H., van den Heuvel-Panhuizen, M., and Robitzsch, A. (2008). Gender differences in the mathematics achievements of German primary school students: results from a German large-scale study. *ZDM* 40, 601–616. doi: 10.1007/s11858-008-0124-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.

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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 October 2021
doi: 10.3389/fpsyg.2021.727848

Assessing the Nomological Network of the South African Personality Inventory With Psychological Traits

Carin Hill[1]*, Jan Alewyn Nel[2], Leon T. de Beer[3], Velichko H. Fetvadjiev[3,4], Lyle I Stevens[1] and Monique Bruwer[3]

[1] Department of Industrial Psychology and People Management, College of Business and Economics, University of Johannesburg, Johannesburg, South Africa, [2] Department of Human Resource Management, University of Pretoria, Pretoria, South Africa, [3] WorkWell Research Unit, North-West University, Potchefstroom, South Africa, [4] Department of Social Psychology, University of Amsterdam, Amsterdam, Netherlands

OPEN ACCESS

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

Reviewed by:
Xiqin Liu, University of Electronic Science and Technology of China, China
Sara Filipiak, Maria Curie-Skłodowska University, Poland

***Correspondence:**
Carin Hill
chill@uj.ac.za

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

Received: *19 June 2021*
Accepted: *17 September 2021*
Published: *11 October 2021*

Citation:
Hill C, Nel JA, de Beer LT, Fetvadjiev VH, Stevens LI and Bruwer M (2021) Assessing the Nomological Network of the South African Personality Inventory With Psychological Traits. Front. Psychol. 12:727848. doi: 10.3389/fpsyg.2021.727848

The purpose of this study was to expand internal construct validity and equivalence research of the South African Personality Inventory (SAPI), as well as to investigate the nomological validity of the SAPI by examining its relationship with specific and relevant psychological outcomes. The internal and external validity of the SAPI was assessed within three separate samples (N = 936). Using the combined data from all three samples, Exploratory Structural Equation Modelling (ESEM) indicated that the six-factor SAPI model fit proved to be excellent. Measurement invariance analyses showed that the SAPI dimensions in the ESEM model were invariant across gender and race groups. Next, two separate studies explored the associations of the SAPI factors with relevant psychological outcomes. An ESEM-within-CFA (set ESEM) method was used to add the factors into a new input file to correlate them with variables that were not part of the initial ESEM model. Both models generated excellent fit. In Study 1, psychological well-being and cultural intelligence were correlated with the SAPI factors within a sample of students and working adults. All of the psychological well-being dimensions significantly correlated with the SAPI factors, while for cultural intelligence, the highest correlations were between Meta-cognition and Openness and Meta-cognition and Positive Social-Relational Disposition. In Study 2, work locus of control and trait anxiety was correlated with the SAPI factors within a sample of adults from the general South African workforce. Work Locus of Control correlated with most factors of the SAPI, but more prominently with Positive Social-Relational Disposition, while Neuroticism correlated strongly with trait anxiety. Finding an appropriate internal structure that measures personality without bias in a culturally diverse context is difficult. This study provided strong evidence that the SAPI meets the demanding requirements of personality measurement in this context and generated promising results to support the relevance of the SAPI factors.

Keywords: South African Personality Inventory, nomological network, psychological traits, general anxiety, work locus of control, psychological well-being, cultural intelligence

INTRODUCTION

Indigenous measures of personality have been developed over recent decades to meet local needs in various non-Western cultures and to reduce the prevailing reliance on imported instruments (Fetvadjiev and Van de Vijver, 2015; Cheung and Fetvadjiev, 2016; Church, 2017). Early indigenous research has devoted extensive attention to the specific cultural (emic) interpretation of local personality constructs. More recent lines of research have sought to expand the culturally specific focus of early studies by applying a broader comparative approach that includes both local and presumed universal (etic) elements in a combined emic–etic approach (Cheung et al., 2011). This approach is characterised by direct comparisons of indigenous instruments to universal concepts, an assessment of indigenous measures' cross-cultural replicability, and examining the predictive value of indigenous instruments for locally relevant outcomes. Notably the second and third aspect have received relatively less attention in the literature (Church, 2017). The present study aims to examine the predictive value of an indigenously developed instrument, the South African Personality Inventory (SAPI; Fetvadjiev et al., 2015), for consequential outcome variables in the domains of cultural intelligence, well-being, and personal growth across three multi-ethnic samples in South Africa.

Personality is known to be related to a range of consequential life outcomes (Ozer and Benet-Martínez, 2006). Evidence has also started accumulating that indigenous or emic–etic measures have a role to play in predicting relevant outcomes. For example, Katigbak et al. (2002) found that Philippine personality scales were associated with various self-reported behaviours and attitudes; the indigenous scales offered improved prediction over and above a Big Five instrument notably for praying. Based on the extensive research programme on the Chinese Personality Assessment Inventory (CPAI), Cheung et al. (2013) reported that the CPAI was related to behaviours indicating variety of interests (e.g., learning languages), variety in social networks (e.g., seeking and giving advice), and interpersonal behaviours with family and friends (e.g., quarrelling and gift-giving). It is worth noting that these associations were observed both in Asian countries (China, South Korea, and Japan) and in the United States, highlighting the emic–etic aspects of the CPAI. In one of the few indigenous studies outside Asia so far, Burtăverde et al. (2018) found that a Romanian indigenous personality instrument explained variance in social adaptation (e.g., career satisfaction), risky social behaviours (e.g., driving fines), and status-striving (e.g., materialism). These research programmes illustrate the value of examining the nomological networks of indigenously derived measures by assessing their associations with relevant criterion variables. Still, this field of research has remained limited and has mostly been confined to Asian samples. The present study aims to advance the field by analysing important criterion variables in the nomological network of an African-derived instrument, the SAPI.

THE SOUTH AFRICAN CONTEXT AND THE SAPI

In South Africa, a markedly multicultural society, the government requires that psychological assessments comply with specific legislation, including the Employment Equity Act (EEA) Section 8 (Act 55 of 1998). The EEA states that psychological assessments need to be scientifically shown to be valid and reliable, fair to all employees, and should not discriminate based on language, race, gender, or culture in any way (see The Republic of South Africa, 1998).

The SAPI project's goal has been to provide South Africa with a personality model that takes into account the implicit concepts of personality found across the 11 official spoken languages (Afrikaans, English, isiNdebele, isiXhosa, isiZulu, Northern Sotho, Setswana, Siswati, Southern Sotho, Tshivenda, and Xitsonga) and that substantiates a psychometrically sound inventory in terms of reliability and validity (Nel et al., 2012; Fetvadjiev et al., 2015). The SAPI was initially conceptualised as a nine-factor model that included Conscientiousness, Emotional Stability, Extraversion, Facilitating, Integrity, Intellect, Openness, Relationship Harmony, and Soft-Heartedness (Nel et al., 2012). Building on this conceptual model, Fetvadjiev et al. (2015) found a factor structure that contains 18 facet scales representing six factors labelled Conscientiousness, Extraversion, Neuroticism, Openness, Negative Social-Relational Disposition, and Positive Social-Relational Disposition. *Conscientiousness* is defined as an individual's orientation toward success, precision, and conventionalism, while *Extraversion* is an individual's tendency toward spontaneous interactions while entertaining others through jokes and stories. The *Neuroticism* factor represents the tendency to be impulsive and have fluctuating emotions, whereas the *Openness* factor describes the quality of being well-informed, rational, and a progressive thinker. The two social-relational factors address how a person typically approaches their relationships with others: *Negative Social-Relational Disposition* describes the extent to which a person typically approaches relations with others in a contentious manner, whereas *Positive Social-Relational Disposition* illustrates a person's inclination toward a positive approach in managing relations with others. Fetvadjiev et al. (2015) established that the SAPI factors were equivalent across various ethnic groups and correlated with impression-management qualities of social desirability while producing weak correlations with deceitful qualities of social desirability. The SAPI's social-relational factors remained relatively distinct when compared to measures of the Big Five (see Valchev et al., 2014; Fetvadjiev et al., 2015). Morton et al. (2019) used a 20-facet version of the SAPI and confirmed the same six-factor structure. Finally, the SAPI model has been replicated in two cultural groups in New Zealand, where the SAPI was found to add incremental value above a Big Five instrument in the prediction of family orientation and well-being (Fetvadjiev et al., 2021).

External Construct Validity

While it is important for newly developed measuring instruments to produce valid and reliable factors, construct validity-related evidence is essential to ensure the actual use of such an instrument in the relevant field (Ziegler et al., 2013). Cronbach and Meehl (1955) introduced the "nomological network" in 1955, stating that such an interlocking system provides researchers with the opportunity to learn more about and enrich a theory-based construct through certain methodological principles which allow for the scientific confirmation of the construct validity of psychological tests. These principles include amongst others that constructs should exhibit frequent lawful relationships with other constructs, and lawful relationships include establishing connections between observable manifestations, between theoretical and observable constructs, or between various theoretical constructs – either statistically or deterministically (Cronbach and Meehl, 1955; Belkhamza and Hubona, 2018). Furthermore, a nomological network provides evidence on how a construct predicts outcome criteria and increases the definiteness of the factors of the theoretical construct (Cronbach and Meehl, 1955; Zettler et al., 2020).

Forming a nomological network is, therefore, a significant way to assess construct validity, and it involves both the internal and external examination of a particular construct (Cronbach and Meehl, 1955; Byrne, 1984). Internal examination studies the relationships among the various construct facets and indicates the legitimacy and replicability of the results, while external examination studies the relationships between the construct and other presumably mutually exclusive constructs (Byrne, 1984; Ziegler et al., 2013). The current study aimed to expand the investigation into the psychometric properties of the SAPI through (1) examining the internal construct validity and equivalence of the SAPI, and (2) examining the external construct validity of the SAPI by way of establishing a nomological network between the SAPI factors and relevant psychological outcomes.

External construct validity within the current study was assessed by examining to what extent the SAPI factors are related to other psychological traits that should be theoretically related (concurrent validity), as well as to what extent the SAPI factors are different from other psychological traits that should be theoretically unrelated (discriminant validity) (Cronbach and Meehl, 1955). Psychological traits can be defined as the "...relatively stable or enduring individual differences in thoughts, feelings and behaviour..." (Church, 2000; p. 651) and a literature search on PsycINFO regarding meta-analytical studies of the relationship between personality and various psychological traits produces a vast amount of research papers. For example, the major factors of personality have been linked with psychological traits such as anxiety (Kotov et al., 2010; McKinney et al., 2021), humour (Mendiburo-Seguel et al., 2015), mindfulness (Giluk, 2009; Ortet et al., 2020), narcissism (Grijalva and Newman, 2015), subjective well-being (Anglim et al., 2020; Ortet et al., 2020), and values (Fischer and Boer, 2015; Nei et al., 2018), to name but a few (see also Ozer and Benet-Martínez, 2006). Within the South African context, various personality traits have been linked with psychological traits such as anxiety (Van Jaarsveld and Schepers, 2007), social adjustment (Papageorgiou and Callaghan, 2018), emotional competence (Coetzee et al., 2006), cultural intelligence (Nel et al., 2015), locus of control (Schepers and Hassett, 2006; Van Wyk et al., 2009), psychological well-being (Jones et al., 2015), and self-esteem (Coetzee et al., 2006). The majority of previous research has used the established Big Five model. The present study broadens this scope by examining several important correlates of an indigenously derived, emic–etic instrument in South Africa. Four external criterion variables were included in the current study to assess the external construct validity of the SAPI: cultural intelligence, general anxiety, psychological well-being, and work locus of control.

OVERVIEW

In this study, we used three separate samples to examine the SAPI's internal and external validity. Sample A was derived using purposive non-probability sampling and included industrial psychologists, intern industrial psychologists, psychometrists, and students in industrial psychology. Sample B was based on non-probability convenience sampling design to collect data from South Africa's general workforce. The participants' work contexts were varied and included financial, accounting and banking industries, sports and medicine, fast-moving consumer goods, law, events, education, engineering, marketing, IT, and non-profit organisations. Sample C was also obtained using a non-probability convenience sampling strategy and included students in a higher education institution in South Africa. The demographical details for each sample can be found in **Table 1**.

A combined investigation including all three samples was done to examine the internal validity and measurement invariance (based on gender and ethnicity). The external validity was examined in two separate studies by determining the relations of the SAPI factors with various psychological traits. Study 1 (using Sample A) focused on cultural intelligence (CQ) and psychological well-being (PWB) as criterion variables. In Study 2 (using Sample B), we directed our focus to trait anxiety and to work locus of control (WLC) as an important aspect of functioning in the work environment.

Preliminary Study: Psychometric Properties of the SAPI

Since the purpose of the SAPI project included developing an assessment measure that could be used across ethnic groups within South Africa, it is important to investigate the construct equivalence of the inventory across groups and samples. Fetvadjiev et al. (2015) determined that the equivalence of the six-factor structure of the SAPI was at least fair, and even very good in most comparisons across the four official ethnic groups within South Africa (Black[1],

[1] An official South African term used for people from African descent.

TABLE 1 | Sample characteristics.

Sample	Total *n*	Gender			Race				
		Men	Women	Missing	Black	Coloured	Indian	White	Missing
A	400	152	248	0	62	45	8	284	1
B	422	146	269	7	80	34	43	244	21
C	114	43	71	0	15	17	3	79	0
Total	936	341	588	7	157	96	54	607	22

Coloured[2], Indian, and White), as well as within a replication study amongst a sample containing only Black and White participants. Fetvadjiev et al. (2015) study was done only amongst university students and adults from security or insurance companies and used the relatively lenient framework of exploratory factor analysis. Morton et al. (2018) investigated the factor structure among various industries and managerial positions using a more stringent structural-equation-modelling approach and found the equivalence of the factor structure of the SAPI to be very good. However, Morton et al. (2018) sample size was relatively small (n = 313), and therefore the further investigation of the psychometric properties of the SAPI in a larger and more varied sample is warranted. Building on this previous work, the current study examines the SAPI's properties in a large, multiethnic sample employing a structural-equation-modelling approach. We examine the SAPI's measurement equivalence across three of the country's four ethnic groups (Blacks, Coloureds, and Whites) as well as both genders.

Investigations of construct equivalence given a previously established factor structure are typically done using confirmatory factor analysis (CFA; Van de Vijver and Leung, 2021). However, CFA presents certain limitations regarding the latent variable measurement specification that warrant using a different approach, namely exploratory structural equation modelling (ESEM; Asparouhov and Muthén, 2009; Van de Vijver and Leung, 2021). There are two main constraints of CFA: (1) it has a stringent requirement of zero cross-loadings that causes the data to produce misfitting models, and subsequent researchers tend to modify their models extensively in search of model fit; and (2) the consequence of misspecified zero loadings in CFA is the distortion of factors, over-estimated factor correlations, and distorted structural relationships (Asparouhov and Muthén, 2009). ESEM, on the other hand, incorporates some of "the best features of exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and structural equation modeling (SEM)" (Marsh et al., 2013, p. 2). While CFA requires the rigorous adherence to item cross-loadings fixed to zero, ESEM allows the estimation of item cross-loadings. As such, theoretically, the latent factor inter-correlations of independent personality factors will be considerably smaller compared to CFA-estimated correlations (Ginns et al., 2014). Therefore, in the current study, the model fit of the six-factor SAPI model was tested using ESEM.

[2] An official South African term used for people from mixed descent.

Method

The SAPI was administered to 936 students and working adults in Samples A, B, and C. Apart from meeting the general target sample descriptions of the respective sample (working adults and students in the fields of industrial psychology and psychometrics in Sample A; working adults from the general work force in Sample B; students in Sample C), participants had to be 18 years or older to complete the questionnaires. Research proposals concerning the studies were presented to research committees at the various supervising universities, and ethical clearance was granted for each study. Participation was voluntary, and the purpose of the research was clearly explained. Each participant was provided with a letter of consent, and ethical aspects such as confidentiality were explained and assured, as well as the option to withdraw at any given moment. Due to its small sample size, the Indian race group was excluded from the current analyses.

The SAPI version used in this study consisted of 146 items grouped into 19 facet scales, representing the six SAPI factors. The version used in this study was a preliminary version of the SAPI that was adapted in the articles by Fetvadjiev et al. (2015) and Morton et al. (2019). The responses are provided on a Likert scale that ranges from 1 (strongly disagree) to 5 (strongly agree). The facet scores were used as indicators of the factors in this study. The ESEM analyses were executed using robust maximum likelihood estimation (MLR) in Mplus 8.6 (Muthén and Muthén, 2021). To assess the measurement invariance of the model, we tested the configural, metric, and scalar invariance. Due to the sensitivity of chi-square to sample size, we used the rule of thumb for maximum change in CFI (0.01), SRMR (0.030), and RMSEA (0.015) (Chen, 2007).

