

CLINICAL PSYCHOMETRICS: OLD ISSUES AND NEW PERSPECTIVES

EDITED BY: Michela Balsamo, Marco Innamorati and Dorian A. Lamis
PUBLISHED IN: Frontiers in Psychology





frontiers

Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-956-8

DOI 10.3389/978-2-88945-956-8

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

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

CLINICAL PSYCHOMETRICS: OLD ISSUES AND NEW PERSPECTIVES

Topic Editors:

Michela Balsamo, "G. d'Annunzio" University of Chieti-Pescara, Italy

Marco Innamorati, Università Europea di Roma, Italy

Dorian A. Lamis, Emory University School of Medicine, United States

Clinical Psychometrics can be defined as a discipline that deals with the definition and measurement of clinical constructs. Among its interests, it includes dimensions, such as skills, behavior, psychopathology, quality of life, and personality. Indeed, this discipline focuses on individual differences, the theory of measurement, the construction of measure instruments and their application in an international context.

Clinical Psychometrics can be considered as an essential tool in many fields of research related to psychological and psychiatric interventions: for example, it is useful for diagnostic assessment (in various fields, including clinical and forensic areas), for the design and evaluation of specific psychological and pharmacological treatments. Therefore, Clinical Psychometrics is an applied discipline using psychometric tools to develop evidence-based type procedures relating to the understanding and improvement of the psychological conditions of individuals.

This Research Topic on "Clinical Psychometrics" is interested in several aspects of measurement of psychological variables, focusing on the two fundamental paradigmatic aspects of the discipline, the Classical Test Theory and the Item Response Theory.

This Research Topic seeks to stimulate a scientific debate between psychotherapists and psychometricians in this area. It could have applicative fallouts, such as designing trans-cultural studies in order to: 1) investigate the invariance of new instruments for measuring clinical variables; 2) test the invariance of existing instruments used in clinical research; 3) develop more refined measure instruments for the evaluation of clinical dimensions, similarly to work conducted by the Obsessive Compulsive Cognitions Working Group in identifying domains considered central to OCD and developing the 87-item Obsessive Beliefs Questionnaire; 4) evaluate therapeutic outcomes and processes (such as, states stress, psychological distress, psychological adjustment to illness, health-related quality of life, mood disorders, sexual functioning, etc.).

The goal of this Research Topic is to disseminate a culture of integration between "psychometric model" and "clinical model", promoting the scientific debate about the deepening of the existing methods and/or the proposal of new methods capable of combining clinical significance with quantitative rigor.

This Research Topic welcomed all types of articles, with the exception of case reports. We were particularly interested in:

1. Systematic reviews shedding new lights on the psychometric properties of the most used psychological measures in clinical psychology, neuroscience, psychiatry, psychosomatics, etc.;
2. Guidelines and suggestions on the correct use and gold standards in psychological assessment in the form of research studies and brief reports on the development of new measures and adaptation of existing ones.

Citation: Balsamo, M., Innamorati, M., Lamis, D. A., eds. (2019). *Clinical Psychometrics: Old Issues and New Perspectives*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-956-8

Table of Contents

- 06** *Editorial: Clinical Psychometrics: Old Issues and New Perspectives*
Michela Balsamo, Marco Innamorati and Dorian A. Lamis

SECTION: INVESTIGATING THE MEASUREMENT INVARIANCE OF NEW AND EXISTING INSTRUMENTS FOR MEASURING CLINICAL VARIABLES

- 10** *Consequences of Disregarding Metric Invariance on Diagnosis and Prognosis Using Psychological Tests*
David Blanco-Canitrot, Jesús M. Alvarado and Daniel Ondé
- 13** *Testing Factor Structure and Measurement Invariance Across Gender With Italian Geriatric Anxiety Scale*
Laura Picconi, Michela Balsamo, Rocco Palumbo and Beth Fairfield
- 25** *Dimensions of Anxiety, Age, and Gender: Assessing Dimensionality and Measurement Invariance of the State-Trait for Cognitive and Somatic Anxiety (STICSA) in an Italian Sample*
Leonardo Carlucci, Marley W. Watkins, Maria Rita Sergi, Fedele Cataldi, Aristide Saggino and Michela Balsamo
- 37** *Psychometric Properties and Measurement Invariance of the Brief Symptom Inventory-18 Among Chinese Insurance Employees*
Mingshu Li, Meng-Cheng Wang, Yiyun Shou, Chuxian Zhong, Fen Ren, Xintong Zhang and Wendeng Yang
- 45** *Is Parent–Child Disagreement on Child Anxiety Explained by Differences in Measurement Properties? An Examination of Measurement Invariance Across Informants and Time*
Thomas M. Olino, Megan Finsaas, Lea R. Dougherty and Daniel N. Klein

SECTION: DEVELOPING MORE REFINED INSTRUMENTS FOR MEASURING CLINICAL VARIABLES

- 56** *Assessment of Affect Lability: Psychometric Properties of the ALS-18*
Anna Contardi, Claudio Imperatori, Italia Amati, Michela Balsamo and Marco Innamorati
- 64** *Psychometric Properties of the Cognitive Emotion Regulation Questionnaire (CERQ) in Patients With Fibromyalgia Syndrome*
Albert Feliu-Soler, Elvira Reche-Camba, Xavier Borràs, Adrián Pérez-Aranda, Laura Andrés-Rodríguez, María T. Peñarrubia-María, Mayte Navarro-Gil, Javier García-Campayo, Juan A. Bellón and Juan V. Luciano
- 77** *Confirmatory Factor Analysis of the French Version of the Savoring Beliefs Inventory*
Philippe Golay, Bénédicte Thonon, Alexandra Nguyen, Caroline Fankhauser and Jérôme Favrod
- 84** *Measuring the Capacity to Love: Development of the CTL-Inventory*
Nestor D. Kapusta, Konrad S. Jankowski, Viktoria Wolf, Magalie Chéron-Le Guludec, Madlen Lopatka, Christopher Hammerer, Alina Schnieder, David Kealy, John S. Ogrodniczuk and Victor Blüml

- 97** *Italian Validation of the Capacity to Love Inventory: Preliminary Results*
Giorgia Margherita, Anna Gargiulo, Gina Troisi, Francesca Tessitore and Nestor D. Kapusta
- 104** *The Intolerance of Uncertainty Inventory: Validity and Comparison of Scoring Methods to Assess Individuals Screening Positive for Anxiety and Depression*
Marco Lauriola, Oriana Mosca, Cristina Trentini, Renato Foschi, Renata Tambelli and R. Nicholas Carleton
- 115** *Selfie Expectancies Among Adolescents: Construction and Validation of an Instrument to Assess Expectancies Toward Selfies Among Boys and Girls*
Valentina Boursier and Valentina Manna
- 129** *Psychometric Properties of the Italian Version of the Young Schema Questionnaire L-3: Preliminary Results*
Aristide Saggino, Michela Balsamo, Leonardo Carlucci, Veronica Cavalletti, Maria R. Sergi, Giorgio da Fermo, Davide Dèttore, Nicola Marsigli, Irene Petruccelli, Susanna Pizzo and Marco Tommasi
- 142** *Measuring Intimate Partner Violence and Traumatic Affect: Development of VITA, an Italian Scale*
Gina Troisi
- 156** *The Reliability of the DEM Test in the Clinical Environment*
Alessio Facchin and Silvio Maffioletti
- 167** *Could Time Detect a Faking-Good Attitude? A Study With the MMPI-2-RF*
Paolo Roma, Maria C. Verrocchio, Cristina Mazza, Daniela Marchetti, Franco Burla, Maria E. Cinti and Stefano Ferracuti

SECTION: EVALUATING METHODOLOGICAL ISSUES INVOLVED INTO THERAPEUTIC OUTCOMES AND PROCESSES

- 176** *Using Item Response Theory for the Development of a New Short Form of the Eysenck Personality Questionnaire-Revised*
Daiana Colledani, Pasquale Anselmi and Egidio Robusto
- 189** *Using Overt and Covert Items in Self-Report Personality Tests: Susceptibility to Faking and Identifiability of Possible Fakers*
Giulio Vidotto, Pasquale Anselmi, Luca Filipponi, Marco Tommasi and Aristide Saggino
- 198** *New Perspectives in the Adaptive Assessment of Depression: The ATS-PD Version of the QuEDS*
Andrea Spoto, Francesca Serra, Ivan Donadello, Umberto Granziol and Giulio Vidotto



Editorial: Clinical Psychometrics: Old Issues and New Perspectives

Michela Balsamo^{1*}, Marco Innamorati² and Dorian A. Lamis³

¹ Department of Psychological, Health, and Territorial Sciences, “G. d’Annunzio” University of Chieti-Pescara, Chieti, Italy, ² Department of Human Science, Università Europea di Roma, Rome, Italy, ³ Emory University School of Medicine, Atlanta, GA, United States

Keywords: psychological testing, psychometrics, clinical assessment, quantitative measurement, psychological assessment

Editorial on the Research Topic

Clinical Psychometrics: Old Issues and New Perspectives

Clinical Psychometrics is defined as a discipline that deals with the definition and measurement of clinical constructs. It focuses on the theory of measurement, the construction and validation of psychological measures, and their application in the assessment of individual differences. Therefore, Clinical Psychometrics is an applied discipline, which uses psychometric tools in order to develop evidence-based procedures aimed at understanding and improving the psychological well-being of individuals.

Clinical Psychometrics can be considered as an essential tool in many fields of research related to psychological and psychiatric interventions: for example, it is useful for diagnostic assessment (in various fields, including clinical and forensic areas), and for the design and evaluation of specific psychological and pharmacological treatments.

In the Research Topic “Clinical Psychometrics: Old Issues and New Perspectives,” we were interested in disseminating a culture of integration between the “psychometric model” and the “clinical model,” promoting a scientific debate around existing measures and methods, and proposing new methods capable of combining clinical significance with quantitative rigor (Balsamo et al., 2015a,b).

Therefore, we brought together, within this research topic, contributions from researchers investigating factor invariance of new and existing instruments for measuring clinical variables; research studies developing more refined instruments for the evaluation of clinical dimensions; as well as research studies evaluating methodological issues involved in therapeutic outcomes and processes.

OPEN ACCESS

Edited and reviewed by:

Pietro Cipresso,
Istituto Auxologico Italiano (IRCCS),
Italy

*Correspondence:

Michela Balsamo
m.balsamo@unich.it

Specialty section:

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

Received: 02 April 2019

Accepted: 09 April 2019

Published: 07 May 2019

Citation:

Balsamo M, Innamorati M and
Lamis DA (2019) Editorial: Clinical
Psychometrics: Old Issues and New
Perspectives. *Front. Psychol.* 10:947.
doi: 10.3389/fpsyg.2019.00947

INVESTIGATING THE MEASUREMENT INVARIANCE OF NEW AND EXISTING INSTRUMENTS FOR MEASURING CLINICAL VARIABLES

An area of interest in this Research Topic was the investigation of factor invariance of psychological tests and questionnaires (e.g., Saggino et al., 2017). In fact, psychological tests are frequently administered to different populations and ethnic groups without ever testing the assumption that scores are comparable and interpretable when tests are administered to males and females, adolescents and late adults, or different populations (Balsamo et al., 2015a, 2016, 2018). As reported in “Consequences of disregarding metric invariance on diagnosis and prognosis using psychological tests,” this assumption could have severe consequences when using psychological measures in clinical contexts. In this simulation study, the authors have shown that the lack of

measurement invariance can lead in different samples to over-diagnose a measured condition or diagnose it randomly without any consideration about its real presence.

In this Research Topic, papers directly tested measurement invariance of several questionnaires. For example, two articles investigated factor invariance across sex of two different instruments assessing anxiety severity (“Testing factor structure and measurement invariance across gender with Italian Geriatric Anxiety Scale”); (“Dimensions of anxiety, age, and gender: assessing dimensionality and measurement invariance of the State-Trait for Cognitive and Somatic Anxiety (STICSA) in an Italian sample”). Moreover, in another paper (“Psychometric Properties and Measurement Invariance of the Brief Symptom Inventory-18 Among Chinese Insurance Employees”), the authors investigated factor invariance for the Brief Symptom Inventory-18, a common screening tool for psychological symptoms. Lastly, the paper “Is parent-child disagreement on child anxiety explained by differences in measurement properties? An examination of measurement invariance across informants and time” aimed to longitudinally investigate measurement invariance between maternal and child reports across ages in anxiety assessment. The authors moved from the evidence that agreement between parent-reports of youth and youth self-reports of anxiety problems is modest at best and demonstrated that inter-informant agreement could be compromised for most of the dimensions of anxiety.

DEVELOPING MORE REFINED INSTRUMENTS FOR MEASURING CLINICAL VARIABLES

The majority of the contributions was related to the development and refinement of psychological tests using a transcultural approach. Some Authors presented national adaptation of questionnaires assessing emotional regulation in clinical and non-clinical populations (“Assessment of Affect Lability: Psychometric Properties of the ALS-18,” “Psychometric properties of the Cognitive Emotion Regulation Questionnaire (CERQ) in patients with fibromyalgia syndrome”; “Confirmatory Factor Analysis of the French Version of the Savoring Beliefs Inventory”).

Other papers were illustrative of the psychometric functioning of questionnaires assessing dispositional traits, such as the capacity to love (“Measuring the Capacity to Love: development of the CTL-Inventory,” “Italian validation of the Capacity to Love Inventory: preliminary results”), that is an important diagnostic marker in clinical contexts (e.g., in pathological narcissism), was and a significant outcome parameter of psychotherapeutic treatment; the intolerance of uncertainty (“The Intolerance of Uncertainty Inventory: validity and comparison of scoring methods to assess individuals screening positive for anxiety and depression”), which was found to be associated with a difficulty to tolerate absence of sufficient information and sustain the perception of uncertainty (Carleton, 2016a,b); the expectations correlated with selfies-taking and posting in adolescents (“Selfie expectancies among adolescents: construction and validation

of an instrument to assess expectancies toward selfies among boys and girls”); or assessing the cognitive self-defeating schemas (“Early Maladaptive Schemas” conceptualized by Young et al., 2003), associated with the development of personality disorders and many axis-I disorders (“Psychometric properties of the Italian version of the Young Schema Questionnaire L-3: preliminary results”).

Particularly, instruments such as the CTL-Inventory (Kapusta et al.), composed of six dimensions of the human disposition to establish relationships strictly connected to a person’s psychic development, yields good internal consistency with stable and consistent results in three culturally different (Austrian, Poland, and Italian) samples, and very good test-retest reliability, as well as negative associations with depression, narcissism and promiscuity, and positive associations with relationship qualities such as conflicts, support and depth. Correlated with this disposition, in the paper “Measuring intimate partner violence and traumatic affect: development of VITA, an Italian scale” the author proposed an interesting self-report questionnaire (VITA Scale: Intimate Violence and Traumatic Affects Scale) for measuring intensity of post-traumatic affect, derived from intimate partner violence, the most widespread form of violence against women (World Health Organization [WHO], 2013).

Finally, “The reliability of the DEM test in the clinical environment” paper represents an example of adaptation of psychological test with medical outcome using a transcultural approach. The developmental eye movement (DEM) test could represent a practical and easy method for assessing and quantifying ocular motor skills and evaluating performance over time in children in clinical settings.

One of the common issues for practitioners or those using self-report inventories of personality and psychopathology concerns the susceptibility to malingering or faking. In the “Could time detect a faking-good attitude? A study with the MMPI-2-RF” paper, the authors addressed the role of time in detecting the intentional and deliberate behaviors that helps an individual achieve personal goals (Faking-Good attitude).

EVALUATING METHODOLOGICAL ISSUES INVOLVED INTO THERAPEUTIC OUTCOMES AND PROCESSES

Reflecting state-of-the-art scientific literature, all the papers described above are based on the classical test theory (CTT; Spearman, 1904; Novick, 1965; Gulliksen, 2013). The CTT relies on the evaluation of the reliability, validity, and factor structure of a defined psychological measure (e.g., Innamorati et al., 2013, 2014b, 2015), but within this framework it is impossible to distinguish and compare the parameters related to the individuals (abilities or traits or clinical dimensions, such as depression, anxiety; e.g., Balsamo, 2013; Balsamo and Saggino, 2014; Balsamo et al., 2014) and those relative to the items (difficulties).

Two additional papers presented important contributions from two different methodological frameworks, the Item Response Theory (IRT; Rasch, 1960; Lord, 1980), and the Formal Psychological Assessment (FPA; Spoto, 2011; Spoto et al., 2013).

IRT has been found to offer a useful approach to address some drawbacks of the CTT-based instruments (e.g., to develop new assessment measures to use in psychiatric settings; to shorten full-length tools or refine existing instruments, to address content redundancy). In the paper “Using Item Response Theory for the Development of a New Short Form of the Eysenck Personality Questionnaire-Revised,” the IRT was used to develop a new version of a short form of the Eysenck Personality Questionnaire-Revised (EPQ-R), which includes Psychoticism, Extraversion, Neuroticism, and Lie scales. It outperformed the original instrument (EPQ-R; Eysenck et al., 1985), providing further evidence toward the usefulness of assessing personality traits in clinical settings via IRT.

One intriguing IRT feature concerns the ability to detect respondents in the faking condition from those in the sincere condition. In the study “Using overt and covert items in self-report personality tests: susceptibility to faking and identifiability of possible fakers”, a one-parameter Rasch model, Rasch, 1960; Andrich, 1988) was applied for analyzing items of the alexithymia scale categorized as overt or covert by expert psychotherapists in order to investigate the influence of faking on overt and covert items, and to identify these possible fakers.

An interesting perspective in the assessment of emotional psychopathology was provided by authors of the paper “New perspectives in the adaptive assessment of depression: the ATS-PD version of the QuEDS.” They proposed an Adaptive Testing System for Psychological Disorders (ATS-PD) version of the Qualitative-Quantitative Evaluation of Depressive Symptomatology questionnaire (QuEDS). Adaptive testing could be used to shorten questionnaires without loss of information, reducing the assessment time and focusing on the specific clinical configuration presented from the patients (Petersen et al., 2006).

REFERENCES

- Andrich, D. (1988). *Rasch Models for Measurement*. Beverly Hills, CA: Sage Publications.
- Balsamo, M. (2010). Anger and depression: evidence of a possible mediating role for rumination. *Psychol. Rep.* 106, 3–12. doi: 10.2466/PRO.106.1.3-12
- Balsamo, M. (2013). Personality and depression: evidence of a possible mediating role for anger trait in the relationship between cooperativeness and depression. *Compr. Psychiatry* 54, 46–52. doi: 10.1016/j.comppsy.2012.05.007
- Balsamo, M., Carlucci, L., Sergi, M. R., Murdock, K. K., and Saggino, A. (2015c). The mediating role of early maladaptive schemas in the relation between co-rumination and depression in young adults. *PLoS ONE* 10:e0140177. doi: 10.1371/journal.pone.0140177
- Balsamo, M., Carlucci, L., Sergi, M. R., Romanelli, R., D'Ambrosio, I., and Fairfield, B. (2016). A new measure for trait and state anxiety: the State Trait Inventory of Cognitive and Somatic Anxiety (STICSA). Standardization in an Italian population. *Ital. J. Cogn. Behav. Psychol.* 22, 229–232.
- Balsamo, M., Cataldi, F., Carlucci, L., and Fairfield, B. (2018). Assessment of anxiety in older adults: a review of self-report measures. *Clin. Interv. Aging* 13, 573–593. doi: 10.2147/CIA.S114100

CONCLUSION

Scientists were invited to submit contributions that could facilitate sharing of knowledge among clinicians and researchers engaged in the metric evaluation of clinical phenomena. The ultimate goal is to disseminate a culture of the integration between “psychometric model” and “clinical model,” promoting the scientific debate about the enhancement of the existing methods and/or the proposal of new methods capable of combining clinical significance with the quantitative rigor (Balsamo, 2010; Balsamo et al., 2015c).

Much work needs to be done, but some major issues have been raised by several authors committed to this discipline and have some answers have been obtained in this Research Topic. The response to the call for papers yielded a wealth of proposals with 19 accepted papers by 92 contributing authors.

Our Research Topic included important studies which provide a state-of-the-art scientific compendium of recent and sound psychometric tools useful for improving evidence-based procedures. To the extent that we managed to counter the widespread tendency of the research in clinical psychology and psychiatry to persevere in using inadequate measurement instruments for the diagnosis of disorders and the evaluation process of treatment, we have attained the goal we set ourselves. Only in this way, results derived from clinical research will be no more purely formal and academic, but will have a significant impact on patients' well-being (Nierenberg and Sonino, 2004).

To our delight, several of the articles included have already been accessed thousands of times, indicating a genuine interest in the topics covered.

AUTHOR CONTRIBUTIONS

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

- Balsamo, M., Giampaglia, G., and Saggino, A. (2014). Building a new Rasch-based self-report inventory of depression. *Neuropsychiatr. Dis. Treat.* 10, 153–165. doi: 10.2147/NDT.S53425
- Balsamo, M., Innamorati, M., Van Dam, N. T., Carlucci, L., and Saggino, A. (2015a). Measuring anxiety in the elderly: psychometric properties of the state trait inventory of cognitive and somatic anxiety (STICSA) in an elderly Italian sample. *Int. Psychogeriatr.* 27, 999–1008. doi: 10.1017/S1041610214002634
- Balsamo, M., Macchia, A., Carlucci, L., Picconi, L., Tommasi, M., Gilbert, P., et al. (2015b). Measurement of external shame: an inside view. *J. Pers. Assess.* 97, 81–89. doi: 10.1080/00223891.2014.947650
- Balsamo, M., and Saggino, A. (2014). Determining a diagnostic cut-off on the Teate Depression Inventory. *Neuropsychiatr. Dis. Treat.* 10, 987–995. doi: 10.2147/NDT.S55706
- Carleton, R. N. (2016a). Fear of the unknown: one fear to rule them all? *J. Anxiety Disord.* 41, 5–21. doi: 10.1016/j.janxdis.2016.03.011
- Carleton, R. N. (2016b). Into the unknown: a review and synthesis of contemporary models involving uncertainty. *J. Anxiety Disord.* 39, 30–43. doi: 10.1016/j.janxdis.2016.02.007
- Eysenck, S. B. G., Eysenck, H. J., and Barrett, P. (1985). A revised version of the Psychoticism scale. *Pers. Individ. Diff.* 6, 21–29.
- Gulliksen, H. (2013). *Theory of Mental Tests*. New York, NY: Routledge.

- Innamorati, M., Imperatori, C., Balsamo, M., Tamburello, S., Belvederi Murri, M., Contardi, A., et al. (2014a). Food Cravings Questionnaire-Trait (FCQ-T) discriminates between obese and overweight patients with and without binge eating tendencies: the Italian Version of the FCQ-T. *J. Pers. Assess.* 96, 632–639. doi: 10.1080/00223891.2014.909449
- Innamorati, M., Imperatori, C., Meule, A., Lamis, D. A., Contardi, A., Balsamo, M., et al. (2015). Psychometric properties of the Italian food cravings questionnaire-trait-reduced (FCQ-Tr). *Eat. Weight Disord.* 20, 129–135. doi: 10.1007/s40519-014-0143-2
- Innamorati, M., Lester, D., Balsamo, M., Erbuto, D., Ricci, F., Amore, M., et al. (2014b). Factor validity of the Beck Hopelessness Scale in Italian medical patients. *J. Psychopathol. Behav.* 36, 300–307. doi: 10.1007/s10862-013-9380-3
- Innamorati, M., Tamburello, S., Contardi, A., Imperatori, C., Tamburello, A., Saggino, A., et al. (2013). Psychometric properties of the attitudes toward self-revised in Italian young adults. *Depress. Res. Treat.* 2013:209216. doi: 10.1155/2013/209216
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Nierenberg, A. A., and Sonino, N. (2004). From clinical observations to clinimetrics: a tribute to Alvan R. Feinstein, MD. *Psychother. Psychosomat.* 73, 131–133. doi: 10.1159/000076447
- Novick, M. R. (1965). The axioms and principal results of classical test theory. *ETS Res. Bull. Ser.* 1965, 1–31.
- Petersen, M. A., Groenvold, M., Aaronson, N., Fayers, P., Sprangers, M., Bjorner, J. B., et al. (2006). Multidimensional computerized adaptive testing of the EORTC QLQ-C30: basic developments and evaluations. *Qual. Life Res.* 15, 315–329. doi: 10.1007/s11136-005-3214-z
- Rasch, G. (1960). *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment tests*. Oxford: Nielsen & Lydiche.
- Saggino, A., Carlucci, L., Sergi, M. R., D'Ambrosio, I., Fairfield, B., Cera, N., et al. (2017). A validation study of the psychometric properties of the Other as Shamer Scale–2. *SAGE Open* 7:2158244017704241. doi: 10.1177/2158244017704241
- Spearman, C. (1904). "General Intelligence," objectively determined and measured. *Am. J. Psychol.* 15, 201–292.
- Spoto, A. (2011). *Formal Psychological Assessment Theoretical and Mathematical Foundations*. Doctoral Dissertation, University of Padua. Available online at: <http://paduaresearch.cab.unipd.it/3477>
- Spoto, A., Bottesi, G., Sanavio, E., and Vidotto, G. (2013). Theoretical foundations and clinical implications of formal psychological assessment. *Psychother. Psychosom.* 82, 197–199. doi: 10.1159/000345317
- World Health Organization [WHO] (2013). *Responding to Intimate Partner Violence and Sexual Violence against Women: WHO Clinical and Policy Guidelines*. Geneva: World Health Organization.
- Young, J. E., Klosko, J. S., and Weishaar, M. E. (2003). *Schema Therapy: A Practitioner's Guide*. New York, NY: Guilford Press.