Results

The six-factor model of the SAPI was fitted to the observed data using ESEM. The fit of the six-factor model proved to be excellent: χ^2 = 333.77 (df = 72, p < 0.001), RMSEA = 0.062 (90% CI: 0.056, 0.069), CFI = 0.975, TLI = 0.942, SRMR = 0.015. The Cronbach's alpha reliabilities of the facets ranged between 0.71 and 0.89, with the exception of Straightforwardness (0.58) and Deceitfulness (0.60), both of which have only three items. At the factor level, the following reliability coefficients were found: Conscientiousness (0.93), Extraversion (0.89), Neuroticism (0.84); Openness (0.88); Negative Social-Relational Disposition (0.90); Positive Social-Relational Disposition (0.96). **Table 2** provides the correlations between the factors. The results of the measurement invariance testing showed that the SAPI structure was invariant at the scalar

TABLE 2 | Correlation matrix of the latent variables.

	M	SD	1	2	3	4	5
Conscientiousness	4.07	0.79	–				
Extraversion	3.78	0.90	0.30	–			
Neuroticism	2.57	0.98	−0.30	−0.01	–		
Openness	4.01	0.78	0.56	0.55	−0.24	–	
Negative Social-Relational Disposition	2.00	0.92	−0.30	−0.01	0.34	−0.09	–
Positive Social-Relational Disposition	3.98	0.72	0.65	0.58	−0.15	0.65	−0.28

All displayed correlations above and below −0.01 are significant at p < 0.05 or lower.

TABLE 3 | Results of the measurement invariance testing.

Model group: gender	χ^2	*df*	CFI	Δ CFI	RMSEA	Δ RMSEA	SRMR	Δ SRMR
Configural invariance	402.39	144	0.976	–	0.062	–	0.017	–
Metric invariance	484.91	222	0.976	0.000	0.050	0.012	0.045	−0.028
Scalar invariance	496.74	235	0.976	0.000	0.049	0.001	0.044	0.001
Model group: ethnicity	χ^2	*df*	**CFI**	**Δ CFI**	**RMSEA**	**Δ RMSEA**	**SRMR**	**Δ SRMR**
Configural invariance	414.80	216	0.980	–	0.057	–	0.018	–
Metric invariance	631.20	372	0.974	0.006	0.049	0.008	0.044	−0.026
Scalar invariance	675.09	398	0.972	0.002	0.049	0.000	0.047	−0.003

Gender, male and female; ethnicity, black, coloured, and white.

level across both gender and race groups according to the adopted cut-off criteria (see **Table 3**).

Study 1: The SAPI, Cultural Intelligence, and Psychological Well-Being

Construct validation research focuses on finding empirical confirmation that certain hypothesised relationships exist within a construct's nomological network (Byrne, 1984). Both CQ and PWB are known to be related to personality (Ryff, 1989; Ang et al., 2006; Grant et al., 2009; Cheung et al., 2012; Nel et al., 2015) and were examined as relevant outcomes.

Cultural intelligence is defined as an individual's ability to function in a multicultural setting or situation (Earley and Ang, 2003). According to Earley and Ang (2003) and Ang et al. (2006), CQ consists of four dimensions: Behavioural CQ (an individual's ability to act appropriately following multicultural aspects, such as values and beliefs of different cultures); Cognitive CQ (an individual's knowledge of multicultural aspects); Meta-cognitive CQ (an individual's thought processes in order to understand cultural contexts); and Motivational CQ (the amount of energy an individual invests in understanding multicultural aspects). Research focusing on the relationship between personality and CQ has increased over the last decade (see Ott and Michailova, 2018). The four CQ dimensions tend to correlate with Agreeableness mostly, but also with Conscientiousness, Extraversion, and Openness (Huff et al., 2014; Li et al., 2016; Presbitero, 2016; Shu et al., 2017; Wang et al., 2019; Camargo et al., 2020). Li et al. (2016) found correlations between Emotional Stability and Behavioural CQ, Meta-cognitive CQ, and Motivational CQ. Huff et al. (2014) identified relationships between Intellect and Cognitive CQ, Meta-cognitive CQ, and Motivational CQ, whereas Shu et al. (2017) established that the HEXACO's Honesty-Humility factor significantly correlated with Meta-cognitive CQ and Motivational CQ. To date, only Nel et al. (2015) provided evidence on the ability of SAPI factors and facets using the initially conceptualised nine-factor SAPI model to predict CQ. Their study concluded that Intellect and Facilitating predicted Meta-cognitive CQ; Soft-heartedness, Facilitating, and Extraversion predicted Motivational CQ; while Soft-heartedness and Conscientiousness predicted Behavioural CQ. CQ is of central importance for functioning in multicultural contexts such as South Africa. It is thus highly relevant to examine the role of local personality measures, particularly the SAPI with its emphasis on social-relational aspects, in accounting for individual differences in CQ.

Psychological well-being, in turn, refers to individuals' need to function optimally, realise attributes and talents unique to themselves, and focus on identity, purpose and meaning, and relations to others (Ryff and Keyes, 1995; Ryff and Singer, 1996). Ryff (1989) model of PWB consists of six core factors: Autonomy (going about following one's standards rather than the opinions of others); Environmental Mastery (participation in external activities); Personal Growth (to advance in knowledge, skills, and potential); Positive Relations with Others (the presence of close relationships with others in one's life); Purpose in Life (having a sense of determination and significance in one's life); and Self-acceptance (maintaining a positive attitude toward oneself). Correlational relationships have been established between the six PWB factors and personality traits. Neuroticism and negative affect tend to be negatively correlated with all of the PWB factors (Schmutte and Ryff, 1997; Burns and Machin, 2010; Jones et al., 2015; Anglim and Grant, 2016), while Schmutte and Ryff (1997) found no correlation between negative effect and Personal Growth. Extraversion and positive affect also tend to be positively correlated with all PWB factors (Schmutte and Ryff, 1997; Burns and Machin, 2010; Anglim and Grant, 2016), although Jones et al. (2015) found no correlation between extraversion and Autonomy. Openness has been found to be correlated with all PWB factors except Environmental Mastery, and Agreeableness, with all factors except Autonomy (Anglim and Grant, 2016). Lastly, Conscientiousness positively correlated with all of the PWB factors (Jones et al., 2015; Anglim and Grant, 2016).

It was speculated that the components of the SAPI would relate to certain CQ and PWB dimensions, contributing to convergent validity. Based on the previous research on CQ, it was expected that CQ as a whole would be most systematically related to the Positive Social-Relational domain, conceptually related to Agreeableness. Furthermore, associations between Openness and the cognitive and motivational aspects of CQ could be expected. For example, people who possess the Openness characteristics of being well-informed, a quick learner, adaptable, articulate, innovative, and perceptive would in all likelihood be more knowledgeable of customs in cultures different from their own and have a higher level of deliberate cultural consciousness when interacting with people from different cultures (i.e., high on Cognitive CQ and Meta-cognitive CQ). With respect to PWB, the most consistent relationships could be expected for Neuroticism and Conscientiousness, with more varied relationships for the other factors. It was also anticipated that a person who tends to be accommodating and loyal, compassionate and encouraging, as well as understanding and considerate (high on Positive Social-Relational Disposition), would probably also rank high on having close relationships with others in one's life (positive relations with others). Finally, people who tend to be indiscreet and deceitful, who tend to exclusively focus on their own needs and see themselves as more important than others (Negative Social-Relational Disposition), will presumably have less satisfying interpersonal relationships and less engagement with other cultures (hence, negative relationships with CQ and PWB aspects).

Method

The SAPI, a cultural intelligence measure, and a psychological well-being measure were administered to 400 students and working adults in Sample A. The SAPI factor scales had the following values of Cronbach's alpha: Conscientiousness (0.93), Extraversion (0.88), Neuroticism (0.81), Openness (0.86), Negative Social-Relational (0.89), and Positive Social-Relational (0.96).

A 20-item Cultural Intelligence Scale developed by Van Dyne et al. (2015) was used. The scale consists of four factors labelled Meta-cognitive CQ (4 items; e.g., "I am conscious of the cultural knowledge I apply to cross-cultural interactions"), Cognitive CQ (6 items; "I know the arts and crafts of other cultures"), Motivational CQ (5 items; "I enjoy interacting with people from different cultures"), and Behavioural CQ (5 items; "I vary the rate of my speaking when a cross-cultural situation requires it"). Each of these dimensions is measured on a five-point response scale ranging from 1 (strongly disagree) to 5 (strongly agree). The internal consistency of the CQ constructs within a South African sample (Nel et al., 2015) were more than adequate, with alpha coefficients ranging between 0.82 (Motivational CQ) and 0.91 (Behavioural CQ).

We used the Psychological Well-being Scale (PWBS) developed by Ryff (1989), which consists of 84 items representing the six dimensions. Each item within the dimensions is answered on a five-point response scale ranging from 1 (strongly disagree) to 5 (strongly agree). Studies using the 84-item version of the PWBS found acceptable Cronbach's alpha coefficients ranging between 0.77 and 0.93 for the six dimensions (Ryff, 1989; Van Dierendonck, 2004; Davidson, 2006). The Cronbach's alpha coefficients of the CQ and PWB measures in the current study are presented in **Table 4**.

Results and Discussion

The six-factor model of the SAPI was fitted to the observed data using ESEM. The fit of the six-factor model proved to be excellent: $\chi^2 = 202.04$ ($df = 72$; $p < 0.001$), RMSEA = 0.067 (90% CI: 0.056, 0.078), CFI = 0.969, TLI = 0.928, SRMR = 0.018. Then an ESEM-within-CFA (set ESEM) method was used to add the SAPI factors into a new input file in order to be able to correlate the SAPI factors with variables that were not part of the ESEM model. This specification allows for the retention of the ESEM parameters for the SAPI model whilst not allowing any cross-loadings from the other variables' items in the model, which retain their traditional CFA structure. For the outcome variables, latent variables were specified by using the composite score as a single indicator for the respective latent variable, with the residual variance constrained to 1-reliability. The assumption was reliability of at least 0.70, thus imposing a constraint on the residual variance of the composite score for the specified latent variable. All outcome variables were included in the same model. The fit of this model was also excellent: $\chi^2 = 421.64$ ($df = 202$; $p < 0.001$), RMSEA = 0.052 (90% CI: 0.045, 0.059), CFI = 0.968, TLI = 0.937, SRMR = 0.21.

In order to place CQ and PWB in the nomological network of the SAPI, product-moment correlation analysis was used to determine the relationships between the SAPI and the constructs of CQ and PWB. The results are presented in **Table 4**. As can be seen from the table, CQ was most consistently correlated with PSR and Openness, with low to moderate correlations. The correlations with the other personality factors were generally smaller and limited to individual CQ components. The pattern is broadly consistent with previous research, although it appears that in the South African context, CQ can be best understood by its association with positive social-relational traits and openness. The limited correlations with extraversion and conscientiousness are different from some previous studies in Western samples and from Nel et al. (2015) results in South Africa using the early, conceptual SAPI model. These findings suggest that the broad PSR dimension may have subsumed some of the variance that could be attributed to other factors.

For the PWB factors, the most consistent and generally highest correlations were observed for Conscientiousness and Neuroticism, followed by Openness and the two social-relational

TABLE 4 | Correlations of the latent variables of Study 1.

	M	SD	α	C	E	N	O	NSR	PSR
Meta-cognition CQ	3.67	0.68	0.79	0.18	0.15	−0.24	0.38	−0.08	0.38
Cognitive CQ	3.04	0.75	0.84	0.10	0.17	−0.16	0.20	−0.05	0.14
Motivational CQ	3.59	0.71	0.81	0.07	0.20	−0.25	0.37	−0.10	0.26
Behavioural CQ	3.24	0.79	0.84	0.04	0.15	−0.12	0.28	−0.01	0.25
Autonomy	4.05	0.56	0.81	0.30	0.13	−0.60	0.34	−0.31	0.24
Environmental mastery	4.08	0.54	0.84	0.56	0.20	−0.66	0.29	−0.37	0.42
Purpose in life	4.37	0.55	0.84	0.61	0.20	−0.53	0.35	−0.38	0.46
Personal growth	4.45	0.51	0.83	0.36	0.31	−0.46	0.63	−0.27	0.50
Positive relations	4.28	0.55	0.82	0.31	0.51	−0.37	0.34	−0.33	0.47
Self-acceptance	4.22	0.60	0.87	0.42	0.19	−0.66	0.26	−0.32	0.32

All displayed correlations below −0.10 and above 0.10 are significant at $p < 0.05$ or lower.
C, conscientiousness; E, extraversion; N, neuroticism; O, openness; NSR, Negative Social-Relational Disposition; PSR, Positive Social-Relational Disposition.

TABLE 5 | Correlations of the latent variables of Study 2.

	M	*SD*	α	1	2	3	4	5	6	7	8
Conscientiousness	4.04	0.75	0.92	–							
Extraversion	3.72	0.88	0.89	0.35	–						
Neuroticism	2.64	0.93	0.84	−0.30	−0.16	–					
Openness	3.99	0.73	0.89	0.51	0.49	−0.36	–				
NSR	2.01	0.89	0.89	−0.30	−0.05	0.29	−0.05	–			
PSR	3.91	0.68	0.95	0.58	0.56	−0.21	0.56	−0.30	–		
GAD	0.98	0.79	0.91	−0.14	−0.07	0.61	−0.17	0.17	−0.15	–	
WLC	2.49	0.68	0.84	−0.29	−0.17	0.22	−0.28	0.21	−0.30	0.21	–

All displayed correlations below −0.07 and above −0.05 are significant at $p < 0.05$ or lower.
WLC, work locus of control; GAD, trait anxiety; NSR, Negative Social-Relational Disposition; PSR = Positive Social-Relational Disposition.

factors. Consistently with previous research, Extraversion and PSR had their lowest correlation with Autonomy, and Openness had one of its lowest correlations with Environmental Mastery (Jones et al., 2015; Anglim and Grant, 2016). Also consistent with expectations, PSR and NSR were meaningfully related to the Relations with Other component of PWB. Finally, NSR tended to have meaningful associations with PWB, but was essentially not related to CQ.

Study 2: The SAPI, Trait Anxiety, and Work Locus of Control

Study 3 extended the investigation into the convergent and discriminant validity of the SAPI and the establishment of its nomological network by first of all investigating the correlations between the SAPI and measures of anxiety and WLC.

Saviola et al. (2020) describe anxiety as "…a mental state characterised by an intense sense of tension, worry or apprehension, relative to something adverse that might happen in the future" (p. 1). Anxiety is usually studied either as a trait or a state (Wilt et al., 2011). Trait anxiety is a personality trait that can be identified as an individual's general inclination to be anxious or the natural anxiety levels exhibited by a person (Vreeke and Muris, 2012; Leal et al., 2017). In contrast, state anxiety refers to a person's anxiety levels over a short period without the presence of particular pathological conditions (Vreeke and Muris, 2012; Saviola et al., 2020). The present study focused on anxiety as a trait. Research has repeatedly shown that trait anxiety is positively related to the Big Five's Neuroticism/Negative Emotionality (Kotov et al., 2007; Karsten et al., 2012; Vreeke and Muris, 2012; Watson and Naragon-Gainey, 2014; Fowler et al., 2017; Goldstein et al., 2018; Naragon-Gainey and Watson, 2018; Watson et al., 2019; Qu et al., 2020), and negatively related to Conscientiousness (Vreeke and Muris, 2012; Watson and Naragon-Gainey, 2014; Qu et al., 2020), and to Extraversion/Positive Emotionality (Kotov et al., 2007; Qu et al., 2020). Goldstein et al. (2018) found trait anxiety to be positively related to Openness, while Qu et al. (2020) found a significant negative relationship between the two constructs. In the HEXACO PI-R, Anxiety is a subscale of the Emotionality factor, suggesting that a person with very high scores on the Emotionality scale experiences anxiety in response to stressors (Lee and Ashton, 2004). Ashton et al. (2007) found that an Anxiety scale loaded strongly on low Agreeableness and Emotionality.