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

Copyright © 2019 Balsamo, Innamorati and Lamis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Consequences of Disregarding Metric Invariance on Diagnosis and Prognosis Using Psychological Tests

David Blanco-Canitrot, Jesús M. Alvarado* and Daniel Ondé

Department of Psychobiology - Behavioral Sciences Methods & Institute of Biofunctional Studies from Complutense University of Madrid, Madrid, Spain

Keywords: invariance, differential item functioning, predictive value of tests, reliability and validity

INTRODUCTION

Guenole and Brown (2014) have shown how failure to meet invariance criteria affects to path coefficients in SEM. In applied research context, these authors suggest testing non-invariance to detect possible undesired effects in the subsequent model evaluation. According to this line of argument, this work intends to show the negative consequences of ignoring the property of invariance when a scale is used with selection or diagnostic purposes.

A scale is invariant when subjects from different groups with the same level on the latent variable have the same probability of obtaining equal test score. However, invariance is not an all-or-nothing judgment. In multi-group Confirmatory Factor Analysis (CFA), four levels of invariance are defined (Meredith, 1993): configural invariance (prerequisite of same factorial structure), metric invariance (MI) or weak invariance (equality of factor loadings), scalar or strong invariance (equality of factor loadings and intercepts), and strict invariance (equality of factor loadings, intercepts and residuals). When a multi-group CFA is conducted, the evaluation of these types of invariance consists on a stepwise procedure from the least restrictive solutions (configural vs. MI) to the most restrictives (MI vs. strong and strong vs. strict), using nested χ^2 tests (Brown, 2015). Consequently, the evaluation of MI is a necessary requirement to compare group scores (Millsap, 2011).

In the parallel model of Classical Test Theory (CTT), MI is directly related to reliability¹. In this model all items have the same standardized factor loading (λ), and the communality (λ^2) is equal to the average correlation of the scale. Consequently, for a scale of n items, reliability of a given value of λ can be calculated from the standardized alpha coefficient: $\alpha = n\lambda^2 / (1 + (n - 1)\lambda^2)$.

Relationship between reliability and predictive validity was first established by Gulliksen (1950) and his attenuation formula. However, the effect of loss of reliability in one of the groups of the sample over the predictive validity is not sufficiently known. What happens when discriminability of some items (i.e., their factor loadings) is different between groups and the instrument is used to make predictions on a dichotomous pass/fail test criterion? How can this MI problem interfere with the correct classification of subjects? This paper aims to explore common practices in applied research that usually ignore MI evaluation (Borsboom, 2006). In this paper, we will try to show the need to reconsider the practical usefulness of psychological tests and scales in decision-making, due to the biased in the correct classification of the subjects.

¹It should be noted that, when data does not fit to the parallel model (i.e., equal true scores and equal standard errors), to estimate reliability it is necessary to know error variances in addition to factorial loadings (see Steenkamp and Baumgartner, 1998).

OPEN ACCESS

Edited by:

Marco Innamorati,
Università Europea di Roma, Italy

Reviewed by:

Claudio Barbaranelli,
Sapienza Università di Roma, Italy

*Correspondence:

Jesús M. Alvarado
jmalvara@ucm.es

Specialty section:

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

Received: 12 December 2017

Accepted: 31 January 2018

Published: 15 February 2018

Citation:

Blanco-Canitrot D, Alvarado JM and
Ondé D (2018) Consequences of
Disregarding Metric Invariance on
Diagnosis and Prognosis Using
Psychological Tests.
Front. Psychol. 9:167.
doi: 10.3389/fpsyg.2018.00167

METHODS

Simulation Procedure

Common values in applied research of reliability and sample size were simulated via Monte Carlo study with 500 sample replications (Harwell et al., 1996) of 100 statistical units for each group ($N = 200$), where factor loadings of a ten-item scale were simulated between 0.44 and 0.50, with an associated reliability of 0.71 and 0.77. In applied psychology, the median sample size of non-students is 200 (Shen et al., 2011). There are between 1 and 10 items per scale in more than 90% of the studies (Hinkin, 1995). To reach Nunnally (1978) recommendation regarding reliability in applied research contexts (minimum of 0.70), for a 10-item scale, a factor loading of 0.44 per item is needed. Following the parallel model, $\alpha = 10(0.44)^2 / (1+9(0.44)^2) = 0.706$.

The database was generated based from the factorial model that is defined in Equation (1).

$$X_{ij} = \sum_{k=1}^k \lambda_{jk} F_k + \sqrt{(1 - \sum_{k=1}^k \lambda_{jk}^2)} \times e_j \quad (1)$$

Where X_{ij} is the simulated response of subject i on a given item j , λ_{jk} is the loading item j in a factor k (which was generated by a unifactorial model), F_k is the latent factor generated by a standardized normal distribution (mean 0 and variance 1) and e_j is the random measurement error of each item.

Predictive validity was evaluated through a generated criterion variable with normal distribution $N(0,1)$, correlation = 0.7 with the 10-item scale, and dichotomized by an established cut point of $Z = 1$ ($p = 1 - 0.8413$), a simulation situation in which only about 15% of subjects with best scores in the criterion have been selected.

Lack of MI was manipulated replacing progressively discriminant items (Group 1) for items with factor loadings equal to zero for the second sample (Group 2). In other words, Differential Item Functioning (DIF) was introduced progressively on the 10-item scale, so that in these items all variance in the second sample would be attributed to error and, thus, all responses would be entirely random.

A Receiver Operating Characteristic (ROC) curve analysis was used to evaluate the effect of the number of items with DIF on the correct classification in criterion variable. This analysis is a fundamental tool to evaluate predictive validity of psychological tests, since allows to detect cases correctly classified in the criterion and identify the cut point that maximizes sensitivity or true positives and specificity or true negatives (Swets and Pickett, 1982).

RESULTS

First row of **Table 1** shows that, when simulated scale have no DIF, sensibility and specificity both Group 1 and Group 2 are between 0.75 and 0.77. Rest of the rows of **Table 1** show the progressive negative effect over sensibility and specificity as the number of DIF items in the scale increases. For example, with 1 DIF item total scale sensitivity is 0.744, with 5 DIF items is 0.710, and with 10 DIF items is 0.605.

This decrease in sensitivity (and specificity) may seem an acceptable loss of discriminative capacity, although overall results are masking its true effects. It can be observed that both Group 2 sensitivity and specificity values have a more pronounced decrease than that observed in the total results. Conversely, in Group 1 sensitivity and specificity increases as the number of items with DIF increases, which is undetectable when observing total results. Both tendencies are undesired effects of lack of MI.

DISCUSSION

In this paper we have exposed that the presence of DIF in the items of a scale implies an important violation of the MI of the instrument, and this lack of MI has significant negative effects on predictive validity.

The results show that when reliability of the scale decreases in one of the subsamples (due to the presence of non-discriminating items), the probability that the subjects of this sample exceed the cut point decreases. When this situation occurs, the cut point for the total sample will also

TABLE 1 | Total, Group 1 and Group 2 sensibility and specificity regarding the number of DIF items manipulated in the simulated 10-item scale.

Reliability	Number of DIF items	Total sensitivity	Group 1 sensitivity	Group 2 sensitivity	Total specificity	Group 1 specificity	Group 2 specificity
0.769	0	0.749	0.748	0.751	0.763	0.773	0.750
0.745	1	0.744	0.756	0.733	0.760	0.773	0.745
0.720	2	0.740	0.764	0.716	0.751	0.775	0.724
0.696	3	0.736	0.772	0.699	0.743	0.783	0.702
0.670	4	0.726	0.783	0.669	0.739	0.794	0.681
0.645	5	0.710	0.788	0.632	0.726	0.804	0.645
0.622	6	0.698	0.801	0.595	0.711	0.811	0.607
0.604	7	0.678	0.807	0.549	0.689	0.820	0.554
0.591	8	0.655	0.812	0.497	0.668	0.821	0.510
0.585	9	0.632	0.819	0.444	0.648	0.827	0.464
0.585	10	0.605	0.825	0.384	0.624	0.833	0.410

decrease and, therefore, subjects of the subsample without DIF will see their options of exceeding the corrected cut point increased.

The loss of discrimination in one or more items from which the lack of MI has been generated is related to non-uniform DIF defined in the Item Response Theory (IRT) framework. Non-uniform DIF usually goes unnoticed as it does not affect the mean of the groups. However, as we have shown in this paper, non-uniform DIF (and consequently, the lack of MI), can have serious consequences when the test is used for predictive or diagnostic purposes. The results imply that one of the two groups (Group 2) would be randomly diagnosed, without any consideration about the real presence of the measured condition, while the other group (Group 1), would be over-diagnosed. Within a selection process, such as an exam, tests scores clearly loses reliability in Group 2 (situation that illegitimately denying the participants any chance of passing the test according to their skills), while increasing those chances on Group 1.

REFERENCES

- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika* 71, 425–440. doi: 10.1007/s11336-006-1447-6
- Brown, T. (2015). *Confirmatory Factor Analysis for Applied Research, 2nd Edn.* New York, NY: Guilford Press.
- Guenole, N., and Brown, A. (2014). The consequences of ignoring measurement invariance for path coefficients in structural equation models. *Front. Psychol.* 5:980. doi: 10.3389/fpsyg.2014.00980
- Gulliksen, H. (1950). *Theory of Mental Tests.* New York, NY: Wiley.
- Harwell, M., Stone, C. A., Hsu, T.-C., and Kirisci, L. (1996). Montecarlo studies in item response theory. *Appl. Psychol. Meas.* 20, 101–125. doi: 10.1177/014662169602000201
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *J. Manage.* 21, 967–988. doi: 10.1177/014920639502100509
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58, 525–543. doi: 10.1007/BF02294825
- Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance.* New York, NY: Routledge.
- Nunnally, J. (1978). *Psychometric Theory.* New York, NY: McGraw-Hill.

Consequently, researchers should be conscious of the serious implications of using scales and tests that might have non-invariant items when approaching diagnostic and selective processes. It is surprising to find through a simple search that, in the 123,000 studies from 2014 to 2017 that are shown in Google Scholar with the term “gender differences,” only 3.73% does the term “metric invariance.” Despite warnings from psychometricians, research works that regards DIF analysis as an important step in the process of developing a scale are scarce, so it becomes this paper’s goal to increase awareness of the necessity and usefulness of such analysis.

AUTHOR CONTRIBUTIONS

DB-C proposed the project, JA developed the theoretical aspects and DB-C performed computations and analysis under JA’s supervision. DO contributed to expand theoretical explanation, as well as interpretation of data. All authors discussed the results and contributed to the final manuscript.

Shen, W., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., and Ones, D. S. (2011). Samples in applied psychology: over a decade of research in review. *J. Appl. Psychol.* 96, 1055–1064. doi: 10.1037/a0023322

Steenkamp, J. E. M., and Baumgartner, H. (1998). Assessing measurement invariance in crossnational consumer research. *J. Cons. Res.* 25, 78–90. doi: 10.1086/209528

Swets, J. A., and Pickett, R. M. (1982). *Evaluation of Diagnostic Systems: Methods From Signal Detection Theory.* New York, NY: Academic Press.

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

Copyright © 2018 Blanco-Canitrot, Alvarado and Ondé. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Testing Factor Structure and Measurement Invariance Across Gender With Italian Geriatric Anxiety Scale

Laura Picconi^{1*}, Michela Balsamo¹, Rocco Palumbo² and Beth Fairfield^{1,3}

¹ Department of Psychological, Health & Territorial Sciences, University of Chieti, Chieti, Italy, ² Department of Neurology, Boston University, Boston, MA, United States, ³ Centro Scienze dell'Invecchiamento e Medicina Traslazionale, University of Chieti, Chieti, Italy

OPEN ACCESS

Edited by:

Pietro Cipresso,
Istituto Auxologico Italiano (IRCCS),
Italy

Reviewed by:

Cesar Merino-Soto,
Universidad de San Martín de Porres,
Peru

Lina Pezzuti,
Sapienza Università di Roma, Italy

*Correspondence:

Laura Picconi
l.picconi@unich.it

Specialty section:

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

Received: 07 March 2018

Accepted: 18 June 2018

Published: 05 July 2018

Citation:

Picconi L, Balsamo M, Palumbo R
and Fairfield B (2018) Testing Factor
Structure and Measurement
Invariance Across Gender With Italian
Geriatric Anxiety Scale.
Front. Psychol. 9:1164.
doi: 10.3389/fpsyg.2018.01164

Late-life anxiety is an increasingly relevant psychiatric condition that often goes unnoticed and/or untreated compared to anxiety in younger populations. Consequently, assessing the presence and severity of clinical anxiety in older adults an important challenge for researchers and clinicians alike. The Geriatric Anxiety Scale is a 30-item geriatric-specific measure of anxiety severity, grouped in three subscales (Somatic, Affective, and Cognitive), with solid evidence for the reliability and validity of its scores in clinical and community samples. Translated into several languages, it has been proven to have strong psychometric properties. In Italy only one recent preliminarily investigative study has appeared on its psychometric properties. However, sample data was largely collected from one specific Italian region (Lombardy) alone. Here, our aim in testing the items of the GAS in a sample of 346 healthy subjects (50% females; 52% from Southern Italy), with mean age of 71.74 years, was 2-fold. First, we aimed to determine factor structure in a wider sample of Italian participants. Confirmatory factor analysis showed that the GAS fits an originally postulated three-factor structure reasonably well. Second, results support gender invariance, entirely supported at the factorial structure, and at the intercept level. Latent means can be meaningfully compared across gender groups. Whereas the means of F1 (Somatic) and F3 (Affective) for males were significantly different from those for females, the means for F2 (Cognitive) were not. More specifically, in light of the negative signs associated with these statistically significant values, the finding showed that F1 and F3 for males appeared to be less positive on average than females. Overall, the GAS displayed acceptable convergent validity with matching subscales highly correlated, and satisfactory internal discriminant validity with lower correlations between non-matching subscales. Implications for clinical practice and research are discussed.

Keywords: geriatric anxiety scale, late-life anxiety, factor structure, measurement invariance, gender differences

INTRODUCTION

Late-life anxiety is an increasingly relevant psychiatric condition and will become an increasing cause of health care utilization, contributing to elevated personal and societal costs, as numbers of older adults constantly increase in diverse countries across the developing world (Wolitzky-Taylor et al., 2010; Baxter et al., 2013). In Italy, for example, 7.3% of the older adults showed symptoms of

chronic anxiety in 2013 (Istat, 2013). Additionally, due to a combination of declining fertility and increased life expectancy, the percentage of people older than 65 years will likely reach 33% of the total population by 2056 (Istat, 2011) and will further increase the percentage of chronic anxiety.

The detection of anxiety disorders in older adults, however, can be complicated by cognitive impairment, newly emergent changes in life circumstances, high age-related medical and psychiatric comorbidity, and a symptom presentation that is markedly different from younger age groups (Magni and DeLeo, 1984; Kogan et al., 2000; Cully et al., 2006; Seignourel et al., 2008; Balsamo et al., 2010; Wolitzky-Taylor et al., 2010; Therrien and Hunsley, 2012). For these reasons, late-life anxiety is more likely to go unnoticed and untreated compared to anxiety in younger populations and makes assessing the presence and severity of clinical anxiety in older adults an important challenge for researchers and clinicians alike. Nonetheless, relatively little is known about the assessment of anxiety in older adults (Ayers et al., 2007; Balsamo et al., 2018).

Among assessment methods adopted for anxiety assessment in both research and clinical practice, self-report measures are by far the most common (Alwahhabi, 2003; Dennis et al., 2007; Antony and Barlow, 2011). Self-report inventories are easy-to-use and time-saving tools for screening psychopathology, measuring the severity of illness, limit patient/participant burden, and for monitoring treatment outcome. Approximately 12 anxiety measures have been identified as frequently used for the assessment of anxiety in older adults (Therrien and Hunsley, 2012). Importantly, most of these measures were originally developed and validated in college samples and therefore lack specific norms and sufficient psychometric evidence for use with older adults. The remaining instruments are new measures created specifically for use with older adults, such as the Geriatric Anxiety Inventory (GAI; Pachana et al., 2007), the Adult Manifest Anxiety Scale-Elderly Version (Reynolds et al., 2003), and the Geriatric Anxiety Scale (GAS; Segal et al., 2010).

Among the age-specific instruments of anxiety, the GAS provided solid evidence for the reliability and validity of its scores in clinical and community samples of older adults in the US (Segal et al., 2010; Yochim et al., 2011, 2013). Already translated in many languages such as German, Persian and Chinese (Bolghan-Abadi et al., 2013; Gottschling et al., 2015; Lin et al., 2016), this questionnaire has been shown to have good psychometric properties among Italian community-dwelling older adults (Gatti et al., 2017). However, its factorial structure has not yet been well-investigated in a large geographically varied sample. Indeed, in the study by Gatti et al. (2017), sample data was largely collected from one specific Italian region (Lombardy) alone.

In light of its promising psychometric properties functioning, including its ability to capture several components of anxiety (somatic, affective, and cognitive symptoms), our study aims to investigate the factor structure of the Italian version of the GAS within the structural equation modeling (Confirmatory Factor Analysis) framework and, to assess internal consistency, convergent and discriminant validity with measures of anxiety,

depression, and personality, in a large Italian sample of healthy community-dwelling older adults. The latter feature of this measure is most important, because it allows clinicians to easily assess whether a patient is experiencing primarily somatic symptoms versus affective or cognitive symptoms, and thus to conclude whether the symptoms are related to a physical health problem instead of an anxiety disorder (Yochim et al., 2011). Moreover, since theoretical and empirical studies have presented mixed results concerning gender differences in experiencing anxiety in older adults (Mueller et al., 2015), we conducted a multiple-group CFA to assess (configural, metric, and scalar) measurement invariance of the GAS and latent means differences across gender groups. Gender, in fact, is a variable which has been identified as a risk factor for anxiety (see, for example, De Beurs et al., 2000; McLean et al., 2011; Mueller et al., 2015). Specifically, women tend to report higher levels of anxiety than men. So, lower scores on GAS scales for males than for females were expected in this sample (Owens et al., 2000).

MATERIALS AND METHODS

Participants and Procedure

Three hundred and forty-six community-dwelling older adults (50% females) from different regions in Italy, recruited from student family members, friends and volunteers, participated in the study. Mean age of the sample was 71.74 ($SD = 6.78$) years. Participants did not receive monetary reimbursement for participation. Exclusion criteria were the presence of current treatment for memory problems, head injuries resulting in hospitalization for more than 24 h and/or medical conditions that could potentially affect cognitive functioning (e.g., Alzheimer's disease, multiple sclerosis, and Parkinson's disease) and, thus, the ability to take the assessment. Moreover, all participants reported being in good mental and physical health.

Initially, 436 questionnaires were returned. Seventeen did not contain answers to all of the GAS items (showing 10% or more missing values). In addition, 73 univariate outliers were detected and removed from the initial dataset by using standard z -score (Tabachnick and Fidell, 2007). Considering levels of education, most participants (27.3%) had a High School diploma and 13.9% a university degree. Most participants came from Central (40.8%) and Southern Italy (52%). Participant characteristics are described in detail in **Table 1**.

For the construct validation of the GAS dimensions, 345 participants from the larger sample also completed the Big-Five Questionnaire 2 (BFQ-2), 327 completed the Teate Depression Inventory (TDI) and 346 completed the Geriatric Anxiety Inventory (GAI). Each participant anonymously completed the questionnaire packet and gave informed consent prior to their inclusion in the study. The study was approved by the Psychological Science Departmental ethical committee at the University of Chieti. All participants provided written, informed consent, in accordance with the Ethical Standards of the Helsinki Declaration.

TABLE 1 | Demographic characteristics of participants ($n = 346$).

Characteristics	
Age (years); mean (SD)	71.74 (6.78)
Sex; % (n)	
Men	50 (173)
Women	50 (173)
Marital status; % (n)	
Single	6.2 (20)
Married	68.6 (223)
Divorced	3.7 (12)
Separated	1.5 (5)
Live-in partner	0.6 (2)
Widowed	19.4 (63)
Education; % (n)	
Primary or lower	36.4 (123)
Upper secondary	22.5 (76)
High School diploma	27.3 (92)
University degree	13.9 (47)
Italian geographic areas; % (n)	
Northern Italy	2.4 (8)
Central Italy	40.8 (136)
Southern Italy	52 (173)
South-islands	4.8 (16)

With the exception of gender, percentages were calculated on the number of subjects who answered questions: 325 for marital status, 338 for Education, 333 for Italian geographic areas.

Measures

Geriatric Anxiety Scale (GAS)

The GAS (Segal et al., 2010) is a 30-item self-report measure used to assess and quantify anxiety symptoms among older adults. Individuals are asked to indicate how often they have experienced each symptom during the immediately preceding week, including today. Respondents answer using a 4-point Likert scale ranging from 0 (not at all) to 3 (always), with higher scores indicating higher levels of anxiety. The GAS includes three theoretically-derived subscales: Cognitive symptoms, Somatic symptoms, and Affective symptoms. The number of items for each subscale ranges from 8 to 9. The GAS total score is based on the first 25 items. The additional 5 content items assess areas of anxiety often reported to be of concern for older adults (health and financial concerns, fear of dying, and so on). These items are for clinical use alone and therefore do not load on the total GAS score.

The GAS was translated from English into Italian through a 6-stage procedure, including an initial translation and a back-translation process carried out by a group of researchers at the University of Bergamo (Gatti et al., 2017). At stage 1, two bilingual translators with Italian mother tongue carried out an independent forward translation. At stage 2, the two translators and a research group discussed and synthesized the results to develop a single forward translation. At stage 3, two bilingual translators with English mother tongue translated the GAS back into English. At stage 4, all translators (2 forward translators + 2 back translators) together with the research group took part in a focus group discussion. Another expert in geriatric psychology,

without any previous knowledge of translation procedures, also participated in the focus group. At stage 5, the pre-final version of the questionnaire was administered to a sample of 15–20 older adults. At stage 6, the research group generated a final report to provide a description of all translations and cultural adaptations made. In the original validation study (Segal et al., 2010), internal consistency of the measure was excellent for the GAS Total score and the 3 Subscales (Total score $\alpha = 0.93$; Cognitive $\alpha = 0.90$; Somatic $\alpha = 0.80$; Affective $\alpha = 0.82$). Cronbach's alphas for the GAS in the present sample were good: 0.88 for Total score, 0.76 for Cognitive scale, 0.77 for Somatic scale, and 0.75 for Affective scale.

Geriatric Anxiety Inventory (GAI)

The Geriatric Anxiety Inventory (Pachana et al., 2007; Italian version by Rozzini et al., 2009) is a 20-item self-report measure used to assess dimensional anxiety among older adults. It has a dichotomous yes/no response format and therefore provides an easy to use response format for mild cognitively impaired older adults. The total score of the GAI ranges from 0 to 20, with higher scores corresponding to higher levels of anxiety. Its internal consistency has been shown to be excellent in samples of community-dwelling older adults and older adults receiving psychiatric services (Andrew and Dulin, 2007; Pachana et al., 2007; Diefenbach et al., 2009; Byrne et al., 2010). Evidence regarding the concurrent validity of the GAI showed moderate to strong correlations with other anxiety measures (Pachana et al., 2007; Yochim et al., 2011) and worry (Pachana et al., 2007; Diefenbach et al., 2009). Divergent validity with measures of depression varied across studies ($r = 0.38$ in Byrne et al., 2010; $r = 0.74$ in Yochim et al., 2011). The Italian version of the GAI exhibited high test-retest reliability ($r = 0.86$), good internal consistency (Cronbach's $\alpha = 0.76$), as well as a high level of concurrent validity with the Anxiety Status Inventory (ASI, Zung, 1971) ($r = 0.85$) (Rozzini et al., 2009). In the present sample, Cronbach's α was 0.90.