Locus of control refers to a person's general level of expectancy toward a situation they have experienced (Aubé et al., 2007; Burger, 2008; Omari et al., 2012). People can present either an internal locus of control or an external locus of control. An internal locus of control refers to a person's belief that the results of certain events are due to his or her personal ability, efforts, and dedication (Aubé et al., 2007; Omari et al., 2012). Individuals with an external locus of control tend to interpret events due to chance, luck, fate, authoritative others, or circumstantial complexity (Rotter, 1966; Van Praag et al., 2004; Burger, 2008; Aghaei et al., 2013). In an attempt to produce a work-specific measure of locus of control, Spector (1988) developed the Work Locus of Control Scale (WLCS) that measures generalised beliefs regarding whether or not people can control reinforcements within the work context, tapping into both internal control and external control (Spector and O′Connell, 1994; Oliver et al., 2006; Aubé et al., 2007). According to Bosman et al. (2005), a person's assessment of the relationship between how they behave at work and the subsequent rewards or punishments represents that person's work locus of control. External locus of control has been found to be positively correlated with Neuroticism and negatively correlated with the remainder of the Big Five factors (see Chen et al., 2016; Lovell and Brown, 2017; Smidt et al., 2018; Žitný and Halama, 2011). The more situation-specific external work locus of control trait has been positively correlated with psychological traits such as Neuroticism, Trait anxiety, work anxiety, Negative Affectivity, and Type A impatience (see Cook et al., 2000; Johnson et al., 2009), while negatively correlated with Extraversion, Agreeableness, Conscientiousness, emotional intelligence, Type A achievement, and autonomy (see Cook et al., 2000; Johnson et al., 2009; Spector and O′Connell, 1994).

Based on the previous research, it is expected that anxiety and external WLOC would be positively correlated with Neuroticism and would tend to have negative correlations with the rest of the SAPI factors except the Negative Social-Relational. As for the latter, because it shares aspects of negative valence with Neuroticism and is known to be moderately negatively correlated with the Positive Social-Relational factor (Fetvadjiev et al., 2015; Table 2), it can be expected to correlate positively with both criterion variables; the association with anxiety should be weaker than for Neuroticism, given the limited conceptual correspondence.

Method

The SAPI, a trait anxiety measure, and a work locus of control measure were administered to 422 adults from the general South African workforce in Sample B. Work locus of control was only assessed in the subsample of working adults.

We used the Generalised Anxiety Disorder 7-item scale (GAD-7), developed by Spitzer et al. (2006) to be used as a clinical measure to assess Generalised Anxiety Disorder. It can also be used to assess non-clinical trait anxiety. The GAD-7 is a seven-item scale that makes use of four response options, namely: "not at all" (0), "several days" (1), "more than half the days" (2), and "nearly every day" (3). No items are reverse scored. Scores on the GAD-7 range between zero and 21 and represent either mild ($\geq$5), moderate ($\geq$10) or severe ($\geq$15) levels of symptoms of anxiety (Löwe et al., 2008). The Cronbach's alpha coefficients of the GAD-7 suggests excellent internal consistency (0.92) (Spitzer et al., 2006).

To measure WLC, we used the Work Locus of Control Scale (WLCS) developed by Paul Spector (1988) to assess workplace beliefs that relate to both internal and external locus of control (Spector and O′Connell, 1994). The WLCS is a 16-item questionnaire rated on a six-point Likert scale ranging from 1 (disagree very much) to 6 (agree very much). Half of the questions/items relate to the internal rewards domain, and the other half relate to the external rewards domain (Spector and O′Connell, 1994). Therefore, half of the items (8 items) are reverse scored. In the interpretation of the WLCS, a high score is indicative of a more external locus of control, while a low score is indicative of a more internal locus of control (Macan et al., 1996; Aubé et al., 2007). The internal consistency of the WLCS coefficient alphas ranges from 0.75 to 0.85 (Spector, 1988; Spector et al., 2002). The Cronbach's alpha coefficients of the SAPI, GAD-7, and WLCS in the current study are presented in **Table 5**.

Results

The six-factor model of the SAPI was fitted to the observed data using ESEM. The fit of the six-factor model proved to be excellent: $\chi^2 = 217.81$ ($df = 72$; $p < 0.001$), RMSEA = 0.069 (90% CI: 0.059, 0.080), CFI = 0.967, TLI = 0.921, SRMR = 0.019. Then an ESEM-within-CFA (set ESEM) method was used to add the factors into a new input file, using the same specification as described for Study 1. The fit of this model was also excellent: $\chi^2 = 280.21$ ($df = 98$; $p < 0.001$), RMSEA = 0.066 (90% CI: 0.057, 0.076), CFI = 0.961, TLI = 0.917, SRMR = 0.022.

The product-moment correlation between the SAPI, the WLCS, and the GAD-7 was examined to place Work Locus of Control and Trait Anxiety in the SAPI's nomological network. **Table 5** presents the correlations for the variables. In line with expectations, anxiety was strongly positively correlated with Neuroticism and had weaker correlations with the other factors. The sizes of WLC's correlations varied little across factors and were also in line with previous research: External WLC was positively related to Neuroticism (as well as the Negative Social-Relational factor) and negatively to the other factors.

GENERAL DISCUSSION

The current study aimed to examine the psychometric properties of the SAPI (Fetvadjiev et al., 2015) by evaluating the internal structure and the external relations with various psychological traits. A newly developed instrument needs to demonstrate salient and consistent validity and reliability to be viewed as a dependable instrument in the relevant field (Ziegler et al., 2013). The SAPI model not only displayed consistent internal validity and consistency across three separate student and working adults samples, but also demonstrated meaningful relations with various psychological traits, adding to its nomological network.

The SAPI showed scalar measurement invariance pertaining to gender and ethnicity. Testing for measurement invariance plays a vital part in personality research since it is crucial to ensure that no elements are measured that may demonstrate bias and difference in meaning (Van de Vijver and Leung, 2021). Therefore, in this study, we did not detect bias in measuring personality based on gender and ethnicity, highlighting the potential of the SAPI as a bias-free instrument for cross-cultural comparisons.

The findings on the network of associations between the SAPI and relevant psychological traits are encouraging. We examined the correspondence of cultural intelligence, psychological well-being, work locus of control, and trait anxiety with the SAPI factors. We found that the SAPI factors correlated with various cultural intelligence factors on a small or medium effect, which is not surprising since the SAPI was developed by keeping in mind cultural and social factors, and the prominence of the social-relational orientation that was found in the initial phases of the SAPI development (Nel et al., 2012; Fetvadjiev et al., 2015). The strongest correlations seem to be between Openness with Meta-cognition CQ and Motivational CQ, while Positive Social-Relational Disposition corresponded strongly with Meta-cognition CQ. Both factors had associations with all CQ components It is evident that the cognitive adaptability in a multicultural context is associated with the traits of being open and managing constructive relations.

The SAPI factors correlated moderately to highly with psychological well-being factors. Conscientiousness (positively) and Neuroticism (negatively) corresponded strongly with all psychological well-being factors. Conscientiousness is associated with a person's direction, organisation or positioning of life, while Neuroticism focuses on the emotional management of a person. Therefore, psychological well-being seems to relate strongly to elements captured in SAPI's Conscientiousness and Neuroticism, in line with previous studies (Anglim et al., 2020; Ortet et al., 2020). Mastering and managing one's environment and emotions are thus key qualities to obtain overall well-being. Another SAPI factor that correlated with all psychological well-being factors (except with Autonomy) is Positive Social-Relational Disposition. It is illustrative of the importance of the Positive Social-Relational Disposition that it correlated with the factors of subjective well-being since enhancing social relations is fundamental to achieve overall well-being (Anglim and Grant, 2016). Openness, in turn, showed strong correspondence with Personal Growth. Personal growth is close to the meta-cognition process of cultural intelligence, which corresponded closely with SAPI's Openness. This further enforces the relation that a person needs to demonstrate Openness to manage self-relation and how one views the social environment and process information for appropriate conduct. This finding highlights the practical relevance of Openness, a factor whose replication has sometimes been questioned in etic studies where Western instruments have been used in non-Western populations. Taken as a whole, our findings illustrate the strong potential of the SAPI to predict individual differences in elements important to well-being.

In line with expectations, Neuroticism correlated strongly with trait anxiety, which confirms previous findings (Fowler et al., 2017; Goldstein et al., 2018; Naragon-Gainey and Watson, 2018; Watson et al., 2019; Qu et al., 2020). Finally, work locus of control correlated with most factors of the SAPI, particularly with Conscientiousness, Openness, and Positive Social-Relational Disposition. The understanding of work locus of control pertains to the internal and external control of elements in the workplace and the outcomes thereof (Oliver et al., 2006; Aubé et al., 2007). Whereas Positive Social-Relational Disposition pertains more to the person's constructive handling and managing of relationships with others (Fetvadjiev et al., 2015), work locus of control may be partly related the tendency to maintain positive relations, in whether a person can control aspects of the social work environment to facilitate constructive relations.

LIMITATIONS

The current study is not without limitations. The cross-sectional nature of the various studies implies that causal relationships between the various psychological constructs and the SAPI could not be determined. Longitudinal studies could identify trends and changes in the relationship between personality and these constructs over time. Another limitation is the lack of an appropriate sample size for the Indian ethnic group; it is recommended that a more stratified approach be taken in future SAPI research to ensure equal representation of the various race groups, and that overall larger samples are studied. Considering the high correlation of neuroticism with anxiety in the current study, it is also interesting to explore the potential application of the SAPI in clinical samples.

CONCLUSION

This study generated promising results to support the saliency of the SAPI factors. Finding an appropriate internal structure that measures personality without bias in a diverse context is not easy, and this study provided strong evidence that the SAPI is on the right track to be a dependable and sound personality instrument. The SAPI factors had meaningful associations with relevant psychological outcomes. This study adds to the growing field of emic–etic personality research by embedding the indigenously derived SAPI in a broader network of previously established psychological traits. Future emic–etic research should seek to further examine the nomological networks of indigenous personality measures with reference to both universal and locally salient psychological outcomes.

DATA AVAILABILITY STATEMENT

The datasets generated during and analysed during the current study are not publicly available due to the copyrighted items. Requests to access the datasets should be directed to CH, chill@uj.ac.za.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Department of Industrial Psychology and People Management, Research Ethics Committee, University of Johannesburg and WorkWell Research Unit Research Ethics Committee, North-West University. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CH and JN contributed to conception and design of the study. CH, LS, and MB organised the database. LB performed the statistical analysis. CH, LS, and MB wrote the first draft of the manuscript. CH, JN, LB, and VF wrote sections of the manuscript. CH, JN, LB, and VF contributed to manuscript revision, read, and approved the submitted version. All authors contributed to the article and approved the submitted version.

FUNDING

The South African National Research Foundation Incentive Funding for Rated Researchers supported this research (Grant IFR13011816248).

ACKNOWLEDGMENTS

The current study is an extension and amalgamation of LS and MB's Masters dissertations.

REFERENCES

Aghaei, M., Asadollahi, A., Moezzi, A. D., Beigi, M., and Parvinnejad, F. (2013). The relation between personality type, locus of control, occupational satisfaction and occupational exhaustion and determining the effectiveness of Stress Inoculation Training (SIT) on reducing it among staffers of Saipa Company. *Res. J. Recent Sci.* 2, 6–11.

Ang, S., Van Dyne, L., and Koh, C. (2006). Personality correlates of the four-factor model of cultural intelligence. *Group Organ. Manag.* 31, 100–123. doi: 10.1177/1059601105275267

Anglim, J., and Grant, S. (2016). Predicting psychological and subjective well-being from personality: incremental prediction from 30 facets over the Big 5. *J. Happiness Stud.* 17, 59-80. doi: 10.1007/s10902-014-9583-7

Anglim, J., Horwood, S., Smillie, L. D., Marrero, R. J., and Wood, J. K. (2020). Predicting psychological and subjective well-being from personality: a meta-analysis. *Psychol. Bull.* 146:279. doi: 10.31234/osf.io/gupxj

Ashton, M. C., Lee, K., and Goldberg, L. R. (2007). The IPIP–HEXACO scales: an alternative, public-domain measure of the personality constructs in the HEXACO model. *Pers. Individ. Differ.* 42, 1515–1526. doi: 10.1016/j.paid.2006.10.027

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

Aubé, C., Rousseau, V., and Morin, E. M. (2007). Perceived organisational support and organisational commitment: the moderating effect of locus of control and work autonomy. *J. Manag. Psychol.* 22, 479–495. doi: 10.1108/02683940710757209

Belkhamza, Z., and Hubona, G. S. (2018). "Nomological networks," in *Proceedings of the Twenty-fourth Americas Conference on Information Systems: Emergent Research Forum (ERF), New Orleans, LA, August 16-18, 2018*. Available online at: https://www.researchgate.net/profile/Zakariya-Belkhamza/publication/344622707_Nomological_Networks_in_IS_Research/links/5f85317592851c14bcc3934b/Nomological-Networks-in-IS-Research.pdf

Bosman, J., Buitendach, J. H., and Rothmann, S. (2005). Work locus of control and dispositional optimism as antecedents to job insecurity. *S. Afr. J. Ind. Psychol.* 31, 17–23. doi: 10.4102/sajip.v31i4.217

Burger, J. M. (2008). *Personality*, 7th Edn. Belmont, CA: Thomson Wadsworth.

Burns, R. A., and Machin, M. A. (2010). Identifying gender differences in the independent effects of personality and psychological well-being on two broad affect components of subjective well-being. *Pers. Individ. Differ.* 48, 22–27. doi: 10.1016/j.paid.2009.08.007

Burtăverde, V., De Raad, B., and Zanfirescu, A.-Ş. (2018). An emic-etic approach to personality assessment in predicting social adaptation, risky social behaviors, status striving and social affirmation. *J. Res. Pers.* 76, 113–123. doi: 10.1016/j.jrp.2018.08.003

Byrne, B. M. (1984). The general/academic self-concept nomological network: a review of construct validation research. *Rev. Educ. Res.* 54, 427–456. doi: 10.3102/00346543054003427

Camargo, A., Storme, M. and Çelik, P. (2020). Further investigation of the relationship between cultural intelligence and expatriate career intentions. *Int. J. Educ. Vocat. Guidance* 20, 101–122. doi: 10.1007/s10775-019-09396-1

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Modeling* 14, 464–504. doi: 10.1080/10705510701301834

Chen, Y. T., Wang, G., and Xu, G. (2016). Relationships among personality, locus of control, and economic confidence of Chinese migrant workers. *Soc. Behav. Pers.* 44, 415–430. doi: 10.2224/sbp.2016.44.3.415

Cheung, F. M., Cheung, S. F., and Fan, W. (2013). "From Chinese to cross-cultural personality inventory: a combined emic–etic approach to the study of personality in culture," in *Advances in Culture and Psychology*, Vol. III, eds M. Gelfand, C.-Y. Chiu, and Y.-Y. Hong (New York, NY: Oxford University Press), 117–179.

Cheung, F. M., Fan, W., and Yao, J. (2012). "Chinese personality and vocational behavior," in *Handbook of Chinese Organizational Behaviour: Integrating Theory, Research and Practice*, eds X. Huang and M. H. Bond (Northampton: Edward Elgar Publishing Limited), 359–379.

Cheung, F. M., and Fetvadjiev, V. H. (2016). "Indigenous approaches to testing and assessment," in *The ITC International Handbook of Testing and Assessment*, eds F. T. L. Leong, D. Bartram, F. M. Cheung, K. F. Geisinger, and D. Iliescu (Oxford: Oxford University Press), 333–346.

Cheung, F. M., Van de Vijver, F. J. R., and Leong, F. T. L. (2011). Toward a new approach to the assessment of personality in culture. *Am. Psychol.* 66, 593–603. doi: 10.1037/a0022389

Church, A. T. (2000). Culture and personality: toward an integrated cultural trait psychology. *J. Pers.* 68, 651–703. doi: 10.1111/1467-6494.00112

Church, A. T. (2017). "Personality across cultures: historical overview and current topics," in *The Praeger Handbook of Personality Across Cultures*, ed. A. T. Church (Santa Barbara, CA: Praeger), 1–45.

Coetzee, M., Martins, N., Basson, J.S. and Muller, H. (2006). The relationship between personality preferences, self-esteem and emotional competence. *S. Afr. J. Ind. Psychol.* 32, 64–73. doi: 10.4102/sajip.v32i2.233

Cook, K. W., Vance, C. A., and Spector, P. E. (2000). The relation of candidate personality with selection-interview outcomes. *J. Appl. Soc. Psychol.* 30, 867–885. doi: 10.1111/j.1559-1816.2000.tb02828.x

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychol. Bull.* 52, 281–302. doi: 10.1037/h0040957

Davidson, D. (2006). *Psychological, Physical and Social Well-Being in An Individual and Team Sport: A Phenomenological and Quantitative Study*. Ph.D. thesis. Mhlathuze: University of Zululand.