Teate Depression Inventory (TDI)

The TDI (Balsamo and Saggino, 2013a; Balsamo et al., 2014b) is a 21-item self-report instrument designed to assess symptoms of Major Depressive Disorder as specified in the latest edition of the DSM (DSM-5, American Psychiatric Association, 2013), in order to overcome psychometric weaknesses of existing measures of depression (Balsamo and Saggino, 2007). It was developed via Rasch logistic analysis of responses, within the framework of Item Response Theory (Rasch, 1960; Andrich, 1995). Each item is rated on a 5-point Likert-type scale, ranging from 0 (always) to 4 (never). Growing literature suggests that the TDI has strong psychometric properties in both clinical and nonclinical samples, including an excellent Person Separation Index, no evidence of bias due to item-trait interaction, good discriminant and convergent validity, and control of major response sets (Balsamo et al., 2013b, 2015a,b,c; Innamorati et al., 2013, 2014; Saggino et al., 2014, 2017; Contardi et al., 2018). Additionally, three cut-off scores were recommended in terms of sensitivity, specificity and classification accuracy for screening for varying levels (minimal, mild, moderate, and severe) of depression severity in a group of patients diagnosed with Major Depressive Disorder

(Balsamo and Saggino, 2014a). In the present sample, Cronbach's alpha was 0.88.

Big Five Questionnaire (BFQ-2)

Personality traits were assessed via the Big Five Questionnaire (BFQ-2; Caprara et al., 1993, 2007) which comprises 134 items rated on a 5-point Likert scale (1 = very false for me, 5 = very true for me). The BFQ has been shown to be a valid and reliable measure of the Big Five traits in large samples of Italian respondents as well as in cross-cultural comparisons (e.g., Caprara et al., 2000). In the present study, the internal consistencies of the five traits were 0.83 (for Extraversion), 0.90 (for Agreeableness), 0.83 (for Conscientiousness), 0.91 (for Openness), and 0.89 (for Emotional Stability).

Data Analysis

Factorial structure of the GAS was examined within the framework of structural equation modeling (CFA) analyzed by EQS 6.0 (Bentler, 2006), allowing for correlation among error terms.

The analyses were performed on covariance matrices, since SEM statistical theory relies on the distributional properties of the elements of a covariance matrix.

The method of estimation used in all models was the robust maximum likelihood estimator, which yields corrected standard errors using the Satorra-Bentler method (Satorra and Bentler, 1994; Rhemtulla et al., 2012). Accordingly, we reported the Satorra-Bentler chi square statistic, with the following robust indices: robust comparative fit index (CFI), robust root mean square error of approximation (RMSEA), and robust standardized root-mean-square residual (SRMR). The following heuristic labels were used to describe model fit: acceptable when CFI was 0.90–0.94, RMSEA was 0.08 and SRMR was 0.08, while good when CFI is equal to or above 0.95, RMSEA is 0.06 or below and SRMR is 0.05 (Hu and Bentler, 1998; Yu, 2002; Byrne, 2006; Steiger, 2007). Lagrange multiplier test (LM) was used to identify which fixed parameters, if freely estimated, would lead to a significantly better fitting model. The LM test operates multivariately in determining misspecified parameters in a model. EQS produces univariate and multivariate χ^2 statistics that permit evaluation of the appropriateness of the specific restrictions; it also yields a parameter change statistic that represents the value that would be obtained if a particular fixed parameter were freely estimated in a future run. Statistically significant LM χ^2 values would argue for the presence of factor cross-loadings and error covariances, respectively. Decisions regarding possible misspecification followed by respecification of the model are based on the incremental univariate statistics. The user typically looks for parameters whose χ^2 values stand apart from the rest and probabilities <0.05 (Byrne, 2006). We used the Expected Parameter Change (EPC) in combination with the Modification Index (MI) (Sarlis et al., 2009). For each parameter tested via the LM Test, the parameter change statistic represents its estimated value if this parameter is freely estimated in a subsequent test of the model. If the EPC is rather small, one concludes that there is no serious misspecification. However,

when the EPC is large, for example larger than 0.2, it is concluded that there is a relevant misspecification in the model.

In addition, Multigroup Confirmatory Factor Analysis (MG-CFA; Meredith, 1993; van de Schoot et al., 2012) was performed to test measurement invariance of the GAS with respect to gender on a set of nested models, that begin with the separate determination of a baseline model for each group. Estimation is based on the robust statistics (ML, robust; the S-B χ^2) and analyses are based on the covariance matrix. The intercepts in addition to variances and covariances will be modeled. Associated with each constraint is a cumulative multivariate LM Test χ^2 value, and an incremental univariate χ^2 value, along with their probability values. To locate parameters that are noninvariant across groups, we look for probability values associated with the incremental univariate χ^2 values that are <0.05 . Invariance was tested for configural (M1), metric (M2) and scalar (M3) invariance. According to Cheung and Rensvold (Cheung and Rensvold, 2000), the Δ CFI is a robust statistic for testing the between-group invariance of CFA models. They recommended that invariance can be assumed when this value is 0.01 or less, in absolute values. Finally, the invariance of Latent Factor Means was to be examined in a CFA framework.

We used the value of the critical ratio (CR) to assess latent mean differences. CR is calculated by parameter estimate divided by its standard error, which tests whether the coefficient is significantly different from 0. A CR value larger than 1.96 indicates statistically significant differences in the latent means (Byrne, 2006).

Using IBM SPSS (2010), internal consistency was estimated by Cronbach's alpha (Cronbach, 1951), McDonald's omega (ω ; Zinbarg et al., 2005; Dunn et al., 2014), and mean corrected item-total correlations. The homogeneity assumption stating that the population variances are equal for gender was tested by Levene's Test (Barbaranelli, 2006). Corrected item-total correlations were calculated to examine how each item contributed to the overall scale. Cronbach's alpha below 0.60 are unacceptable, whereas item inter-correlation coefficients higher than 0.30 are adequate (Nunnally and Bernstein, 1994).

To assess convergent and discriminant validity, relationships between the GAS total, its subscales, and all other measures were investigated using correlation coefficients (Pearson's r). The point biserial correlation (r_{pb}) is the value of Pearson's product moment correlation when one of the variables is dichotomous and the other variable is metric. However, when the values of the two categories of the dichotomous variables are 0 and 1, $r_{pb} = r$ (Pearson's) (p. 143, Ercolani et al., 2001). Mathematically, the Point-Biserial Correlation Coefficient is calculated just as the Pearson's Bivariate Correlation Coefficient would be calculated, where in the dichotomous variable of the two variables is either 0 or 1 - which is why it is also called the binary variable.

This was followed by application of the Fisher r -to- z transformation (Cohen and Cohen, 1983) to examine one-tailed differences in the magnitude of the correlation coefficients to determine whether correlations were significantly different from each other. If r_a is greater than r_b , the resulting value of z will have a positive sign; if r_a is smaller than r_b , the sign of z will be negative.

RESULTS

The descriptive statistics of all GAS items, arranged for the three subscales, are presented in **Table 2**. The means of the 3-point Likert GAS items were relatively low with values ranging from 0.17 (Item 4) to 1.10 (Item 23).

Inspection of skewness and kurtosis indexes indicated that departures from normality were not severe, so no variable transformations were deemed necessary (West et al., 1995).

Confirmatory Factor Analysis, Invariance Measurement and Invariance of Latent Factor Means

Prior to model testing, Mardia's test of normality was used to assess the normality of data by evaluating the kurtosis (Mardia's normalized estimate = 798.113; Mardia, 1974). The high Mardia's normalized estimate of kurtosis suggested non full normality of data. Thus, all analyses were based on the robust maximum likelihood estimator (Satorra and Bentler, 1994).

Confirmatory factor analysis (CFA) was used to validate both the originally postulated three factor structure of the GAS (Model 1: Cognitive, Affective and Somatic; Segal et al., 2010), a one general anxiety factor solution (Model 2), and to test the two-factor structure (Model 3), found by Picconi, Balsamo and Fairfield (report not published, 2017)¹, through a Principal Axis Factoring (PAF) with Direct Oblimin rotation, in which Cognitive/Affective and Somatic factors emerge (see **Table 3**). Goodness-of-fit statistics for all tested structural models were presented in **Table 4**. The SB χ^2 goodness-of-fit tests were significant for each of the CFA models (SB χ^2 ranged from 431.80, $df = 271$, to 406.15, $df = 269$, $p < 0.001$).

Together, results supported both the two factor Cognitive/Affective and Somatic and the one factor solution implied by the GAS item pool.

However, Model 1 (three factor structure) demonstrated significantly better fit compared to Model 2 (one general anxiety factor solution) (Satorra-Bentler Scaled Chi-Square

¹All technical data is available from the authors.

TABLE 2 | Descriptive statistics of the Italian GAS ($N = 346$).

	Item	Subscale	Mean (SD)	Min-Max	Skewness (SE)	Kurtosis (SE)
My heart raced or beat strongly.	1	Somatic	0.66 (0.59)	0–2	0.27 (0.13)	–0.66 (0.26)
My breath was short.	2	Somatic	0.63 (0.60)	0–2	0.38 (0.13)	–0.67 (0.26)
I had an upset stomach.	3	Somatic	0.53 (0.61)	0–2	0.71 (0.13)	–0.45 (0.26)
I had difficulty falling asleep.	8	Somatic	0.91 (0.80)	0–3	0.59 (0.13)	–0.15 (0.26)
I had difficulty staying asleep.	9	Somatic	0.77 (0.81)	0–3	0.75 (0.13)	–0.24 (0.26)
I had a hard time sitting still.	17	Somatic	0.33 (0.56)	0–2	1.45 (0.13)	1.15 (0.26)
I felt tired.	21	Somatic	1.08 (0.67)	0–3	0.38 (0.13)	0.45 (0.26)
My muscles were tense.	22	Somatic	0.64 (0.67)	0–3	0.73 (0.13)	0.22 (0.26)
I had back pain, neck pain, or muscle cramps.	23	Somatic	1.10 (0.78)	0–3	0.41 (0.13)	–0.09 (0.26)
I felt like things were not real or like I was outside of myself.	4	Cognitive	0.17 (0.38)	0–1	1.73 (0.13)	0.99 (0.26)
I felt like I was losing control.	5	Cognitive	0.29 (0.48)	0–2	1.23 (0.13)	0.26 (0.26)
I had difficulty concentrating.	12	Cognitive	0.65 (0.56)	0–2	0.14 (0.13)	–0.74 (0.26)
I felt like I was in a daze.	16	Cognitive	0.25 (0.45)	0–2	1.47 (0.13)	0.90 (0.26)
I worried too much.	18	Cognitive	0.82 (0.67)	0–3	0.34 (0.13)	–0.29 (0.26)
I could not control my worry.	19	Cognitive	0.57 (0.73)	0–3	1.24 (0.13)	1.20 (0.26)
I felt like I had no control over my life.	24	Cognitive	0.18 (0.38)	0–1	1.71 (0.13)	0.92 (0.26)
I felt like something terrible was going to happen to me.	25	Cognitive	0.25 (0.47)	0–2	1.63 (0.13)	1.72 (0.26)
I was afraid of being judged by others.	6	Affective	0.53 (0.59)	0–2	0.62 (0.13)	–0.55 (0.26)
I was afraid of being humiliated or embarrassed.	7	Affective	0.32 (0.53)	0–2	1.34 (0.13)	0.84 (0.26)
I was irritable.	10	Affective	0.75 (0.58)	0–2	0.10 (0.13)	–0.46 (0.26)
I had outbursts of anger.	11	Affective	0.48 (0.60)	0–2	0.84 (0.13)	–0.28 (0.26)
I was easily startled or upset.	13	Affective	0.41 (0.56)	0–2	0.99 (0.13)	0.01 (0.26)
I was less interested in doing something I typically enjoy.	14	Affective	0.45 (0.58)	0–2	0.87 (0.13)	–0.24 (0.26)
I felt detached or isolated from others.	15	Affective	0.34 (0.53)	0–2	1.24 (0.13)	0.55 (0.26)
I felt restless, keyed up, or on edge.	20	Affective	0.62 (0.65)	0–2	0.57 (0.13)	–0.65 (0.26)
Somatic subscale	9		6.65 (3.64)	0–18	0.42 (0.13)	–0.13 (0.26)
Cognitive subscale	8		3.17 (2.60)	0–11	0.71 (0.13)	–0.28 (0.26)
Affective subscale	8		3.91 (2.79)	0–12	0.51 (0.13)	–0.48 (0.26)

Somatic subscale (9 items) = sum of items 1, 2, 3, 8, 9, 17, 21, 22, 23. Cognitive subscale (8 items) = sum of items 4, 5, 12, 16, 18, 19, 24, 25. Affective subscale (8 items) = sum of items 6, 7, 10, 11, 13, 14, 15, 20; SD = Standard Deviation; SE = Standard error.

TABLE 3 | Factor structure extracted-Efa.

ITEM	Extracted factors and loadings (n = 346)		h ²
	COGNITIVE/AFFECTIVE	SOMATIC	
Item 5. I felt like I was losing control.	0.664	-0.072	0.394
Item 24. I felt like I had no control over my life.	0.612	-0.080	0.328
Item 25. I felt like something terrible was going to happen to me.	0.576	0.007	0.336
Item 7. I was afraid of being humiliated or embarrassed.	0.557	-0.049	0.283
Item 6. I was afraid of being judged by others.	0.549	-0.076	0.261
Item 15. I felt detached or isolated from others.	0.531	-0.046	0.258
Item 13. I was easily startled or upset.	0.475	0.162	0.336
Item 19. I could not control my worry.	0.460	0.260	0.409
Item 4. I felt like things were not real or like I was outside of myself.	0.429	-0.042	0.166
Item 16. I felt like I was in a daze.	0.421	0.110	0.239
Item 11. I had outbursts of anger.	0.400	0.118	0.225
Item 14. I was less interested in doing something I typically enjoy.	0.325	0.082	0.141
Item 12. I had difficulty concentrating.	0.297	0.222	0.209
Item 10. I was irritable.	0.272	0.245	0.206
Item 8. I had difficulty falling asleep.	-0.061	0.672	0.411
Item 21. I felt tired.	-0.030	0.618	0.363
Item 9. I had difficulty staying asleep.	-0.059	0.594	0.318
Item 22. My muscles were tense.	0.048	0.576	0.364
Item 23. I had back pain, neck pain, or muscle cramps.	-0.054	0.516	0.239
Item 3. I had an upset stomach.	0.033	0.440	0.211
Item 20. I felt restless, keyed up, or on edge.	0.380	0.423	0.497
Item 18. I worried too much.	0.353	0.379	0.413
Item 2. My breath was short.	0.177	0.326	0.200
Item 1. My heart raced or beat strongly.	0.293	0.322	0.291
Item 17. I had a hard time sitting still.	0.183	0.232	0.133
% explained variance	0.162	0.127	0.289

h² is communality. All factor loadings of ≥ 0.30 are in bold; % of variance explained is in bold.

Difference = 6.84; $df = 21$; $p = 0.998$) (Brown, 2006; Satorra and Bentler, 2010; Barbaranelli and Ingoglia, 2013), and respect to Model 3 (two-factor structure) (Satorra-Bentler

Scaled Chi-Square Difference = 48.14; $df = 2$; $p < 0.001$), with the presence of three error covariances between the items (GAS9 and GAS8, GAS7 and GAS6, GAS25 and GAS24), suggested by Lagrange multiplier test (MI) and by the expected parameter change statistic (EPC) (Saris et al., 1987). Factor loadings, standardized solution of the items and factor structure coefficients, which can be essential for the accurate interpretation of CFA results, are shown in **Table 4** (Graham et al., 2003).

In Model 1, all factor loadings were statistically significant and ranged from 0.36 to 0.75, with an average standardized factor loading of 0.51. Squared multiple correlations ranged from 0.13 to 0.56, with an average SMC of 0.27 indicating that, on average, 27% of the variance in observed variables was accounted for by latent factors. The latent factor correlations were very high, ranging between 0.73 and 0.96. We added also the structure coefficients, which are merely the correlations between the measured variables and the latent factors. Measured variables are correlated with *all* factors when the factors are correlated, even for variables with CFA pattern parameters fixed to be zeroes. The estimation of these structure coefficients does not cost additional degrees of freedom, since the coefficients are fully determined by the pattern and the factor correlation coefficients already being estimated. The structure coefficients are analogous to the zero-order bivariate Pearson correlations without isolating the overlapping relationships among the factors (Thompson, 1997; Graham et al., 2003).

Then, a multiple-group approach was used to test measurement invariance across gender (see **Table 5**).

Measurement invariance across gender groups was entirely supported at the factorial structure, and at the intercept level. The Δ CFIs are lower than 0.01 in all models, suggesting that invariance can be assumed. Based on the establishment of the full scalar invariance across gender, we can compare the latent mean differences across this group. To obtain an estimate of this difference, the female group was chosen as a reference group. Thus, since the female group was designated as the reference group, their factor means were fixed to zero, and we concentrated solely on estimates as they relate to the male group. Because analyses were based on the robust statistics, these estimates are interpreted in terms of robust standard errors and the resulting z-statistics. Accordingly, these results indicate that whereas the means of F1 (Somatic; females = 7.13; males = 6.17; CR = -2.246; small effect size, Cohen's $d^2 = -0.27$) and F3 (Affective; females = 4.20; males = 3.62; CR = -2.128; small effect size, Cohen's $d = -0.21$) for males were significantly different from those for females, the means for F2 (Cognitive; females = 3.35; males = 3.00; CR = -1.332; zero or near zero effect, Cohen's $d = -0.14$) were not. More specifically, considering the negative signs associated with these statistically significant values, the finding showed that F1 and F3 for males appeared to be less positive on average than for females.

A positive CR implies that the comparison group has higher latent mean than the reference group. Conversely, a negative CR suggests that the comparison group's latent mean is smaller than

²Effect sizes were estimated by Cohen's d , where 0.2 is indicative of a small effect, 0.5 a medium and 0.8 a large effect size (Cohen, 1992).

TABLE 4 | Fit indices for the structural models ($N = 346$).

MODEL	SB χ^2	df	CFI	SRMR	RMSEA	90% CI	AIC
Three theoretical factor (Model 1)	406.15*	269	0.93	0.053	0.038	0.031/0.046	-131.85
One factor (Model 2)	418.84*	248	0.91	0.057	0.045	0.037/0.052	-77.15
Two factor, cognitive/affective-somatic (Model 3)	431.80*	271	0.91	0.055	0.041	0.034/0.049	-110.19

FACTOR LOADINGS, STANDARDIZED SOLUTION AND FACTOR STRUCTURE COEFFICIENTS (R_s) -MODEL 1

	Somatic pattern (R_s)	Cognitive pattern (R_s)	Affective pattern (R_s)
Item 1. My heart raced or beat strongly.	0.586 (0.432)	0	0 (0.430)
Item 2. My breath was short.	0.511 (0.377)	0	0 (0.375)
Item 3. I had an upset stomach.	0.465 (0.343)	0	0 (0.341)
Item 8. I had difficulty falling asleep.	0.521 (0.384)	0	0 (0.382)
Item 9. I had difficulty staying asleep.	0.439 (0.324)	0	0 (0.322)
Item 17. I had a hard time sitting still.	0.375 (0.277)	0	0 (0.275)
Item 21. I felt tired.	0.583 (0.430)	0	0 (0.427)
Item 22. My muscles were tense.	0.626 (0.462)	0	0 (0.459)
Item 23. I had back pain, neck pain, or muscle cramps.	0.502 (0.375)	0	0 (0.368)
Item 4. I felt like things were not real or like I was outside of myself.	0	0.358 (0.264)	0 (0.343)
Item 5. I felt like I was losing control.	0	0.560 (0.413)	0 (0.536)
Item 12. I had difficulty concentrating.	0	0.440 (0.325)	0 (0.421)
Item 16. I felt like I was in a daze.	0	0.480 (0.354)	0 (0.459)
Item 18. I worried too much.	0	0.689 (0.508)	0 (0.659)
Item 19. I could not control my worry.	0	0.688 (0.508)	0 (0.658)
Item 24. I felt like I had no control over my life.	0	0.458 (0.338)	0 (0.438)
Item 25. I felt like something terrible was going to happen to me.	0	0.530 (0.391)	0 (0.507)
Item 6. I was afraid of being judged by others.	0 (0.309)	0 (0.404)	0.422
Item 7. I was afraid of being humiliated or embarrassed.	0 (0.328)	0 (0.428)	0.447
Item 10. I was irritable.	0 (0.356)	0 (0.464)	0.485
Item 11. I had outbursts of anger.	0 (0.380)	0 (0.496)	0.518
Item 13. I was easily startled or upset.	0 (0.430)	0 (0.560)	0.586
Item 14. I was less interested in doing something I typically enjoy.	0 (0.270)	0 (0.352)	0.368
Item 15. I felt detached or isolated from others.	0 (0.332)	0 (0.433)	0.453
Item 20. I felt restless, keyed up, or on edge.	0 (0.549)	0 (0.717)	0.749

* $p < 0.001$. SB χ^2 , Satorra and Bentler chi-squared test; df, degrees of freedom; CFI, comparative fit index; SRMR, standardized root mean square residual; RMSEA, root-mean-square error of approximation; 90% CI, 90% confidence interval of RMSEA; AIC, Akaike's information criterion used in the comparison of two or more models with smaller values representing a better fit of the hypothesized model (Hu and Bentler, 1995); Pattern coefficients constrained and not estimated in the model are presented as "0"; the structure coefficients are added in parentheses next to the pattern coefficients.

the reference group (Byrne, 2006). The population variances are equal for all gender groups ($p =$ not significant).

Reliability

Internal consistency of the subscales was good: $\alpha = 0.76$ (95% Confidence Interval: Lower Bound = 0.722; Upper Bound = 0.798; $p < 0.001$; $\omega = 0.81$; Cognitive), $\alpha = 0.77$, (95% Confidence Interval: Lower Bound = 0.732; Upper Bound = 0.805; $p < 0.001$; $\omega = 0.82$; Somatic) and $\alpha = 0.75$ (95% Confidence Interval: Lower Bound = 0.705; Upper Bound = 0.786; $p < 0.001$; $\omega = 0.83$; Affective). Analysis using Feldt's test (see Feldt, 1969; Feldt et al., 1987) indicating that the Cronbach's alpha doesn't significantly differ.

According to the corrected item-total correlations, no items appeared less suitable as indicators of their respective construct.

This means that no item correlations with the scale, excluding the item itself, fall in the low range of 0.0-0.3, and discriminated well (Kline, 1986; Barbaranelli and Natali, 2005; Barbaranelli and D'Olimpio, 2007). The inter-correlations mean of items within each scale ranged from 0.47 (Cognitive) to 0.44 (Affective).

Scale Intercorrelations

As expected, the GAS total scale was positively and strongly correlated with the Cognitive subscale (see **Table 6**; $r = 0.86$, $p < 0.001$, 74% variance shared), Affective subscale ($r = 0.85$, $p < 0.001$, 72% variance shared), and Somatic subscale ($r = 0.85$, $p < 0.001$, 72% variance shared). As expected, the three subscales were highly correlated, with r varying from 0.52 to 0.71 ($p < 0.001$). In addition, the correlation between the Cognitive and Affective subscales was stronger ($r = 0.71$, $p < 0.001$, 50%

TABLE 5 | Test for measurement invariance of the GAS across gender: Summary of goodness of fit statistics.

Model	SB χ^2	df	CFI	RMSEA	RMSEA 90% CI	Model Comparison	Δ CFI
Baseline model males	356.4657	268	0.913	0.044	0.031, 0.055	-----	-----
Baseline model females	360.0953	270	0.903	0.044	0.031, 0.055	-----	-----
M1	716.5637	538	0.899	0.031	0.025, 0.037	-----	-----
M2*	740.1375	562	0.899	0.030	0.024, 0.036	2 vs. 1	0
M3	784.9105	587	0.892	0.031	0.025, 0.037	3 vs. 2	0.007

M1, model for configural invariance; no constraints; M2, model for full metric invariance with all factor loadings constrained equal. M3, model for scalar invariance; with all intercepts constrained equal.