Earley, P. C., and Ang, S. (2003). *Cultural Intelligence: Individual Interactions Across Cultures*. Stanford, MA: Stanford University Press.

Fetvadjiev, V. H., Meiring, D., Van de Vijver, F. J., Nel, J. A., and Hill, C. (2015). The South African Personality Inventory (SAPI): a culture-informed instrument for the country's main ethnocultural groups. *Psychol. Assess.* 27, 827–837. doi: 10.1037/pas0000078

Fetvadjiev, V. H., Neha, T., Van de Vijver, F. J. R., McManus, M., and Meiring, D. (2021). The cross-cultural relevance of indigenous measures: the South African Personality Inventory (SAPI), family orientation, and well-being in New Zealand. *J. Cross. Cult. Psychol.* 52, 3–21. doi: 10.1177/0022022120969979

Fetvadjiev, V. H., and Van de Vijver, F. J. R. (2015). "Measures of personality across cultures," in *Measures of Personality and Social Psychological Constructs*, eds G. J. Boyle, D. H. Saklofske, and G. Matthews, 2nd Edn (Cambridge, MA: Academic Press), 752–776.

Fischer, R., and Boer, D. (2015). Motivational basis of personality traits: a meta-analysis of value-personality correlation. *J. Pers.* 83, 491–510. doi: 10.1111/jopy.12125

Fowler, J. C., Patriquin, M. A., Madan, A., Allen, J. G., Frueh, B. C., and Oldham, J. M. (2017). Incremental validity of the PID-5 in relation to the five factor model and traditional polythetic personality criteria of the DSM-5. *Inter. J. Methods Psychiatr. Res.* 26:e1526. doi: 10.1002/mpr.1526

Giluk, T.L. (2009). Mindfulness, big five personality, and affect: a meta-analysis. *Pers. Individ. Differ.* 47, 805–811. doi: 10.1016/j.paid.2009.06.026

Ginns, P., Martin, A. J., Liem, G. A., and Papworth, B. (2014). Structural and concurrent validity of the International English Mini-Markers in an adolescent sample: exploring analytic approaches and implications for personality assessment. *J. Res. Pers.* 53, 182–192. doi: 10.1016/j.jrp.2014.10.001

Goldstein, B. L., Kotov, R., Perlman, G., Watson, D., and Klein, D. N. (2018). Trait and facet-level predictors of first-onset depressive and anxiety disorders in a community sample of adolescent girls. *Psychol. Med.* 8:1282. doi: 10.1017/s0033291717002719

Grant, S., Langan-Fox, J., and Anglim, J. (2009). The big five traits as predictors of subjective and psychological well-being. *Psychol. Rep.* 105, 205–231. doi: 10.2466/pr0.105.1.205-231

Grijalva E. and Newman, D. A. (2015). Narcissism and Counterproductive Work Behavior (CWB): meta-analysis and consideration of collectivist culture, big five personality, and narcissism's facet structure. *Appl. Psychol.* 64, 93–126. doi: 10.1111/apps.12025

Huff, K. C., Song, P., and Gresch, E. B. (2014). Cultural intelligence, personality, and cross-cultural adjustment: a study of expatriates in Japan. *Int. J. Intercult. Relat.* 38, 151–157. doi: 10.1016/j.ijintrel.2013.08.005

Johnson, S. J., Batey, M., and Holdsworth, L. (2009). Personality and health: the mediating role of trait emotional intelligence and work locus of control. *Pers. Individ. Differ.* 47, 470–475. doi: 10.1016/j.paid.2009.04.025

Jones, N., Hill, C., and Henn, C. (2015). Personality and job satisfaction: the role of psychological well-being. *J. Psychol. Afr.* 25, 297–304. doi: 10.1080/14330237.2015.1078086

Karsten, J., Penninx, B. W. J. H., Riese, H., Ormel, J., Nolen, W. A., and Hartman, C. A. (2012). The state effect of depressive and anxiety disorders on big five personality traits. *J. Psychiatr. Res.* 46, 644–650. doi: 10.1016/j.jpsychires.2012.01.024

Katigbak, M. S., Church, A. T., Guanzon-Lapenã, M. A., Carlota, A. J., and del Pilar, G. H. (2002). Are indigenous personality dimensions culture specific? philippine inventories and the five-factor model. *J. Pers. Soc. Psychol.* 82, 89–101. doi: 10.1037/0022-3514.82.1.89

Kotov, R., Gamez, W., Schmidt, F., and Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. *Psychol. Bull.* 136, 768–821. doi: 10.1037/a0020327

Kotov, R., Watson, D., Robles, J. P., and Schmidt, N. B. (2007). Personality traits and anxiety symptoms: the multilevel trait predictor model. *Behav. Res. Ther.* 45, 1485–1503. doi: 10.1016/j.brat.2006.11.011

Leal, P. C., Goes, T. C., da Silva, L. C. F., and Teixeira-Silva, F. (2017). Trait vs. state anxiety in different threatening situations. *Trends Psychiatry Psychother.* 39, 147–157. doi: 10.1590/2237-6089-2016-0044

Lee, K., and Ashton, M. C. (2004). Psychometric properties of the HEXACO personality Inventory. *Multivariate Behav. Res.* 39, 329–358. doi: 10.1207/s15327906mbr3902_8

Li, M., Mobley, W. H., and Kelly, A. (2016). Linking personality to cultural intelligence: an interactive effect of openness and agreeableness. *Pers. Individ. Differ.* 89, 105–110. doi: 10.1016/j.paid.2015.09.050

Lovell, B., and Brown, R. (2017). Burnout in UK prison officers: the role of personality. *Prison J.* 97, 713–728. doi: 10.1177/0032885517734504

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., et al. (2008). Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. *Med. Care.* 46, 266–274. doi: 10.1097/mlr.0b013e318160d093

Macan, T. H., Trusty, M. L., and Trimble, S. K. (1996). Spector's work locus of control scale: dimensionality and validity evidence. *Educ. Psychol. Meas.* 56, 349–357. doi: 10.1177/0013164496056002016

Marsh, H. W., Vallerand, R. J., Lafrenière, M. A. K., Parker, P., Morin, A. J., Carbonneau, N., et al. (2013). Passion: does one scale fit all? Construct validity of two-factor passion scale and psychometric invariance over different activities and languages. *Psychol. Assess.* 25:796. doi: 10.1037/a0032573

McKinney, J., Chang, M. L., and Glassmeyer, D. (2021). Why females choose STEM majors: understanding the relationships between major, personality, interests, self-efficacy, and anxiety. *Int. J. STEM Educ. Res.* 1–23. Advanced online publication. doi: 10.1007/s41979-021-00050-6

Mendiburo-Seguel, A., Páez, D., and Martínez-Sánchez, F. (2015). Humor styles and personality: a meta-analysis of the relation between humor styles and the big five personality traits. *Scand. J. Psychol.* 56, 335–340. doi: 10.1111/sjop.12209

Morton, N., Hill, C., and Meiring, D. (2018). Validating the South African Personality Inventory (SAPI): examining green behavior and job crafting within a nomological network of personality. *Int. J. Pers. Psychol.* 4, 25–38.

Morton, N., Hill, C., Meiring, D., and De Beer, L. T. (2019). Investigating the factor structure of the South African personality inventory-english version. *S. Afr. J. Ind. Psychol.* 45, 1–13. doi: 10.4102/sajip.v45i0.1556

Muthén, L.K., and Muthén, B.O. (2021). *Mplus User's Guide*, 7th Edn. Los Angeles, CA: Muthén and Muthén.

Naragon-Gainey, K., and Watson, D. (2018). What lies beyond Neuroticism? An examination of the unique contributions of social-cognitive vulnerabilities to internalising disorders. Assessment 25, 143–158. doi: 10.1177/1073191116659741

Nei, K. S., Foster, J. L., Ness, A. M., and Nei, D. S. (2018). Rule breakers and attention seekers: personality predictors of integrity and accountability in leaders. *Int. J. Sel. Assess.* 26, 17–26. doi: 10.1111/ijsa.12201

Nel, J. A., Valchev, V. H., Rothmann, S., Van de Vijver, F. J. R., Meiring, D., and De Bruin, G. P. (2012). Exploring the personality structure in the 11 languages of South Africa. *J. Pers.* 80, 915–948. doi: 10.1111/j.1467-6494.2011.00751.x

Nel, N., Nel, J.A., Adams, B.G., and De Beer, L.T. (2015). Assessing cultural intelligence, personality and identity amongst young white Afrikaans-speaking students: a preliminary study. *SA J. Hum. Resour. Manag.* 13:12. doi: 10.4102/sajhrm.v13i1.643

Oliver, J. E., Jose, P. E., and Brough, P. (2006). Confirmatory factor analysis of the work locus of control scale. *Educ. Psychol. Meas.* 66, 835–851. doi: 10.1177/0013164405285544

Omari, S. M., K'Obonyo, P., and Kidombo, H. (2012). Human resource practices, the relationship between locus of control and employee outcomes. *DBA Afr. Manag. Rev.* 3, 90–99.

Ortet, G., Pinazo, D., Walker, D., Gallego, S., Mezquita, L., and Ibáñez, M. I. (2020). Personality and nonjudging make you happier: contribution of the five-factor model, mindfulness facets and a mindfulness intervention to subjective well-being. *PLoS One* 15:e0228655. doi: 10.1371/journal.pone.0228655

Ott, D. L., and Michailova, S. (2018). Cultural intelligence: a review and new research avenues. *Int. J. Manag. Rev.* 20, 99–119. doi: 10.1111/ijmr.12118

Ozer, D. J., and Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol.* 57, 401–421. doi: 10.1146/annurev.psych.57.102904.190127

Papageorgiou, E., and Callaghan, C. W. (2018). Personality and adjustment in South African higher education accounting studies. *SA J. Account. Res.* 32, 189–204. doi: 10.1080/10291954.2018.1442649

Presbitero, A. (2016). Cultural intelligence (CQ) in virtual, cross-cultural interactions: generalizability of measure and links to personality dimensions and task performance. *Int. J. Intercult. Relat.* 50, 29–38. doi: 10.1016/j.ijintrel.2015.11.001

Qu, W., Sun, H., and Ge, Y. (2020). The effects of trait anxiety and the big five personality traits on self-driving car acceptance. *Transportation* 48, 1-17. doi: 10.1007/s11116-020-10143-7

Rotter, J. B. (1966). Generalised expectancies for internal versus external control of reinforcement. *Psycholog. Monogr.* 80, 1–28. doi: 10.1037/h0092976

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *J. Pers. Soc. Psychol.* 57, 1069–1081. doi: 10.1037/0022-3514.57.6.1069

Ryff, C. D., and Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *J. Pers. Soc. Psychol.* 69, 719–727. doi: 10.1037/0022-3514.69.4.719

Ryff, C. D., and Singer, B. (1996). Psychological well-being: meaning, measurement, and implications for psychotherapy research. *Psychother. Psychosom.* 65, 14–23. doi: 10.1159/000289026

Saviola, F., Pappaianni, E., Monti, A., Grecucci, A., Jovicich, J., and De Pisapia, N. (2020). Trait and state anxiety are mapped differently in the human brain. *Sci. Rep.* 10, 1–11. doi: 10.1038/s41598-020-68008-z

Schepers, J. M., and Hassett, C. F. (2006). The relationship between the fourth edition (2003) of the locus of control inventory and the sixteen personality factor questionnaire (version 5). *S. Afr. J. Ind. Psychol.* 32, 9–18. doi: 10.4102/sajip.v32i2.234

Schmutte, P. S., and Ryff, C. D. (1997). Personality and well-being: re-examining methods and meanings. *J. Pers. Soc. Psychol.* 73, 549–559. doi: 10.1037/0022-3514.73.3.549

Shu, F., McAbee, S. T., and Ayman, R. (2017). The HEXACO personality traits, cultural intelligence, and international student adjustment. *Pers. Individ. Differ.* 106, 21–25. doi: 10.1016/j.paid.2016.10.024

Smidt, W., Kammermeyer, G., Roux, S., Theisen, C., and Weber, C. (2018). Career success of preschool teachers in Germany–the significance of the Big Five personality traits, locus of control, and occupational self-efficacy. *Early Child. Dev. Care* 188, 1340–1353. doi: 10.1080/03004430.2017.1314275

Spector, P. E. (1988). Development of the work locus of control scale. *J. Occup. Psychol.* 61, 335–340. doi: 10.1111/j.2044-8325.1988.tb00470.x

Spector, P. E., Cooper, C. L., Sanchez, J. I., O'Driscoll, M., Sparks, K., Bernin, P., et al. (2002). Locus of control and well-being at work: how generalizable are western findings? *Acad. Manag. J.* 45, 453–466. doi: 10.2307/3069359

Spector, P. E., and O′Connell, B. J. (1994). The contributions of personality traits, negative affectivity, locus of control and Type A to the subsequent reports of job stressors and job strains. *J. Occup. Organ. Psychol.* 67, 1–12. doi: 10.1111/j.2044-8325.1994.tb00545.x

Spitzer, R. L., Kroenke, K., Williams, J. B., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Intern. Med.* 166, 1092–1097. doi: 10.1001/archinte.166.10.1092

The Republic of South Africa (1998). *Employment Equity Act, 1998, Government Gazette, 400 (19740), Cape Town, 19 October 1998*. Available online at: https://www.gov.za/sites/default/files/gcis_document/201409/a55-980.pdf (accessed April 4, 2021).

Valchev, V. H., Van de Vijver, F. J. R., Meiring, D., Nel, J. A., Laher, S., Hill, C., et al. (2014). Beyond agreeableness: social-relational personality concepts from an indigenous and cross-cultural perspective. *J. Res. Pers.* 48, 17–32. doi: 10.1016/j.jrp.2013.10.003

Van de Vijver, F. J. R., and Leung, K. (2021). *Methods and Data Analysis for Cross-Cultural Research*, 2nd Edn. Cambridge: Cambridge University Press.

Van Dierendonck, D. (2004). The construct validity of Ryff's Scales of Psychological Well-Being and its extension with spiritual well-being. *Pers. Individ. Differ.* 36, 629–643. doi: 10.1016/s0191-8869(03)00122-3

Van Dyne, L., Ang, S., and Koh, C. (2015). "Development and validation of the CQS: the cultural intelligence scale," in *Handbook of Cultural Intelligence*, eds S. Ang, and L. Van Dyne (New York, NY: Routledge), 34–56.

Van Jaarsveld, B., and Schepers, J. M. (2007). The personality profile of police office recruits who are high on anxiety. *S. Afr. J. Hum. Resour. Manag.* 5, 51–59. doi: 10.4102/sajhrm.v5i2.119

Van Praag, M., Van der Sluis, J., and Van Witteloostuijn, A. (2004). *The Impact of the Locus-of-Control Personality Trait on the Earnings of Employees vis-à-vis Entrepreneurs. Tinbergen Institute Discussion Paper TI 04-130/3*. Available online at: http://dare.uva.nl/document/14663 (accessed June 7, 2021).

Van Wyk, R., Boshoff, A.B., Nieman, G., Van Vuuren, J., and Pretorius, M. (2009). Prediction of type A behaviour: a structural equation model. *S. Afr. J. Ind. Psychol.* 35:9. doi: 10.4102/sajip.v35i1.758

Vreeke, L. J., and Muris, P. (2012). Relations between behavioral inhibition, big five personality factors, and anxiety disorder symptoms in non-clinical and clinically anxious children. *Child Psychiatry Hum. Dev.* 43, 884–894. doi: 10.1007/s10578-012-0302-5

Wang, C. Y. P., Lien, M. C., Jaw, B. S., Wang, C. Y., Yeh, Y. S., and Kung, S. H. (2019). Interrelationship of expatriate employees' personality, cultural intelligence, cross-cultural adjustment, and entrepreneurship. *Soc. Behav. Pers.* 47, 1–16. doi: 10.2224/sbp.8341

Watson, D., and Naragon-Gainey, K. (2014). Personality, emotions, and the emotional disorders. *Clin. Psychol. Sci.* 2, 422–442. doi: 10.1177/2167702614536162

Watson, D., Nus, E., and Wu, K. D. (2019). Development and validation of the Faceted Inventory of the Five-Factor Model (FI-FFM). Assessment 26, 17–44. doi: 10.1177/1073191117711022

Wilt, J., Oehlberg, K., and Revelle, W. (2011). Anxiety in personality. *Pers. Individ. Differ.* 50, 987–993. doi: 10.1016/j.paid.2010.11.014

Zettler, I., Thielmann, I., Hilbig, B. E., and Moshagen, M. (2020). The nomological net of the HEXACO model of personality: a large-scale meta-analytic investigation. *Perspect. Psychol. Sci.* 15, 723–760. doi: 10.1177/1745691619895036

Ziegler, M., Booth, T., and Bensch, D. (2013). Getting entangled in the nomological net. *Eur. J. Psychol. Assess.* 29, 157–161. doi: 10.1027/1015-5759/a000173

Žitný, P., and Halama, P. (2011). Self-esteem, locus of control, and personality traits as predictors of sensitivity to injustice. *Stud. Psychol.* 53, 27–40.