*We included the correlation between errors. Equality constraints are specified for two common error covariance GAS9 and GAS8; GAS7 and GAS6, except the two involving GAS2 and GAS1 and GAS25 and GAS24; unique for males.

TABLE 6 | GAS inter-scale correlations ($n = 346$), correlations with convergent (GAI, $n = 346$; Emotional Stability, $n = 345$) and discriminant scales (TDI, $n = 327$; Extraversion, Openness, Agreeableness, Conscientiousness, $n = 345$).

	S	C	A
Somatic (S)	1		
Cognitive (C)	0.56***	1	
Affective (A)	0.52***	0.71***	1
GAI	0.82***	0.85***	0.83***
TDI	0.39***	0.45***	0.41***
Extraversion	-0.09	-0.17***	-0.01
Openness	-0.17***	-0.17***	-0.10
Agreeableness	0.03	-0.07	-0.10
Conscientiousness	-0.16***	-0.19***	-0.14*
Emotional Stability	-0.33***	-0.42***	-0.48***

S, Somatic Subscale; C, Cognitive Subscale; A, Affective Subscale; GAI, Geriatric Anxiety Inventory; TDI, Teate Depression Inventory. * $p < 0.05$; *** $p < 0.001$.

variance shared) than the correlation between the Cognitive and Somatic subscale ($r = 0.56$, $p < 0.001$, 31% variance shared) and between the Affective and Somatic subscale ($r = 0.52$, $p < 0.001$, 27% variance shared).

Convergent and Discriminant Validity of the Gas

To investigate the convergent and discriminant validity of the Italian version of the GAS, correlations among the GAS total and its subscales with measures of depression, anxiety and personality were computed (see **Table 6**).

The correlation of the depression scale (TDI) with all anxiety dimensions was weaker than the correlation between measure of anxiety (TDI with GAI, $r = 0.48$). As seen in **Table 5**, the GAS total score and GAS subscale scores were significantly positively correlated with the TDI, with medium effect sizes (GAS total, $r = 0.49$; Cognitive, $r = 0.45$; Somatic, $r = 0.39$; Affective, $r = 0.41$) and with the GAI with high effect size (GAS total, $r = 0.97$; Cognitive, $r = 0.85$; Somatic, $r = 0.82$; Affective, $r = 0.83$).

Compared to the anxiety scale (GAI), the correlation of 0.49 between the GAS total and the TDI was significantly lower than the correlation of 0.97 between the GAS total and GAI

($z = -21.04$, $p < 0.001$). The correlation of 0.39 between the Somatic subscale and the TDI was significantly lower than the correlation of 0.82 between the Somatic subscale and the GAI, ($z = -9.48$, $p < 0.001$). The correlation of 0.45 between the Cognitive subscale and TDI was significantly lower than the correlation of 0.85 between the Cognitive subscale and the GAI, ($z = -9.96$, $p < 0.001$). The correlation of 0.41 between the Affective subscale and the TDI was significantly lower than the correlation of 0.83 between the Affective and GAI, ($z = -9.53$, $p < 0.001$). Also, GAS total score and GAS subscale scores were substantially correlated with Emotional Stability (GAS total, $r = -0.47$; Cognitive, $r = -0.42$; Somatic, $r = -0.33$; Affective, $r = -0.48$).

However, the discriminant correlations with the other subscales of the BFQ-2 were rather low and only a few appeared to be significant ($p < 0.001$) (GAS total, ranging from $r = -0.19$ for Conscientiousness to $r = -0.04$ for Agreeableness; Cognitive, ranging from $r = -0.19$ for Conscientiousness to $r = -0.07$ for Agreeableness; Somatic, ranging from $r = -0.16$ for Conscientiousness to $r = 0.03$ for Agreeableness; Affective, ranging from $r = -0.14$ for Conscientiousness to $r = -0.01$ for Extraversion).

Gas Content Items

Finally, as stated above, the GAS includes five additional content items (items 26-30) that do not load on any scales but are used for clinical purposes and provide information about areas of anxiety often reported to be of concern for older adults (e.g., fear of dying, financial or health concerns; Segal et al., 2010). These scores are not included in the GAS total score.

A rank order of the means of these five content items showed that item 28 was the highest ranked item ("I was concerned about my children", $M = 1.50$, $SD = 1.02$), followed by item 27 ("I was concerned about my health", $M = 1.01$, $SD = 0.75$), item 26 ("I was concerned about my finances", $M = 0.80$, $SD = 0.79$), item 30 ("I was afraid of becoming a burden to my family or children", $M = 0.63$, $SD = 0.83$) and item 29 ("I was afraid of dying", $M = 0.35$, $SD = 0.55$).

DISCUSSION

This study aimed to examine the psychometric properties of the GAS, translated into Italian, among a larger, geographically

more varied sample of older adults. Factor structure, internal reliability, convergent and discriminant validity as well as the gender differences were examined.

Regarding the analysis of the GAS factor structure, the CFA confirmed the better fit of the three factors (Cognitive, Somatic, and Affective) originally derived from English version (Segal et al., 2010; Yochim et al., 2011, 2013). The three latent factors are those that best explained the data.

The GAS captured the broad range of anxiety disorder symptoms. The clinician or researcher can easily determine which types of symptoms are more problematic for the respondent (Segal et al., 2010).

Results also provided evidence about gender invariance. The test of the metric and scalar invariance of the model in relation to gender revealed that all the factor loadings showed to be invariant and the intercepts for observed variables loading on the same latent variable. As scalar invariance was established, means can be reliably compared. Sex is a variable which has been identified as a risk factor for anxiety. Analyses of latent mean differences revealed that females exhibited higher means than males on two GAS subscales, Somatic and Affective, where the means for Cognitive Factor were not.

As expected, women tend to report higher levels of anxiety than men, a finding that is reported consistently in literature. Gum et al. (2009) found that community-dwelling individuals who were diagnosed with an anxiety disorder were more likely to be female. Furthermore, female gender has been associated with a greater likelihood of anxiety chronicity in older adults (De Beurs et al., 2000; Gatti et al., 2017), such that anxiety tends to persist in older women compared to older men.

In addition, results suggested that the GAS total score and subscale scores have good internal consistency reliability (Cronbach's alpha, McDonald's omega, and inter-item correlations mean of items). The Cronbach's alpha values compared to values of the Segal et al. (2010) original version did not differ significantly except for the GAS total score (Feldt test = 0.5833, $p < 0.001$; see Feldt, 1969; Feldt et al., 1987) and Cognitive scale (Feldt test = 0.4167, $p < 0.001$) in which the original sample scored higher reliabilities values. Similar results were found when comparing the alpha values of the Italian version of the GAS with both the Persian and German versions. Cronbach's alpha values not differ significantly (all $p = ns$; Feldt et al., 1987).

Regarding interscale correlations, as expected, there were strong positive relationships between the GAS total score and each of the GAS subscales. Therefore, the relatively high intercorrelation of the scales, which especially occurred between the Cognitive and Affective subscales, is not surprising and can be traced back to the fact that symptoms of anxiety disorders are often comorbid with each other (DSM-IV-TR and DSM-V; Kogan et al., 2000; Segal et al., 2010; Wolitzky-Taylor et al., 2010; American Psychiatric Association, 2013).

Convergent validity of the GAS was evidenced via significant and high correlations between the GAS total score, subscale scores and another measure of anxiety (GAI).

With respect to the discriminant validity of the GAS, our findings confirmed the expected low relationships with measures

of constructs that are non-related (i.e., Extraversion, Openness, Agreeableness, Conscientiousness), or negatively related (i.e., Emotional Stability) to anxiety, whereby the relation between the GAS total score, subscale scores and depression (TDI) was lower than the correlation with anxiety measure. Anxiety in older adults is highly co-morbid with depressive symptoms (Beekman et al., 2000).

It is not surprising that Cognitive subscale and following Affective subscale were associated with measure of depression more strongly than Somatic subscale, because cognitive and affective aspect are two important components of many anxiety disorders (Cioffi et al., 2008; Van Dam et al., 2013; Balsamo et al., 2015b).

Together, the present findings support the reliability and validity of the GAS as a measure of anxiety in an Italian geriatric population. These results are important because the detection of anxiety in older adults is generally complicated by the high frequency of medical disorders present in this age group (Balsamo et al., 2015b, 2018).

Several limitations of this study should be addressed in future studies.

First, we did not investigate various aspects of reliability of the questionnaire (e.g., test-retest reliability). Second, our results are based on a general community sample of older adults, which limit the generalizability of these findings to clinical conditions. More specifically, the confirmatory models and the correlational analyses among self-report measures found in nonclinical samples might not be similar to the processes in clinical samples (see, for example, Balsamo, 2013; Balsamo et al., 2013c). In addition, our sample is non-representative of the Italian population. More heterogeneous individuals by age, education level and geographical provenience education level and geographic origin would reduce potential selection biases our data could be affected. Therefore, validity and usefulness of the GAS in clinical samples and non clinical are not fully guaranteed. Lastly, correlations for convergent and discriminant validity could be computed by using the SEM approach in order to control over the measurement error, obtaining higher precision than the computation with Pearson's r , and other concurrent measures should be taken into consideration, such as measures of trait and state anxiety (Balsamo et al., 2016).

Further research should explore the psychometric performance (e.g., its Differential Item Functioning analysis) of the Italian GAS in larger and more diverse samples of Italian older adults, including also clinical samples and groups with more diverse ethnicity, in order to improve the knowledge on this instrument, providing a more specific assessment of cognitive, affective and somatic anxiety symptoms among older adults. Moreover, a hierarchical or bifactor factorial model could be applied to empirically verify the general score of the GAS in future studies (Reis et al., 2007).

In addition, further studies could be conducted to create a short form of the measure such as Mueller et al. (2015). Short forms of screening measures are preferable in busy clinical settings and in lengthy research protocols to reduce the burden of administration time and scoring.

Because the GAS is based on DSM symptoms of anxiety, it can help clinicians arrive at an accurate diagnosis of an anxiety disorder and thus aid in clinically appropriate treatment.

REFERENCES

- Alwahhabi, F. (2003). Anxiety symptoms and generalized anxiety disorder in the elderly: a review. *Harv. Rev. Psychiatry* 11, 180–193. doi: 10.1080/10673220303944
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th Edn* (DSM-5). Arlington, VA: American Psychiatric Association.
- Andrew, D. H., and Dulin, P. L. (2007). The relationship between self-reported health and mental health problems among older adults in New Zealand: experiential avoidance as a moderator. *Aging Ment. Health* 11, 596–603. doi: 10.1080/13607860601086587
- Andrich, D. (1995). *Rasch Models for Measurement*. Thousand Oaks, CA: Sage Publications.
- Antony, M. M., and Barlow, D. H. (2011). *Handbook of Assessment and Treatment Planning for Psychological Disorders*. New York, NY: Guilford Press.
- Ayers, C. R., Sorrell, J. T., Thorp, S. R., and Wetherell, J. L. (2007). Evidence-based psychological treatments for late-life anxiety. *Psychol. Aging* 22, 8–17. doi: 10.1037/0882-7974.22.1.8
- Balsamo, M., and Saggino, A. (2007). Test per l'assessment della depressione nel contesto italiano: un'analisi critica [Tests for depression assessment in Italian context: a critical review]. *Psicoter. Cogn. Comport.* 13, 167–199.
- Balsamo, M., and Saggino, A. (2013a). *Il Teate Depression Inventory-Manuale [Teate Depression Inventory, Manual]*. Firenze: Hoegrefe.
- Balsamo, M., and Saggino, A. (2014a). Determining a diagnostic cut-off on the Teate Depression Inventory. *Neuropsychiatr. Dis. Treat.* 10, 987–995. doi: 10.2147/NDT.S55706
- Balsamo, M., Romanelli, R., and Saggino, A. (2010). The de-differentiation hypothesis in normal elderly subjects. *Psychol. Rep.* 110, 85–88. doi: 10.2466/PMS.110.1.85-88
- Balsamo, M., Romanelli, R., Innamorati, M., Ciccarese, G., Carlucci, L., and Saggino, A. (2013b). The state-trait anxiety inventory: shadows and lights on its construct validity. *J. Psychopathol. Behav. Assess.* 35, 475–486. doi: 10.1007/s10862-013-9354-5
- Balsamo, M., Lauriola, M., and Saggino, A. (2013c). Work values and college major choice. *Learn. Individ. Differ.* 24, 110–116. doi: 10.1016/j.lindif.2012.12.022
- Balsamo, M., Giampaglia, G., and Saggino, A. (2014b). Building a new Rasch-based self-report inventory of depression. *Neuropsychiatr. Dis. Treat.* 13, 153–165. doi: 10.2147/NDT.S53425
- Balsamo, M., Macchia, A., Carlucci, L., Picconi, L., Tommasi, M., Gilbert, P., et al. (2015a). Measurement of external shame: an inside view. *J. Pers. Assess.* 97, 81–89. doi: 10.1080/00223891.2014.947650
- Balsamo, M., Innamorati, M., Van Dam, N. T., Carlucci, L., and Saggino, A. (2015b). Measuring anxiety in the elderly: psychometric properties of the state trait inventory of cognitive and somatic anxiety (STICSA) in an elderly Italian sample. *Int. Psychogeriatr.* 27, 999–1008. doi: 10.1017/S1041610214002634
- Balsamo, M., Carlucci, L., Murdock, K. K., Sergi, M. R., and Saggino, A. (2015c). The mediating role of early maladaptive schemas in the relation between co-rumination and depression in youths. *PLoS ONE* 10:e0140177. doi: 10.1371/journal.pone.0140177
- Balsamo, M. (2013). Personality and depression: evidence of a possible mediating role for anger trait in the relationship between cooperativeness and depression. *Compr. Psychiatry* 54, 46–52. doi: 10.1016/j.comppsy.2012.05.007
- Balsamo, M., Carlucci, L., Sergi, M. R., Romanelli, R., D'Ambrosio, I., Fairfield, B., et al. (2016). A new measure for trait and state anxiety: the State Trait Inventory of Cognitive and Somatic Anxiety (STICSA). Standardization in an Italian population. *Psicoter. Cogn. Comport.* 22, 229–232.
- Balsamo, M., Cataldi, F., Carlucci, L., and Fairfield, B. (2018). Assessment of anxiety in the older adults: a review of self-report measures. *Clin. Interv. Aging* 13, 573–593. doi: 10.2147/CIA.S114100
- Barbaranelli, C. (2006). *Analisi dei Dati con SPSS II. Le Analisi Multivariate*. Milano: LED Edizioni Universitarie.
- Barbaranelli, C., and D'Olimpio, F. (2007). *Analisi Dei Dati Con SPSS, Vol.1*. Milano: Led Edizioni Universitarie.
- Barbaranelli, C., and Ingoglia, S. (2013). *I Modelli di Equazioni Strutturali. Temi e Prospettive*. Milano: LED.
- Barbaranelli, C., and Natali, E. (2005). *I Test Psicologici. Teorie e Modelli Psicometrici*. Roma: Carocci.
- Baxter, A. J., Scott, K. M., Vos, T., and Whiteford, H. A. (2013). Global prevalence of anxiety disorders: a systematic review and meta-regression. *Psychol. Med.* 43, 897–910. doi: 10.1017/S003329171200147X
- Beekman, A. T., de Beurs, E., van Balkom, A. J., Deeg, D. J., van Dyck, R., and van Tilburg, W. (2000). Anxiety and depression in later life: co-occurrence and communality of risk factors. *Am. J. Psychiatry* 157, 89–95. doi: 10.1176/ajp.157.1.89
- Bentler, P. M. (2006). *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software
- Bolghan-Abadi, M., Segal, D. L., Coolidge, F. L., and Gottschling, J. (2013). Persian version of the Geriatric Anxiety Scale: translation and preliminary psychometric properties among Iranian older adults. *Aging Ment. Health* 17, 896–900. doi: 10.1080/13607863.2013.788999
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Press.
- Byrne, B. M. (2006). *Structural Equation Modeling with EQS: Basic Concepts, Applications and Programming, 2nd Edn*. Mahwah, NJ: Erlbaum.
- Byrne, G. J., Pachana, N. A., Goncalves, D. C., Arnold, E., King, R., and Khoo, S. K. (2010). Psychometric properties and health correlates of the geriatric anxiety inventory in Australian community-residing older women. *Aging Ment. Health* 14, 247–254. doi: 10.1080/13607861003587628
- Caprara, G. V., Barbaranelli, C., Borgogni, L., and Perugini, M. (1993). The “big five questionnaire”: A new questionnaire to assess the five factor model. *Pers. Individ. Dif.* 15, 281–288. doi: 10.1016/0191-8869(93)90218-R
- Caprara, G. V., Barbaranelli, C., Bermudez, J., Maslach, C., and Ruch, W. (2000). Multivariate methods for the comparison of factor structures in cross-cultural research: an illustration with the Big Five Questionnaire. *J. Cross Cult. Psychol.* 31, 437–464. doi: 10.1177/0022022100031004002
- Caprara, G. V., Barbaranelli, C., Borgogni, L., and Vecchione, M. (2007). *Big Five Questionnaire-2: Manual*. Firenze: Organizzazioni Speciali.
- Cheung, G. W., and Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equation modeling. *J. Cross Cult. Psychol.* 31, 187–212. doi: 10.1177/0022022100031002003
- Cioffi, R., Balsamo, M., and Saggino, A. (2008). Psychometric characteristic of the Italian clinical depression questionnaire using a 5-point rating scale with undergraduates. *Psychol. Rep.* 102, 305–311. doi: 10.2466/pr0.102.1.305-311
- Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155
- Cohen, J., and Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 2nd Edn*. Hillsdale, NJ: Erlbaum.
- Contardi, A., Imperatori, C., Amati, I., Balsamo, M., and Innamorati, M. (2018). Assessment of affect lability: psychometric properties of the ALS-18. *Front. Psychol.* 9:427. doi: 10.3389/fpsyg.2018.00427
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 3, 297–334. doi: 10.1007/BF02310555
- Cully, J. A., Graham, D. P., Stanley, M. A., Ferguson, C. J., Sharafkhan, A., Soucek, J., et al. (2006). Quality of life in patients with chronic obstructive pulmonary disease and comorbid anxiety or depression. *Psychosomatics* 47, 312–319. doi: 10.1176/appi.psy.47.4.312
- De Beurs, E., Beekman, A. T. F., Deeg, D. J. H., van Dyck, R., and van Tilburg, W. (2000). Predictors of change over three years in anxiety symptoms of older

AUTHOR CONTRIBUTIONS

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

- persons: results from the Longitudinal Aging Study Amsterdam. *Psychol. Med.* 30, 515–527 doi: 10.1017/S0033291799001956
- Dennis, R. E., Boddington, S. J., and Funnell, N. J. (2007). Self-report measures of anxiety: are they suitable for older adults? *Aging Mental Health* 11, 668–677. doi: 10.1080/13607860701529916
- Diefenbach, G. J., Tolin, D. F., Meunier, S. A., and Gilliam, C. M. (2009). Assessment of anxiety in older home care recipients. *Gerontologist* 49, 141–153. doi: 10.1093/geront/gnp019First
- Dunn, T. J., Baguley, T., and Brunsden, V. (2014). From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br. J. Psychol.* 105, 399–412. doi: 10.1111/bjop.12046
- Ercolani, A. P., Areni, A., and Leone, L. (2001). *Statistica per la Psicologia*. Bologna: Il Mulino.
- Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika* 34, 363–373. doi: 10.1007/BF02289364
- Feldt, L. S., Woodruff, D. J., and Salih, F. A. (1987). Statistical inference for coefficient alpha. *Appl. Psychol. Meas.* 11, 93–103. doi: 10.1177/014662168701100107
- Gatti, A., Gottschling, J., Brugnera, A., Adorni, R., Zarbo, C., Compare, A., et al. (2017). An investigation of the psychometric properties of the Geriatric Anxiety Scale (GAS) in an Italian sample of community-dwelling older adults. *Aging Ment. Health*. doi: 10.1080/13607863.2017.1347141. [Epub ahead of print].
- Gottschling, J., Segal, D. L., Häusele, C., Spinath, F. M., and Stoll, G. (2015). Assessment of anxiety in older adults: translation and psychometric evaluation of the German version of the Geriatric Anxiety Scale (GAS). *J. Psychopathol. Behav. Assess.* 38, 136–148. doi: 10.1007/s10862-015-9504-z
- Graham, J. M., Guthrie, A. C., and Thompson, B. (2003). Consequences of not interpreting structure coefficients in published CFA research: a reminder. *Struct. Equation Model. Multidiscipl. J.* 10, 142–153. doi: 10.1207/S15328007SEM1001_7
- Gum, A. M., King-Kallimanis, B., and Kohn, R. (2009). Prevalence of mood, anxiety, and substance-abuse disorders for older Americans in the national comorbidity survey-replication. *Am. J. Geriatr. Psychiatry* 17, 769–781. doi: 10.1097/JGP.0b013e3181ad4f5a
- Hu, L., and Bentler, P. M. (1995). "Evaluating model fit," in *Structural Equation Modeling: Issues, Concepts, and Applications*, ed R. H. Hoyle (Newbury Park, CA: Sage), 76–99.
- Hu, L., and Bentler, P. M. (1998). Fit indices in covariance structure modeling: sensitivity to under-parameterized model misspecification. *Psychol. Methods* 3, 424–453. doi: 10.1037/1082-989X.3.4.424
- IBM SPSS (2010). *Statistics for Windows, Version 19.0*. Armonk, NY: IBM Corp.
- Innamorati, M., Tamburello, S., Contardi, A., Imperatori, C., Tamburello, A., Saggino, A., et al. (2013). Psychometric properties of the attitudes toward self-revised in Italian young adults. *Depress. Res. Treat.* 2013:209216. doi: 10.1155/2013/209216
- Innamorati, M., Lester, D., Balsamo, M., Erbuto, D., Ricci, F., Amore, M., et al. (2014). Factor Validity of the Beck Hopelessness Scale in Italian Medical Patients. *J. Psychopathol. Behav. Assess.* 36, 300–307. doi: 10.1007/s10862-013-9380-3
- Istat (2011). *Il Futuro Demografico Del Paese: Previsioni Regionali Della Popolazione Residente al 2065*. Roma: Istat Istituto Nazionale di Statistica.
- Istat (2013). *Tutela Della Salute E Accesso Alle Cure*. Istat - Istituto Nazionale di Statistica.
- Kline, P. (1986). *A Handbook of Test Construction*. New York, NY: Methuen.
- Kogan, J. N., Edelstein, B. A., and McKee, D. R. (2000). Assessment of anxiety in older adults: current status. *J. Anxiety Disord.* 14, 109–132. doi: 10.1016/S0887-6185(99)00044-4
- Lin, X. L., Lu, D. L., Gottschling, J., Segal, D. L., and Tang, S. Y. (2016). Validation of a Chinese version of the geriatric anxiety scale among community-dwelling older adults in mainland China. *J. Cross Cult. Gerontol.* 32, 57–70. doi: 10.1007/s10823-016-9302-4
- Magni, G., and DeLeo, D. (1984). Anxiety and depression in geriatric and adult medical patients. *Psychol. Rep.* 55, 607–612. doi: 10.2466/pr0.1984.55.2.607
- Mardia, K. V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya Series B* 36, 115–128.
- McLean, C. P., Asnaani, A., Litz, B. T., and Hofmann, S. G. (2011). Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. *J. Psychiatr. Res.* 45, 1027–1035. doi: 10.1016/j.jpsychires.2011.03.006
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika* 58, 525–543.
- Mueller, A. E., Segal, D. L., Gavett, B., Marty, M. A., Yochim, B., June, A., et al. (2015). Geriatric Anxiety Scale: item response theory analysis, differential item functioning, and creation of a ten-item short form (GAS-10). *Int. Psychogeriatr.* 27, 1099–1111. doi: 10.1017/S1041610214000210
- Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*, 3rd Edn. New York, NY: McGraw-Hill.
- Owens, K. M. B., Hadjistavropoulos, T., and Asmundson, G. J. G. (2000). Addressing the need for appropriate norms when measuring anxiety in seniors. *