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.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 12 April 2022
doi: 10.3389/fpsyg.2022.707172

How to Measure the Safety Cognition Capability of Urban Residents? An Assessment Framework Based on Cognitive Progression Theory

Yachao Xiong[1], Changli Zhang[1], Hui Qi[2], Rui Zhang[1] and Yanbo Zhang[3]**

[1] School of Public Policy and Management, China University of Mining and Technology, Xuzhou, China, [2] School of Management, North China Institute of Science and Technology, Langfang, China, [3] School of Management Engineering and Business, Hebei University of Engineering, Handan, China

OPEN ACCESS

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

Reviewed by:
Pin-Chao Liao, Tsinghua University, China
Rainer Leonhart, University of Freiburg, Germany

***Correspondence:**
Hui Qi
msqihui@126.com
Yanbo Zhang
zhangyanbo0530@gmail.com

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

Received: *09 May 2021*
Accepted: *11 March 2022*
Published: *12 April 2022*

Citation:
Xiong Y, Zhang C, Qi H, Zhang R and Zhang Y (2022) How to Measure the Safety Cognition Capability of Urban Residents? An Assessment Framework Based on Cognitive Progression Theory. Front. Psychol. 13:707172. doi: 10.3389/fpsyg.2022.707172

The salience of social risks and the incidence of various crises in China have induced widespread concerns among urban residents. Encountering frequent risks places higher demands on the cognition of urban residents. The concept of safety cognition capability is defined within the context of urban residents' daily life, and measurement instruments are developed and tested to lay the foundation for grasping the current safety cognition capability of urban residents and conducting further research. In this study, the five-dimensional structure of urban residents' safety cognition capability (URSCC) was proposed by using the grounded theory method to sort out the interview transcript of interviews with 30 urban residents, and a 38-item URSCC scale was designed and used for surveys conducted in China. The results show that the scale can be used as a valid tool to measure the URSCC, and it can help city managers to better understand the safety needs of residents, as well as monitor the effectiveness of policy implementation.

Keywords: urban residents, safety cognition capability, conceptual structure, scale development, qualitative analysis

INTRODUCTION

The transition from an industrial to a modern society symbolizes the onset of the "risk society," in which people live with both conventional risks and new man-made uncertainties (Beck, 1992). Cities appear to be the areas with a high incidence of these natural and man-made hazards (Joffe et al., 2013; Singh, 2015). The side effects of urban modernization directly trigger risks or evolve into potential hazards (Frumkin, 2002; Ewing et al., 2016). Urban areas are not only victims but also producers of risks (Hood, 2005). As the coevolution of a sharp urban sprawl and rapid social transition takes place, major cities in China, especially megacities, are facing a surge of social risks and crises, which pose great challenges to local governments (Jinhua, 2018). The city is shrouded in thick smog (Cheng et al., 2017), and the location of some controversial neighboring facilities (Yue et al., 2018) indicate that Chinese urban residents are living in a high environmental hazards context. The continued and rampant public health safety scandals, such as the Sanlu Milk Powder and Changchun Vaccine incidents, vividly show that the Chinese are facing health risks related to food and medical care (Song et al., 2018; Wang and Ding, 2019). The frequent seasonal floods occurring in large cities have severely damaged important infrastructure, such as electricity and transportation, impacting people's daily lives, which is considered one of the most serious natural hazards occurring in Chinese cities. Accidental injuries caused by frequent risks usually occur during driving, in the workplace, and in the home environment (Hazinski et al., 2005). A survey shows that in 97% of emergencies, the first witnesses and community workers arrive at the scene before the professional emergency team (Bogdanski et al., 1999). If the public has a certain understanding of the risk and can handle emergency situations correctly, precious time can be gained. Accordingly, it is imperative to understand the safety cognition of urban residents in the Chinese context and provide a basis for risk mitigation and regulation policies.

Some scholars have put forward the concept of safety cognition capability and explored its measurement dimensions (Eby and Molnar, 2012; Honghai and Xu, 2014). The view they proposed was that safety cognition capability refers to the individual's identification and response to hazards in various activities, emphasizing the consideration of capabilities related to experience, knowledge, individual decisions, and collective behaviors. It is undeniable that accurate judgment and effective responses to hazards are the core elements of safety cognition capability, but one's hazard coping capability cannot be completely equal to one's safety cognition capability. As has been pointed out by the iceberg theory, the classic theory of capability research, capability is not limited to the values of knowledge and skills above the surface. Motivation hidden deep below the surface is the key to distinguishing differences in individual capability (Yu-Jie, 2012). Urban residents' capabilities can be easily observed, e.g., their knowledge, experiences, and behaviors, which are explicit characteristics, but the elements hidden, such as safety values, are rooted in the hearts of residents. These motivations are indispensable for understanding, evaluating, and improving their safety cognition capability.

Cognitive psychology is about processing information (Solso et al., 2005). The model of human information processing stages consists of four stages: sensory processing, perception, response selection, and execution selection (Wickens, 1984). The ladder model further refines the four stages of cognition into eight stages: activation, observation, recognition, interpretation, evaluation, definition of the task, formation of a protocol, and execution (Rasmussen, 1986). The generalized cognitive model divides cognitive processes into three different levels, the skill level, the rule level, and the knowledge level, which are in sequence of increasing levels of cognition. The individuals' cognitive processes are often only on the skill level and rule level (Reason, 1990). In conclusion, cognition has process discontinuity and degree difference, and urban residents' safety cognition also has similar characteristics. Urban residents from different social backgrounds have different cognition of safety, which means they are at different cognitive stages. However, few studies have paid attention to the cognitive gap among different groups. Previous research has focused on how to foster standardized crisis response behaviors among the public. Some researchers have attempted to build a standardized operational procedure for crisis communication that is universally applicable to the public (Fediuk et al., 2010). Standardized policies are also considered to be effective in improving public attitudes and behaviors toward food safety (Ma et al., 2019). In the infrequent scenario of earthquake disasters, disaster risk management agencies should regularly educate the public to maintain belief in the salience of disasters and the importance of preparedness (Zaremohzzabieh et al., 2021). The reality, however, is that there are differences in the upper limits of the individuals' capability to cope with risk and the capabilities needed to deal with hazards across different types of social backgrounds. Due to the constraints of their knowledge, skills, and experience, standardized education and policies of risk and disaster management agencies are not effective in reducing the gap in safety cognition between different groups. An efficient approach is to identify the state of safety cognition capability of various groups and develop targeted measures. Therefore, constructing a stage-based assessment framework to evaluate the safety cognition capabilities of groups from different social backgrounds in different risk situations can target the identification of individuals with deficient capabilities and their cognitive shortcomings.

Based on the statements above, this study introduces safety cognition capability into the daily life of residents and focuses on the development and testing of the urban residents' safety cognition capability (URSCC) scale, specifically: (1) on the basis of existing studies and in-depth interviews, the measurement items of the URSCC scale were refined through qualitative analysis and a preliminary research questionnaire was formed; (2) data were collected through a pre-study, and the scale structure was validated to improve the scale; (3) using formal research data, an exploratory factor analysis and a validation factor analysis were conducted on the scale; (4) the reliability and validity of the URSCC scale were analyzed.

City managers can use the URSCC scale to systematically address the cognitive gaps of residents and formulate targeted policies. This study aims to provide a new perspective for the study of urban residents' safety cognition.

THE CURRENT RESEARCH

Capability is a stable psychological quality that refers to the possibility of an individual achieving various goals (Robbins and Judge, 2013). Studies in the field of psychology, philosophy, and organizational behavior believe that the generation and development of capability must be linked to specific tasks in specific situations (Chien and Tsai, 2012; De Vos et al., 2015; Stephens et al., 2015). Therefore, capability can also be understood as a possibility to accomplish a specific task. The greater the possibility, the stronger the individual's capability. Once a specific task is executed, there are two possibilities, namely success or failure, and the individual's capability is reflected in the transition from inability to ability regarding the task (Chien and Tsai, 2012). Therefore, capability is generally positive (Cavell, 1990). The possibility of the transition from incapability to ability among different individuals varies, that is, there are differences in capabilities between individuals. This conversion encompasses the whole process from the generation of individual capability to the individual's continuous development. Therefore, whether a person can complete a specific task is not only affected by his own knowledge and experience but, more importantly, their perception and attitude toward the task; that is, the value assigned to the task (McClelland, 1973). Values refer to the inherent evaluation of things, the overall view on ideas, customs, and social culture, and the internal generating power of capability (Stern et al., 1999).

The stable development of capability depends on the individual's degree of internal identification with the task. Therefore, whether an individual possesses the value associated with a task is a prerequisite for generating capability. Knowledge and experience are the necessary conditions for the generation and development of capabilities. However, if the individual does not possess values that are consistent with the task goal, even if the individual has perfect knowledge and rich experience, they will not be able to promote the generation of capabilities. Based on this, this article believes that values are the foundation of competence, and knowledge and experience run through the entire competence development process and are important influencing factors for competence development. In addition, feasibility judgments and effective response behaviors based on task recognition are important components of capability.

Capability is not innate. The generation of capability requires behavioral activities as the carrier, which follow the process of value generation, task identification, decision-making, feasibility prediction, and response. In general, there is an upward trend, and the lack of any link will affect the generation and advancement of capabilities. Only the balanced and orderly development of each link can continuously promote the capability to mature. The generation of capability depends on the continuous repetition of the behavior, so the mechanism of individual behavior needs to be considered when discussing the structure of capability. The theory of planned behavior (TPB) will facilitate our exploration of the dimensions of safety cognition capabilities of urban residents. TPB holds that individuals' behavioral decisions are influenced by their psychological characteristics and surroundings and other individuals' behaviors, which means that attitude toward behavior, subjective norm, and perceived behavioral control determine individuals' behavioral intentions. TPB concluded that behavior formation needs to go through three stages, namely, psychological foundation, behavioral intention, and behavior occurrence (Ajzen, 1991). The stage characteristics of behavior formation will contribute to constructing the conceptual structure of URSCC. It is worth noting that there is a significant difference between the capability of an individual to complete a task a single time and the capability to do so multiple times. Repetitive completion of a task will continuously improve the individual's capability. Generally speaking, the capability can gradually sublimate with the continuous development of the individual and eventually form a qualitative change, evolving into a high-quality capability. Therefore, an understanding of capability generation and evolution helps to further explain safety cognition capability.

Cognition can be regarded as a kind of psychological process, including many links, such as perception, thinking, information comparison, and implementation (Mesulam, 1998). After repeating these processes, cognition is transformed in an ascending spiral from the sensible to the rational, and finally to the practical (Stevenson, 2001; Gallese et al., 2004). Safety cognition capability is based on the concept of safe production and is put forward on the basis of general cognition capability. Safety cognition capability refers to people's attitudes, identification, judgment, and response to hazards in various purposeful activities. Previous studies on safety cognition capability are mainly concentrated in the fields of transportation, construction, and coal mines (Hu et al., 2009; Han et al., 2019; Zhang et al., 2020). All recognize the process characteristics of cognition; that is, that the main links of safety cognition include hazard perception, prediction, and response (Guo et al., 2019; Dumbaugh et al., 2020). Safety cognition capability is a special capability. Capability theory holds that the generation and development of any capability must repeat the dynamic process of value formation, information identification, result prediction, and specific response (Helfat and Peteraf, 2003). The safety cognition capability of urban residents also follows this development principle. The implementation of residents' safety behavior is the carrier of their safety cognition capability, which is a specific behavior. The generation of a fixed behavior pattern depends on the stable values driven by individuals, which is also supported by the theory of value-belief-norm (Stern et al., 1999). Safety values are an individual's sensible understanding of the safety climate and constitute the basic premise and core element of urban residents' safety cognition capabilities. In other words, safety values are the threshold level of urban residents' safety cognition capabilities, and the formation of safety values is the embryonic stage of safety cognition capability. After an individual has acquired mature safety values, the first step of safety cognition is the identification of various hazards, which we call the hazard source identification capability, referring to the capability of effectively identifying potential hazard sources after mastering safety knowledge and experience. Therefore, on the basis of the formation of safety values, if an individual has acquired the hazard source identification capability at the same time, this constitutes the perception level of the safety cognition capability of urban residents. The development from a safety value to hazard source identification capability is the formation stage of the safety cognition capability, and hazard identification source capability is a level of safety cognition capability. However, individuals who possess hazard source identification capability are not necessarily able to make safe behavioral choices (Neal and Griffin, 2006). Driven by safety values, individuals can make instantaneous and short-term hazard prediction through decision-making through their hazard source identification capability, which we call their hazard prediction capability. The safety cognition capability develops from the formation stage to the development stage, and the hazard prediction capability is the effective level of safety cognition capability. The progression theory of cognition points out that rational cognition is based on the accumulation of knowledge and the summary of one's experience, and the same is true for the generation and development of one's hazard source identification capability and hazard prediction capability, which constitute the rational stage of urban residents' safety cognition capability.

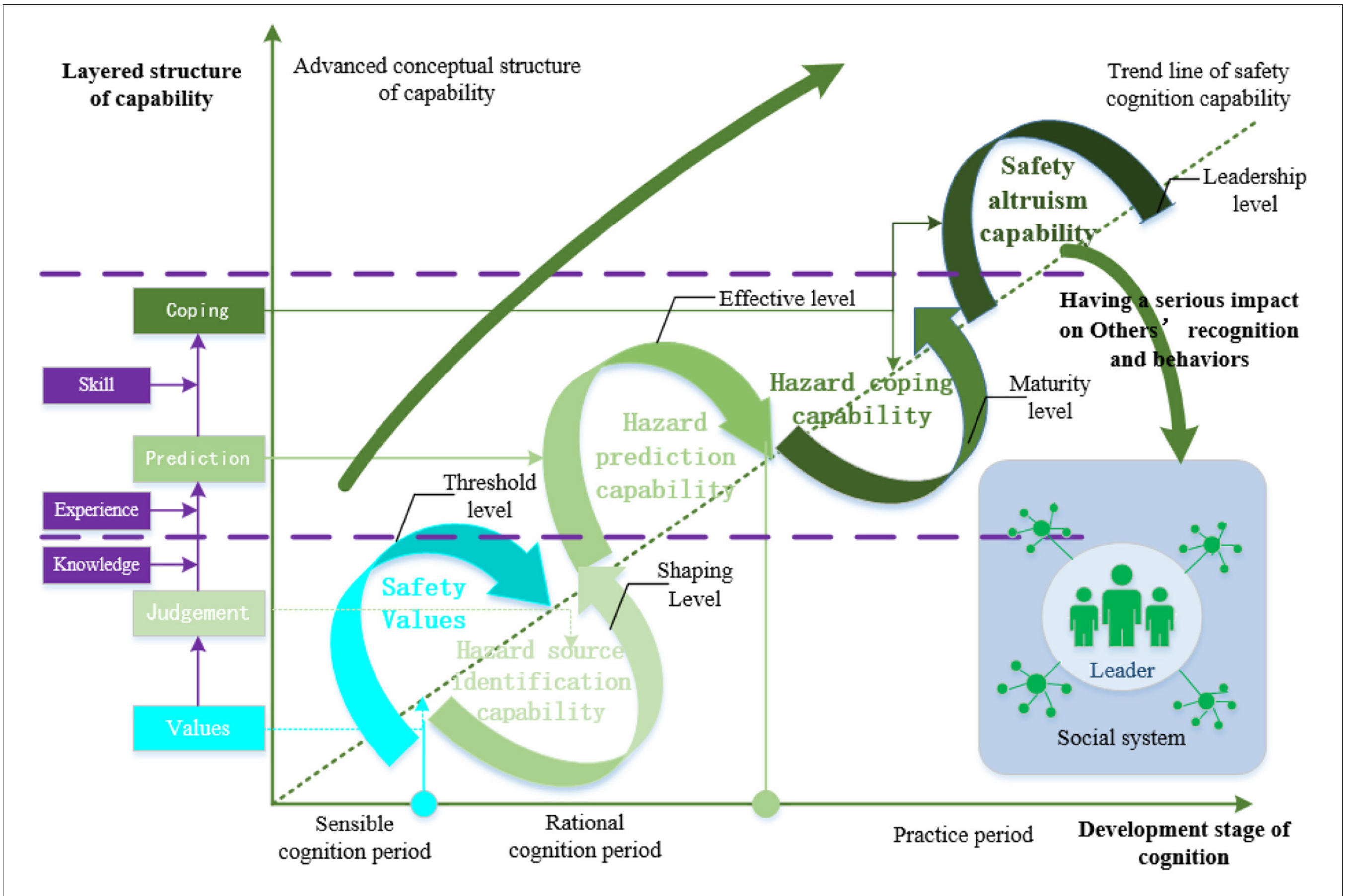

FIGURE 1 | Conceptual structure of safety cognition capability in the whole process.

After predicting the hazards, the individual combines his own safety knowledge and experience to deal with it, that is, the individual uses their hazard coping capability. The evolution from the hazard prediction capability to the hazard coping capability indicates that the development of safety cognition capability has entered a mature period. In addition, safety cognition capability does not stop at individual safety behavior decision-making but spreads the safety values, knowledge, and experience to other people around the individual, who acts as a "missionary" and helps others as a coordinator in the face of public hazards (Jones-Lee, 1991). It can be seen that the linkage effect between the individual and the group, namely one's safety altruism capability, should also be paid attention to when exploring the structure of the safety cognition capability. Although the concept of the safety altruism capability has not been explicitly proposed by scholars, we found that emergency responses of individuals can be adjusted through social practices (Giddens, 1986). We found that safety altruism capability is an intangible but far-reaching safety belief that will continue to affect others and society as a whole. Safety altruism capability is a sublimation of individual capability, and it is the positive diffusion effect of influence between individuals, individuals and groups, and between groups. The performance of safety cognition capability should not stop at one's hazard coping capability but at safety altruism capability as the top level of safety cognition capability. Hazard response capability and safety altruism capability are effective behavioral responses to hazards and constitute the practical stages of the safety cognition capabilities of urban residents. Notable is that not all individuals follow the above capability development process. In real life, there are some individuals who have safety values but do not have hazard identification capabilities, but still show some hazard prediction capabilities, and even show some safety altruism capability, wherein the hazard forecasting and altruistic behavior is an accidental phenomenon with low power and extreme instability. In short, there exists a phenomenon of leapfrogging in some groups regarding urban residents' safety cognition capability. Based on the above analysis, we believe that the generation and development of safety cognition capability follows a process from self-capacity building to group-capacity diffusion, including safety values, hazard source identification capability, hazard prediction capability, hazard coping capability, and safety altruism capability, as shown in **Figure 1**.

MEASUREMENT

Some representative studies on the dimensions and scales of safety cognition capabilities are shown in **Table 1**. Most of the existing studies have taken the safety cognition capability of

TABLE 1 | Safety cognition capability dimension.

References	Dimension	Research object
Abbot et al. (2009)	Self-reported behaviors, psychosocial measures, knowledge	Adolescent
Chen et al. (2011)	Human error, safety performance, accident causes, risk and perception, management actions, safety management and control, accident statistics	Employee
Han et al. (2019)	Implicit social cognition, explicit social cognition, outlet layer artifacts	Employee
Altabbakh (2013)	Safety training, safety knowledge, safety attitude, safety consciousness	College student
Byrd-Bredbenner et al. (2012)	Knowledge, attitudes, self-efficacy.	Middle schoolers
Guo et al. (2019)	Attention, multiple reaction ability, learning ability, short-term memory, performance stability.	Employee
Li and Li (2017)	On-site hazard identification, worker risk behavior identification, occupational safety, regulatory understanding	Employee

specific groups in a single context as the object of study, and few researchers have focused on the URSCC in their daily life and work. In the field of food safety, a measurement framework based on three dimensions of self-reported behavior, psychosocial measures, and knowledge was pioneered (Altabbakh, 2013). Some researchers have argued that psychosocial measures do not fully explain individuals' internal perceptions and evaluations of safety and that self-reported behaviors only reflect some aspects of safety cognitions. Based on this, Byrd proposes to measure food safety cognition along three dimensions: attitude, knowledge, and self-efficacy (Byrd-Bredbenner et al., 2012). The scale developed by Chen is mainly used to measure the safety cognition of construction workers, including human error, safety performance, accident causes, risk perception, management actions, safety management and control, as well as accident statistics, totaling 29 items (Chen et al., 2011). In recent years, several researchers have explored the structure of individuals' safety cognition from the perspective of general abilities. For example, Guo argues that individuals' attention, multiple reaction ability, learning ability, short-term memory, and performance stability constitute their safety cognition capabilities (Guo et al., 2019). Han combines personal and external factors and divides safety cognition into three dimensions: implicit social cognition, explicit social cognition, and outer-layer artifacts (Han et al., 2019). Some researchers have also taken the influence of experience on safety cognition into account by including whether individuals have received safety training as a dimension of safety perceptions. For example, Altabbakh developed a scale to measure safety training, safety knowledge, safety attitude, and safety consciousness (Altabbakh, 2013). The scale developed by Li and Li includes factors, such as on-site hazard identification, worker risk behavior identification, occupational safety, and regulatory understanding of site hazard identification and regulations (Li and Li, 2017).

In terms of an applicable situation, the existing scales mainly focus on the workplace and, as a result, their application areas and situations are restricted. Aside from that, the study focus of other scales has been varied but scattered, with a low degree of recognition and a limited application area and situation. In addition, the URSCC has been enriched through the development of society, and the existing literature is deficient in terms of comprehensive indicators that respond to psychological and individual-group connections. Due to the lack of measurement tools for precision and operability, these scales cannot be directly applied to describing the safety cognition capability of urban residents. Despite these disadvantages, such studies are valuable resources that have led to the development of our scale. We referred to the dimensional settings and statements from the previous scales and modified our self-developed questions with the relevant measurement statements from these scales. For example, for the items of the URSCC scale that relate to the public health domain we refer to this statement: "I eat: raw oysters, clams or mussels, rare hamburgers, raw homemade cookie dough or cake batter, sushi." We replaced the foods mentioned in the statement with foods that are more preferred by Chinese urban residents to ensure the localization of the scale.

Furthermore, grounded theory stresses the use of original data and bridges the gap of theory and reality through methods including literature reviews, interviews, and coding, which can successfully solve flaws in past research in this field. As a result, based on substantial literature research, the URPS scale was developed using a combination of qualitative and quantitative methodologies. We developed the initial scale using grounded theory, and we statistically analyzed the structure of the URPS scale using data acquired from questionnaires.

Initial Scale Construction

To extract the initial question items of the URSCC scale, we conceptualized the specific performance characteristics of the URSCC. We also obtained the initial question items by (1) conducting interviews with selected urban residents and compiling and editing the interviews, and (2) Literature analysis and in-depth analysis of studies on safety cognition and capability evolution to provide a theoretical basis for the scale development.

In addition, these interviews were conducted on the basis of a simple outline that did not include predetermined paradigms and assumptions but which was used as an aid to guide the interviewees' recall and description of the questions, as detailed in **Table 2**. The questions were explained to the interviewees before the interview and could be adjusted during the interview according to the actual situation to elaborate on the topics of the interviewees' responses.

Grounded theory requires that the research subjects are in different age groups, have different education levels, different occupations, and income levels. Therefore, we selected 30 respondents through online publicity. The process of selecting

TABLE 2 | Outline of interview on the safety cognition capability of urban residents.

Theme	Main content
Basic information	Gender, age, education level, address, income level, work level, nature of organization
The current situation of urban residents' sense of safety	a. Do you think it is important to be safe? b. How do you feel about the city you live in?
The structure of URSCC	a. What capabilities do you think are necessary to ensure the safety of yourself and others around you? b. Have you encountered certain risks? And tell us how you responded to them? c. Are there any groups around you that are particularly safety conscious? Describe what qualities they all have in common?

interviewees was carried out according to the theoretical sampling procedure of the grounded theory research method. The method of purposive sampling was used to invite urban residents from different regions as respondents through social software. Considering different cultural backgrounds and regional differences, different groups in eastern, central, and western cities were selected as research participants, including Hebei Province and Jiangsu Province in the eastern region, Anhui Province and Hunan Province in the central region, and Sichuan Province and Xinjiang Province in the western region, giving full consideration to the representativeness of the sample. The basic descriptive statistics of the respondents are as follows: 53% are male and 47% female, 40% are 21–30 years old, 47% are 31–40 years old, and 13% are over 40 years old; 60% of the respondents have a bachelor's degree. In addition, respondents were from cities of different sizes and had different income levels. Details of the interviewees are shown in **Appendix B**. We transformed a representative sample of recordings into text, totaling 42,000 words. In addition, we conducted a theoretical saturation test, which means that when the information obtained from the interview begins to repeat itself and no new important information emerges, the results of the interview have reached theoretical saturation and no further interviews are needed (Glaser and Strauss, 2017). However, the five randomly selected respondents did not provide any new information, which shows that the interview content is theoretically saturated. We invited six researchers to organize the interview texts and collect words and phrases related to safety cognition capabilities. The original statements were then further integrated and simplified by combining them with the literature review.

After initial sorting and categorization, 239 original statements about "safety cognition capability" were collected. Six researchers coded and labeled these expressions and then iteratively discussed them, removing 63 of them that were ambiguous. In view of the diversity of the remaining 176 expressions, we simplified and generalized them based on the analysis of the literature to form specific conceptual indicators. The specific results are summarized in **Table 3**.

TABLE 3 | Classification of semantically similar items.

Original statements	Conceptualization	Frequency
A safe atmosphere is the basic guarantee for my daily life and work; I am willing to preach some safety knowledge; I think a safe atmosphere in the city needs everyone's joint efforts.	Safety values	35
I think the city I live in has more bad weather, serious environmental pollution, frequent man-made accidents, unsafe food, widespread occupational diseases, frequent infectious diseases, unsafe network, and no guarantee of personal safety.	Natural disasters, accidents and disasters, public health events, social security events	30
I think knowing the common hazards can avoid some risks. For some uncertain things, I will understand to ensure their own safety.	Identification of hazard source	21
When I receive a call from an unfamiliar caller and money is involved, I will be vigilant; When buying bagged food, I will pay attention to the date of manufacture and production license.	Hazard prediction	21
When I was followed by a stranger, I quickly moved to a convenience store while calling my family; When I encountered an agitated passenger grabbing the steering wheel on a bus, I stopped it in time; I work in the restaurant industry and can often identify foods that have hygiene problems; When a fire broke out at work, I knew how to use the fire extinguisher and put out the fire in time; Once when a typhoon passed through, I ran to the open outdoor area and was not hit by the collapsed house.	Hazard coping	14
I am surrounded by groups of people who specialize in safety management, who have a high level of safety awareness, are able to anticipate hazards, and are able to correct unsafe behavior in the groups around them; My beloved is a firefighter and often stresses safety awareness to me, and he always handles emergencies appropriately when he encounters them.	Influence and command	11

An individual's level of safety cognition is influenced by various factors, such as their knowledge base, occupation, and social background. The strength of risk perception varies among residents of different social backgrounds (Zhang et al., 2021). It was found that gender, age, ethnicity, education, wealth, job hierarchy, nature of the unit, intelligence, and prestige all have an impact on individuals' safety cognition capability. Based on the collated entries and literature review, we concluded that "gender," "age," "education," "monthly income," and "job level" of urban residents are related to the level of their safety cognition capability, and five questions were formed.

Attitudes toward ideals, customs, and social norms are collectively referred to as values (Aiken, 2010). Values determine individual attitudes, that is, the "stable emergence" of capabilities depends on the individual's internal identification with the task and is essentially determined by a high degree of alignment between values and task goals. Safety cognition capability is a special kind of capability that also follows this rule. We believe that safety cognition capabilities originate from stable safety values, that is, the active maintenance of one's own safety and that of others. Similar expressions were found in the collected entries, such as "safety is a basic guarantee," being "willing to spread safety knowledge," "maintain public safety," and "stop dangerous behavior." This led to the compilation of three scale questions.

The collation revealed that some of the terms were related to urban residents' levels of knowledge about the sources of hazards. The dangers perceived by residents are multifaceted (Yibao Wang et al., 2018), such as "more severe weather (natural disasters)," "frequent man-made accidents (accidents and disasters)," "unsafe food (public health events)," and "life safety is sometimes not guaranteed (social safety events)," which are all sources of hazards that urban residents are exposed to on a daily basis. In addition, we note that many of the collected phrases emphasize the positive effects of having the capability to identify hazards, such as "knowing common hazards can avoid some risks," and that hazard identification is an important part of safety cognition, resulting in 10 scale items.

The study found that the capability to predict risks is an important part of effective safety cognition and that shortening the psychological distance from risk can motivate individuals for this kind of cognition. Combining the frequency of the words and the existing research, the questions of "knowing the level of disaster warning," "being able to recognize the main symptoms of infectious diseases," "being able to recognize crowded people where a trampling accident may occur" were categorized as "hazard prediction capability," and 10 questions were developed.

It has been noted that practice is part of cognition and can correct for biases. Successful risk avoidance experiences can deepen an individual's attitude and understanding of safety, that is, hazard coping is an integral part of safety cognition. According to the collated entries and existing research, the capability to "use fire extinguishers correctly," "getting away from strangers quickly at any time," "knowing how to respond when typhoons pass," and the capability to "distinguish unsanitary food" are attributed to the urban residents' "hazard coping capability" and formed 10 measurement items.

In addition, "I can command and coordinate people around me to deal with danger" appeared three times. The use of this capability should not stop at the individual but has a diffusion effect when it occurs in a broad social group. "I can influence the attitude of the group around me toward safety" and "I can command and coordinate others to respond to hazards" reflect the externalization of individual safety cognition in the group, resulting in the development of three scale items.

The specific steps of the grounded theory analysis include open coding, axial coding, and selective coding (Corbin and Strauss, 1990). During the open coding phase, the researchers debated the statements multiple times before deciding to reclassify them based on semantic similarity and eliminate ambiguous items, leaving 176 statements. Considering the complexity of the remaining 176 statements, at the axial coding step, the researchers integrated and simplified them to create conceptual indicators based on the literature review. Selective coding is a continuation of axial coding at a higher level of abstraction to find out the core category. We followed this criterion to summarize the proposed conceptual indicators, eventually forming a URSCC scale, consisting of 41 items. The purpose of this study is to enhance the theoretical logic and content validity of the structural system of the URSCC through a qualitative research method. In the following section, we will use quantitative analysis to further examine and revise the structural system through data.

QUANTITATIVE METHOD

Preliminary Survey and Extraction of the URSCC Scale

After the initial completion of the URSCC scale, the validity and reliability of the initial scale needed to be analyzed before the formal scale was formed by revising some of the questions. First, through random sampling, researchers promoted and disseminated the web link to the online questionnaire on social media platforms and expanded the number and scope of respondents by continuously forwarding the link. Secondly, in order to make the distribution of the surveyed population reasonable in terms of demographic characteristics, a stratified random sampling method was used to distribute some questionnaires with the help of a professional questionnaire survey website in China. Finally, we compared the selected demographic data with nationally representative demographic data. The demographic data of the survey sample matched well with the national demographic data. At the same time, to ensure the active participation of residents, we provided cash rewards for completing the questionnaire. The preliminary survey was started on 4 February 2020, and a total of 298 questionnaires were collected, of which 53 samples were excluded due to the selection of the same answer for multiple consecutive questions, so that 245 valid questionnaires were obtained, with a valid questionnaire recovery rate of 82.2%. The number of preliminary survey subjects should be three to five times the maximum number of subscale items in the entire scale, and the larger the sample, the better the scale test (DeVellis and Thorpe, 2021). Therefore, the sample size of the preliminary survey should be greater than 30, and the sample size was in line with the standard for scientific research.

First, we conducted reliability tests on the initial scales. Cronbach's alpha coefficient was used to determine the overall reliability of the scale. The results showed that the Cronbach's α value of the URSCC scale was 0.793, indicating that the overall reliability of the scale was acceptable. Item analysis was used to determine the reliability of each item in four ways: (1) Descriptive statistical analysis: Descriptive statistics for each item were used

TABLE 4 | Sample distribution.

Social demographic	variables	Frequency	Percentage	Social demographic	variables	Frequency	Percentage
Gender	Male	404	54.97	Age	20 and below	20	2.72
	Female	331	45.03		21–25	181	24.63
Education	Junior high school and below	17	2.32		26–30	232	31.56
	High School/technical school	66	9.01		31–40	155	21.09
	Junior college	83	11.34		41–50	128	17.41
	Undergraduate	419	57.02		51 and above	19	2.59
	Master's degree	141	19.19	Monthly income(RMB)	<2,000	146	19.82
	Ph.D.	8	1.12		2,000–4,000	185	25.11
Job level	Entry level employee	338	45.95		4,000–6,000	120	16.3
	Grassroots management	158	21.45		6,000–8,000	145	19.86
	Middle management	58	7.93		8,000–10,000	105	14.22
	Senior management	26	3.52		10,000–30,000	22	3.05
	Other	155	21.15		30,000–100,000	12	1.64

to assess the basic quality of the item, and there were no low discrimination items with standard deviations of less than 0.75. (2) Extreme group test: Among the 298 residents surveyed, we selected 27% of the highest total scores and 27% of the lowest total scores and conducted independent sample t-tests for the extreme groups. The t-test values all reached a significance level of 0.05, indicating that each item was effective in identifying high and low scores. (3) Correlation test: Of the 41 questions on the scale, all were significantly correlated with the total score on the scale. (4) Cronbach's α value test: The data showed that the overall reliability of the scale decreased when any of the entries were removed. Thus, 41 items remained in the URSCC scale after item analysis. We conducted a component analysis of these 41 items. During testing, we removed any items with factor loading values of less than 0.5 or with cross-loading values greater than 0.4. After a multi-factor analysis, items 7, 19, and 29 were deleted, and a better discriminant factor structure was obtained. Finally, based on feedback from some respondents and discussions with experts, the linguistic expression of the scale items was improved, thus, further improving the accuracy and clarity of the scale expression and the content validity of the scale summary. We also improved the quality of the initial scale by conducting a pre-study assessment and a formal survey. The final URSCC scale consists of 38 items. The scale was used for the formal research.

Formal Survey and Structural Analysis of the URSCC Scale

The formal investigation was launched in March 2020, and a total of 793 samples with 735 valid survey responses were obtained. For factor analysis, the ratio of the number of items per question to the sample size ranged from approximately 1:5 to 1:10, which was not as important if the total number of subjects was 300 or more (Tinsley and Tinsley, 1987). The structural distribution of the sample is shown in **Table 4**. SPSS20.0 and AMOS16.0 were used to analyze the questionnaire data. The specific analysis is shown in **Table 4**.

Exploratory Factor Analysis

An exploratory factor analysis was performed on half of the sample ($N = 368$) using SPSS 20.0. The KMO value of the scale was $0.909 > 0.8$, and the significance level passed Bartlett's test ($p < 0.001$), indicating that the scale could be subjected to factor analysis. The factor loading matrix was then obtained through principal component analysis and an orthogonal rotation method. As shown in **Table 5**, we selected five eigenvalues greater than 1 based on the Kaiser criterion, with a cumulative variance explained of 60.169%. The definition of each factor is shown in **Table 6**.

Confirmatory Factor Analysis

Using the other half of the data ($N = 367$), the conceptual model obtained by exploratory factor analysis was tested for its fit to the actual observed data. To better verify the accuracy of the model, five competing models are proposed below for comparison with the results of the model obtained from the exploratory factor analysis.

M1: One-factor model, assuming that the common latent variable embraced by the 33 questionnaire items is the URSSC.

TABLE 5 | Exploratory factor analysis results.

Item	Communality	Factor			Item	Communality	Factor	
		S4	S3	S5			S2	S1
URSCC-3	0.753	0.863			URSCC-22	0.579	0.738	
URSCC-2	0.751	0.860			URSCC-18	0.583	0.735	
URSCC-1	0.790	0.843			URSCC-21	0.532	0.718	
URSCC-6	0.747		0.780		URSCC-14	0.573	0.695	
URSCC-13	0.658		0.771		URSCC-16	0.607	0.685	
URSCC-5	0.638		0.770		URSCC-17	0.418	0.679	
URSCC-9	0.542		0.732		URSCC-23	0.578	0.656	
URSCC-11	0.593		0.721		URSCC-15	0.620	0.655	
URSCC-8	0.511		0.714		URSCC-20	0.492	0.614	
URSCC-10	0.582		0.704		URSCC-33	0.603		0.816
URSCC-12	0.494		0.664		URSCC-26	0.563		0.747
URSCC-4	0.612		0.615		URSCC-24	0.628		0.726
URSCC-36	0.703			0.803	URSCC-30	0.507		0.710
URSCC-34	0.684			0.742	URSCC-32	0.501		0.698
URSCC-35	0.739			0.735	URSCC-25	0.580		0.690
					URSCC-28	0.403		0.677
					URSCC-27	0.556		0.629
					URSCC-31	0.735		0.552

TABLE 6 | Definition of each factor.

Factor	Definition
Safety value	Urban residents' attitude, view and internal recognition of safety
Hazard source identification capability	Urban residents' understanding of hazard sources in various fields
Hazard prediction capability	Urban residents' capability to accurately predict danger scenes
Hazard coping capability	Urban residents' capability to continuously and stably effectively deal with various dangerous situations in their daily life and work practice.
Safety altruism capability	Urban residents' capability to influence people around them to improve their safety cognition capability in words or actions

M2: Two-factor model, assuming that 12 items measuring safety values and hazard source identification ability have common latent variables, and 21 items of hazard prediction capability, hazard coping capability, and safety altruism ability have common latent variables.

M3: Three-factor model, assuming that there are common latent variables for 12 items measuring safety values and hazard source identification capability, 18 items measuring hazard prediction capability and hazard coping capability, and 3 items measuring safety altruism capability.

M4: Four-factor model, assuming that 12 items measuring safety values and hazard source identification capability have common latent variables, 9 items measuring hazard prediction capability have common latent variables, 9 items measuring hazard coping capability have common latent variables, and 3 items measuring safety altruism capability have common latent variables.

M5: The five-factor model, based on the results of exploratory factor analysis, assumes five factors for safety values, hazard source identification capability, hazard prediction capability, hazard coping capability, and safety altruism capability.

For each of the above models, the validated factor analysis was conducted with each factor as the latent variable and its corresponding question item as the observed variable. The model fitting results are shown in **Table 7**. The fit results of M1, M2, M3, and M4 are not satisfactory, and the GFI and AGFI of all four models are less than 0.7, while NFI, CFI, TLI, and IFI are less than 0.9, and RMSEA is greater than.07. The χ^2/df of M5 model is 2.828, which is the smallest of the remaining five models, and CFI, TLI, and IFI are all greater than 0.9, so that the first-order model M5 is optimal. However, there are still some indicators that have not reached an excellent level. Once the model parameters were corrected, the correction index was greater than 20 variance coefficients collated, (see **Table 8**).

After three model corrections, the GFI, AGFI, TLI, and CFI values were all greater than 0.9, the RMSEA value was below 0.05, and the χ^2/df value was 2.817. All indicators reached a good range, showing that the model of URSCC has an ideal fit. The standardized path diagram is shown in **Figure 2**.

Reliability and Validity

The assessment of the scale reliability mainly includes two levels: the overall reliability of the scale and the reliability of the latent variables, in which the Cronbach's α value (>0.7) was used to

TABLE 7 | Major fitting degree indices of URSCC.

Model	χ^2	df	χ^2/df	GFI	AGFI	NFI	CFI	TLI	IFI	RMSEA
M1: Single-factor model	6732.452	805	8.363	0.514	0.449	0.503	0.521	0.489	0.522	0.131
M2: Double-factor model	4933.298	804	6.136	0.677	0.633	0.636	0.659	0.636	0.660	0.111
M3: Triple-factor model	4714.427	803	5.871	0.682	0.639	0.652	0.676	0.654	0.677	0.088
M4: Four-factor model	3128.957	799	3.916	0.793	0.798	0.788	0.856	0.834	0.799	0.066
M5: Five-factor model	2254	797	2.828	0.856	0.889	0.851	0.903	0.902	0.909	0.059

TABLE 8 | Overall fitting degree indices of each modification.

Title		Initial model fitting	Release 24-e30	Release e14-e22	Release e11-e12	Assessment
Absolute fitting index	χ^2	2254.216,d f = 797 $P = 0.000$	2243.274, df = 795 $P = 0.000$	2237.125, df = 793 $P = 0.000$	2231.437, df = 792 $P = 0.000$	Great
	GFI	0.856	0.881	0.897	0.909	Great
	RMR	0.066	0.064	0.064	0.061	Good
	RMSEA	0.059	0.057	0.051	0.045	Great
Relative fitting index	AGFI	0.889	0.894	0.899	0.901	Great
	NFI	0.851	0.862	0.871	0.888	Good
	TLI	0.902	0.917	0.922	0.931	Great
	CFI	0.903	0.907	0.911	0.919	Great

test the overall reliability of the scale, and the Cronbach's α value and CR (>0.6) were used to test the reliability of the latent variables. After analyzing the data, it was found that the overall Cronbach's α value of the URSCC scale was 0.928, and that the scale was, thus, reliable as a whole. The Cronbach's α values of the latent variables ranged from 0.799 to 0.901, and the CR values were all above 0.7, both of which were above the acceptable standard, indicating that the scale passed the reliability test.

The assessment of scale validity mainly includes two aspects: content validity and structural validity, in which content validity is mostly measured with qualitative methods, and the validation of structural validity mainly examines the convergent validity and discriminant validity of the scale. In this paper, the initial questionnaire was developed in strict accordance with the scale development procedure, based on a large number of prior studies, and five domain experts were invited to discuss the questionnaire design repeatedly. In total, 298 pre-surveys were conducted, so the content validity of this scale is reliable. In addition, the standardized loadings of the 33 items of the scale on the corresponding latent variables were all greater than 0.5 and reached the significance level, and the corresponding AVE values ranged from 0.581 to 0.701, which satisfied AVE > 0.5, indicating that the convergent validity of the scale was good. In addition, the square roots of the AVEs of the latent variables were all greater than the correlation coefficients between the latent variables, indicating that the potential structural differentiation of the variables was good. The scale passed the validity test. The specific analysis is shown in **Table 9**.

DISCUSSION AND CONCLUSION

Discussion

The URSCC scale measures the safety cognition capability of urban residents regarding the dimensions of safety values, hazard source identification capability, hazard prediction capability, hazard coping capability, and safety altruism capability with objectivity, which truly and clearly reflect the level of urban residents' safety cognition.

The capability theory argues that the emergence and development of any capability follows a dynamic process of value formation, information recognition, outcome prediction, and concrete response, which is repeated over and over again (Wei et al., 2016). Cognition consists of four main processes: information reception, initial analysis, strategy selection, and concrete implementation (Wickens, 1984). In a study of construction workers' cognitions of unsafe behaviors, some scholars have proposed a model of safety cognition that includes four components: hazard identification, reasoning and analysis, decision generation, and implementation response (Goh and Sa'Adon, 2015). This study proposes a safety cognition capability model for urban residents based on the capability theory and the safety cognitive process model.

In addition, this research innovatively proposes two dimensions of safety values and safety altruism capability based on a large number of interviews.

Altabbakh (2013) argues that safety attitudes and awareness are important components of safety cognition capability. However, awareness and attitudes are only the external manifestations of an individual's internal

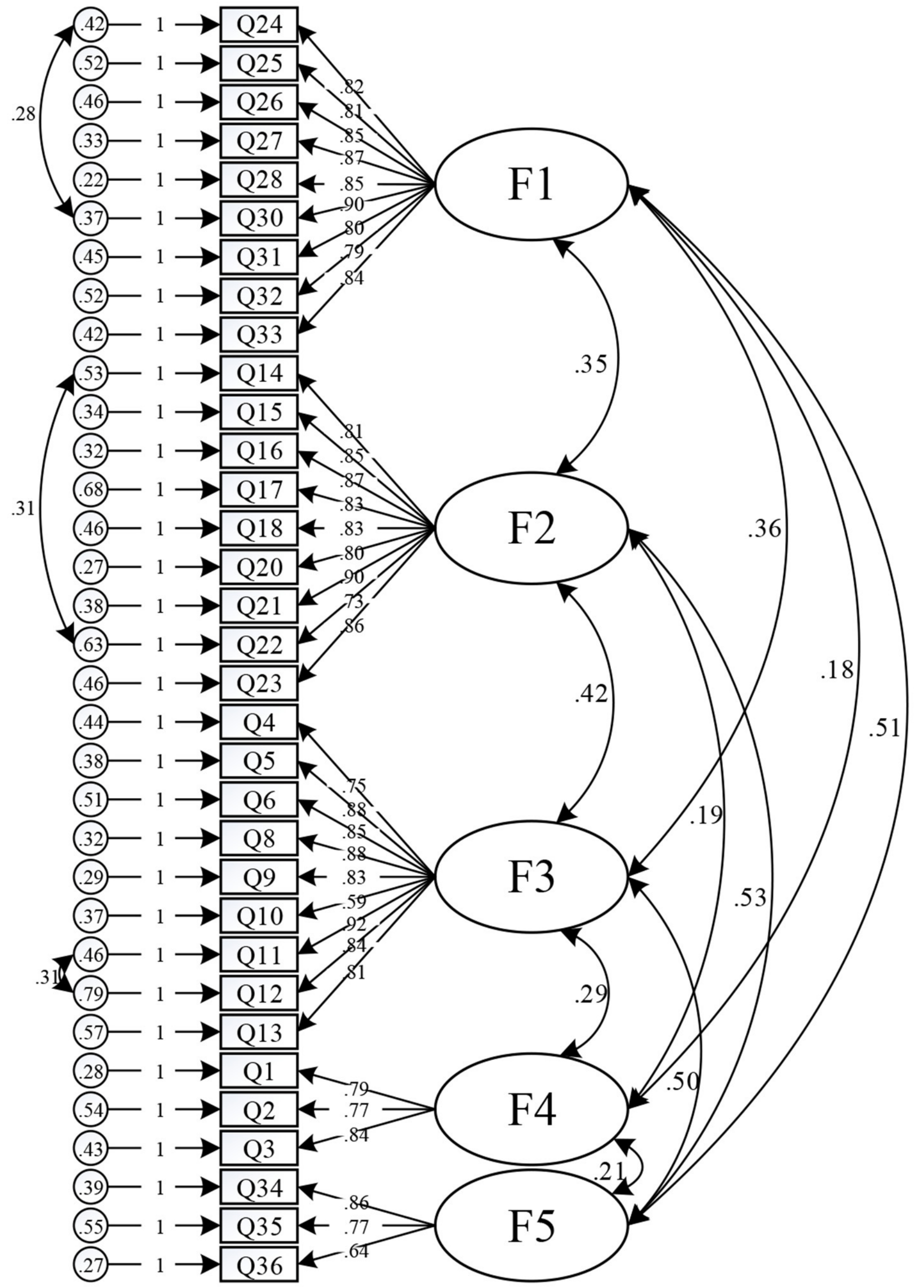

FIGURE 2 | Estimations of the standardized path coefficient of the final confirmatory factor model.

identity, while values are the ultimate origin of behaviors and are the intrinsic motivation for the generation of individual capability. The generation of fixed behavioral patterns depends on stable values within the individual, which is also supported by the "value-idea-norm" theory (Stern et al., 1999). Few researchers have considered the diffusion effect of individual safety cognition capability in groups when developing safety cognition capability scales because the proposed safety altruism capability dimension can also be considered as an innovative contribution of this paper.

In terms of scale applicability, most of the existing scales are applicable to a single context, such as health care settings (Feng et al., 2021), construction sites (Trillo-Cabello et al., 2020), driving (Farrand and Mckenna, 2001), or natural hazards (Crescimbene et al., 2015; Eryilmaz Türkkan and Hirca, 2021). There is no scale that specifically measures the URSCC, and the URSCC scale can be mainly applied to

TABLE 9 | Reliability and validity test of each factor.

	F1	F2	F3	F4	F5
F1	0.889*				
F2	0.494	0.832*			
F3	0.619	0.346	0.822*		
F4	0.488	0.447	0.336	0.801*	
F5	0.557	0.352	0.452	0.395	0.762*
Cronbach'sα	0.897	0.894	0.901	0.845	0.799
CR	0.955	0.953	0.949	0.824	0.804
AVE	0.701	0.693	0.675	0.641	0.581

indicates the square root of the AVE value.

situations that are relevant to the daily lives of urban residents, specifically natural disasters, accidents, public health events, and social security events. In addition, although the survey was conducted in China, the scale is applicable not only to developing countries that have achieved rapid economic growth at the expense of the environment, such as China and India but also to developed countries with strict environmental requirements, such as countries of the European Union and the United States.

Conclusion

First, we conducted in-depth interviews with 30 respondents and developed a URSCC scale consisting of 41 items through qualitative analysis on the basis of existing related studies. Then, we obtained 245 samples through a pre-study, and based on this, we used item analysis and principal component analysis to purify and validate the structure of the scale to finally form a formal research scale of URSCC, consisting of 38 items.

A total of 735 questionnaires was obtained from the formal survey, and we used half of the sample for a principal component analysis to obtain the following six factors: "safety values," "hazard source identification capability," "hazard prediction capability," "ability to respond to hazards," and "safety altruism capability." The KMO of the scale was 0.90, which is greater than 0.7, and the significance was 0.000. The cumulative variance of the six factors was 60.169%. We performed a validated factor analysis of the scale using the other half of the data, and the results showed that the M5 model was superior to the other four models. In addition, we corrected the model parameters because some of the indicators did not meet the requirements. The corrected model had an RMSEA value of 0.059, a χ^2/df value of 2.828, and GFI, AGIF, TLI, CFI, NFI values of 0.856, 0.889, 0.902, 0.903, and 0.851, respectively. The indicators reached the desired range, indicating that the URSCC model has a good fit.

Finally, the reliability of the scale was examined. The Cronbach's α value of the overall reliability of the URSCC scale was 0.928, which is higher than 0.7, and the Cronbach's α values of each latent variable were 0.897, 0.894, 0.901, 0.845, and 0.799, respectively. The CR values were 0.955, 0.953, 0.949, 0.824, and 0.804, respectively. The CRs were 0.955, 0.953, 0.949, 0.824, and 0.804, all of which were within a reasonable range, and the scale had good reliability. In addition, the development of the scale was carried out in strict accordance with the procedures, and the process was rigorous and scientific, which ensured the reliability of the content validity. The standardized loadings of the 33 items of the scale on the corresponding latent variables were all greater than 0.5, and the corresponding AVE values were 0.701, 0.693, 0.675, 0.641, and 0.581, all of which were greater than 0.5. The convergent validity of the scale was also good, and the square roots of the AVEs of the latent variables were greater than the correlation coefficients between the latent variables. The potential structural differentiation of the variables was good, and thus, the scale passed the validity test.

Limitations and Future Studies

The main limitations of this study are as follows: (1) There are local limitations in the sample. During the sampling process, we took the unevenness of urban development levels in China into account, and although the sample was selected to reflect most demographic variables, there were still some areas that could not be covered, and there was no difference in the scales used in cities with different development levels. (2) Since the study focused on urban residents, a large number of rural residents who completed the questionnaire had to be removed, resulting in a lack of comparative analysis of urban and rural residents. (3) The main contribution of this study is the development of the URSCC scale, which has not been empirically tested. Therefore, it is necessary to further validate, revise, and improve the scale. The validity of the scale has only been verified in China. We expect to use this scale to measure and compare the safety cognition capabilities of urban residents in different countries and cities in the future, validating the applicability of the URSCC scale in different countries and regions. Next, we will conduct a large sample survey using the URSCC scale. Then, based on the sample data, we plan to analyze the differences in dimensions and variables across regions to determine whether there are significant differences in the effects of economic development, social development, and technological development on the five main factors in different regions. Meanwhile, studies were conducted in the areas of urban mobility rate, regional integration, and urban crime rates, using the perceived safety capacity of urban residents as a mediating variable.

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 study protocol was reviewed and approved by the Ethics Committee at the Department of Administration, China

University of Mining and Technology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YX obtained the data and wrote the article. CZ conceived the study and performed the field research. HQ contributed to the research design and the data collection. RZ and YZ provided advice for revisions. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the National Natural Science Funding of China (No. 71303233), the Social SciencePlanning Project of Jiangsu Province (No. 18ZZB001), and the Fundamental Research Funds for the Central Universities (No. 3142018050).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.707172/full#supplementary-material

REFERENCES

Abbot, J. M., Byrd-Bredbenner, C., Schaffner, D., Bruhn, C. M., and Blalock, L. (2009). Comparison of food safety cognitions and self-reported food-handling behaviors with observed food safety behaviors of young adults. *Eur. J. Clin. Nutr.* 63, 572–579. doi: 10.1038/sj.ejcn.1602961

Aiken, L. R. (2010). Psychological testing and assessment. *Psyccritiques* 42, 568–568. doi: 10.1117/12.918805

Ajzen, I. (1991). The theory of planned behavior. *Organ. Behav. Hum. Decis. Process.* 50, 179–211. doi: 10.1016/0749-5978(91)90020-t

Altabbakh, H. (2013). *Risk analysis: comparative study of various techniques* (Doctoral thesis). Rolla, MI: Missouri University of Science and Technology.

Beck, U. (1992). From industrial society to the risk society: questions of survival, social structure and ecological enlightenment. *Theory Cult. Soc.* 9, 97–123. doi: 10.1177/026327692009001006

Bogdanski, R., Blobner, M., Hänel, F., and Kochs, E. (1999). Die septische *Enzephalopathie* 34, 123–130. doi: 10.1055/s-1999-178

Byrd-Bredbenner, C., Quick, V., and Corda, K. (2012). Middle schoolers: food safety cognitions and behaviors. *J. Nutr. Educ. Behav.* 44, S51. doi: 10.1016/j.jneb.2012.03.113

Cavell, T. A. (1990). Social adjustment, social performance, and social skills: a tri-component model of social competence. *J. Clin. Child Psychol.* 19, 111–122. doi: 10.1207/s15374424jccp1902_2

Chen, W. T., Lu, C. S., and Huang, Y. H. (2011). Investigating the safety cognition of Taiwan's construction personnel. *J. Mar. Sci. Technol.* 19, 398–408. doi: 10.51400/2709-6998.2181

Cheng, P., Wei, J., Marinova, D., and Guo, X. (2017). Adoption of Protective Behaviours: Residents Response to City Smog in Hefei, China. *J. Conting. Crisis Manag.* 25, 244–255. doi: 10.1111/1468-5973.12148

Chien, S., and Tsai, C. (2012). Dynamic capability, knowledge, learning, and firm performance. *J. Organizat. Change Manag.* 25, 434–444. doi: 10.1108/09534811211228148

Corbin, J., and Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qual. Sociol.* 19, 418–427. doi: 10.1007/BF00988593

Crescimbene, M., Longa, F. L., Camassi, R., and Pino, N. A. (2015). The seismic risk perception questionnaire. *Geol. Soc. Lond. Special Publicat.* 419, 69–77. doi: 10.1144/SP419.4

De Vos, A., De Hauw, S., and Willemse, I. (2015). An integrative model for competency development in organizations: the Flemish case. *Int. J. Hum. Resour. Manag.* 26, 2543–2568. doi: 10.1080/09585192.2014.1003078

DeVellis, R. F., and Thorpe, C. T. (2021). *Scale Development: Theory and Applications:* London: Sage Publications.

Dumbaugh, E., Saha, D., and Merlin, L. (2020). Toward safe systems: traffic safety, cognition, and the built environment. *J. Plann. Educ. Res.* 0739456X20931915. doi: 10.1177/0739456X20931915

Eby, D. W., and Molnar, L. J. (2012). Cognitive impairment and driving safety. *Accident Anal. Prevent.* 49, 261–262. doi: 10.1016/j.aap.2012.05.035

Eryilmaz Türkkan, G., and Hirca, T. (2021). The investigation of flood risk perception as a quantitative analysis from socio-demographic perspective. *Natural Hazards* 106, 715–733. doi: 10.1007/s11069-020-04488-6

Ewing, R., Hamidi, S., and Grace, J. B. (2016). Urban sprawl as a risk factor in motor vehicle crashes. *Urban Stud.* 53, 247–266. doi: 10.1177/0042098014562331

Farrand, P., and Mckenna, F. P. (2001). Risk perception in novice drivers: the relationship between questionnaire measures and response latency. *Trans. Res. F Psychol. Behav.* 4, 201–212. doi: 10.1016/S1369-8478(01)00024-9

Fediuk, T. A., Pace, K. M., and Botero, I. C. (2010). *The Handbook of Crisis Communication:* Malden: Blackwell Publishing.

Feng, P., Yang, H. L., Xu, L., Ojo, O., and Wang, X. H. (2021). Development and psychometric testing of a questionnaire to assess Nurse's perception of risks during enteral. *BMC Nutr.* 20, 6. doi: 10.1186/s12912-020-00520-z

Frumkin, H. (2002). Urban sprawl and public health. *Public Health Rep.* 117, 201–207. doi: 10.1093/phr/117.3.201

Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn. Sci.* 8, 396–403. doi: 10.1016/j.tics.2004.07.002

Giddens, A. (1986). The constitution of societyoutline of the theory of structuration. *Polit. Geogr. Q.* 5, 288–289. doi: 10.1016/0260-9827(86)90040-6

Glaser, B. G., and Strauss, A. L. (2017). *The Discovery of Grounded Theory: Strategies for Qualitative Research.* London: Routledge.

Goh, Y. M., and Sa'Adon, N. B. F. (2015). Cognitive factors influencing safety behavior at height: a multimethod exploratory study. *J. Construct. Eng. Manag.* 141, 04015003.04015001-04015003.04015008. doi: 10.1061/(ASCE)CO.1943-7862.0000972

Guo, M., Hu, L., and Ye, L. (2019). Cognition and driving safety: how does the high-speed railway drivers' cognitive ability affect safety performance? *Transport. Res.* 65, 10–22. doi: 10.1016/j.trf.2019.07.006

Han, Y., Yin, Z., Liu, J., Jin, R., and Yan, L. (2019). Defining and testing a safety cognition framework incorporating safety hazard perception. *J. Construct. Eng. Manag.* 145. 04019081. doi: 10.1061/(ASCE)CO.1943-7862.0001718

Hazinski, M. F., Nadkarni, V. M., Hickey, R. W., O'Connor, R., Becker, L. B., and Zaritsky, A. (2005). Major changes in the 2005 AHA Guidelines for CPR and ECC: reaching the tipping point for change. *Circulation* 112(24 Suppl), IV206. doi: 10.1161/CIRCULATIONAHA.105.170809

Helfat, C. E., and Peteraf, M. A. (2003). The dynamic resource-based view: capability lifecycles. *Strategic Manag. J.* 24, 997–1010. doi: 10.1002/smj.332

Honghai, C., and Xu, M. (2014). Survey and analysis of the cognition of urban residents on the food hygiene and safety in natural disasters. *Appl. Mech. Mater.* 675–677, 233–236. doi: 10.4028/www.scientific.net/AMM

Hood, J. (2005). Emerging risks in the 21st century: an agenda for action. *Risk Manag.* 7, 69–70. doi: 10.1057/palgrave.rm.8240215

Hu, Q., Zhang, S., and Ding, E. (2009). "Environment information cognition and processing in coal mine," in *2009 Fifth International Conference on Semantics, Knowledge and Grid* (Zhuhai: IEEE), 412–413. doi: 10.1109/skg.2009.29

Jinhua, C. (2018). Urban risk generalization in China: space and governance. *Soc. Sci. China* 39, 79–95. doi: 10.1080/02529203.2018.1519215

Joffe, H., Rossetto, T., and Adams, J. (2013). *Cities at Risk: Living with Perils in the 21st Century, Vol. 33*. Berlin: Springer Netherlands.

Jones-Lee, M. W. (1991). Altruism and the value of other people's safety. *J. Risk Uncertain.* 4, 213–219. doi: 10.1007/BF00056126

Li, H., and Li, S. (2017). Research on relationship between safety cognition and unsafe behavior of coal miners. *Safe Coal Mines*. 48, 233–236. doi: 10.13347/j.cnki.mkaq.2017.11.062

Ma, L., Chen, H., Yan, H., Wu, L., and Zhang, W. J. B. P. H. (2019). Food safety knowledge, attitudes, and behavior of street food vendors and consumers in Handan, a third tier city in China. *BMC Public Health* 19, 1–13. doi: 10.1186/s12889-019-7475-9

McClelland (1973). Testing for competence rather than for "intelligence". *Am. Psychol.* 28, 1–14. doi: 10.1037/h0034092

Mesulam, M. M. (1998). From sensation to cognition. *Brain* 121, 1013–1052. doi: 10.1093/brain/121.6.1013

Neal, A., and Griffin, M. A. (2006). A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels. *J. Appl. Psychol.* 91, 946–953. doi: 10.1037/0021-9010.91.4.946

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction. An Approach to Cognitive Engineering*. New York, NY: North-Holland.

Reason, J. (1990). *Human error:* Cambridge: Cambridge University Press.

Robbins, S. P., and Judge, T. A. (2013). *Organizational Behavior, 15th Edn.* Hoboken, NJ: Prentice Hall.

Singh, R. B. (2015). *Urban Development Challenges, Risks and Resilience in Asian Mega Cities.* Tokyo: Springer Japan.

Solso, R. L., MacLin, M. K., and MacLin, O. H. (2005). *Cognitive Psychology*. Boston, MA: Allyn and Bacon.

Song, Y., Yu, H., and Lv, W. (2018). Risk analysis of dairy safety incidents in China. *Food Control* 92, 63–71. doi: 10.1016/j.foodcont.2018.04.007

Stephens, C., Breheny, M., and Mansvelt, J. (2015). Healthy ageing from the perspective of older people: a capability approach to resilience. *Psychol. Health* 30, 715–731. doi: 10.1080/08870446.2014.904862

Stern, P. C., Dietz, T., Abel, T., Guagnano, G. A., and Kalof, L. (1999). A value-belief-norm theory of support for social movements: the case of environmentalism. *Hum. Ecol. Rev*. 6, 81–97. www.jstor.org/stable/24707060

Stevenson, J. C. (2001). The cultural origins of human. *Cognition* 13, 428–428. doi: 10.1002/ajhb.1073

Tinsley, H. E. A., and Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *J. Counsel. Psychol.* 34, 414. doi: 10.1037/0022-0167.34.4.414

Trillo-Cabello, A. F., Carrillo-Castrillo, J. A., and Rubio-Romero, J. C. (2020). Perception of risk in construction. Exploring the factors that influence experts in occupational health and safety. *Safety Sci.* 2021, 104990. doi: 10.1016/j.ssci.2020.104990

Wang, Y., and Ding, Y. (2019). How can china solve the problem of bad vaccines after changchun changsheng? *Biotechnol. Law Rep.* 38, 224–228. doi: 10.1089/blr.2019.29127.yw

Wang, Y., Xu, C., and Cao, M. (2018). *Report of Chinese Urban Safety Sense:* Beijing: Social Science Literature Press.

Wei, J., Chen, H., Cui, X., and Long, R. (2016). Carbon capability of urban residents and its structure: evidence from a survey of Jiangsu Province in china. *Appl. Energy* 173, 635–649. doi: 10.1016/j.apenergy.2016.04.068

Wickens, C. (1984). *Engineering Psychology and Human Performance.* New York, NY: Routledge.

Yue, L. Z., Shu, L. I., and Liang, Z. Y. (2018). New avenues for the development of domain-specific nature of risky decision making. *Adv. Psychol. Sci.* 26, 928–938. doi: 10.3724/SP.J.1042.2018.00928

Yu-Jie, Y. (2012). Problems of quality of migrant workers and countermeasures from the perspective of iceberg model. *J. Anhui Agric. Sci.* 4, 48–50. doi: 10.22004/ag.econ.139474

Zaremohzzabieh, Z., Samah, A. A., Roslan, S., Shaffril, H. A. M., D'Silva, J. L., Kamarudin, S., et al. (2021). Household preparedness for future earthquake disaster risk using an extended theory of planned behavior. *Int. J. Disaster Risk Reduct.* 65, 102533. doi: 10.1016/j.ijdrr.2021.102533

Zhang, Y., Huang, Y., Wang, Y., and Casey, T. W. (2020). Who uses mobile phone while driving for food delivery?: the role of personality, risk perception, and driving self-efficacy. *J. Safety Res.* 73, 69–80. doi: 10.1016/j.jsr.2020.02.014

Zhang, Y., Wang, Y., Ahmad, A. B., Shah, A. A., and Qing, W. (2021). How do individual-level characteristics influence cross-domain risk perceptions among chinese urban residents. *Sage Open* 11, 21582440211003570. doi: 10.1177/21582440211003570

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership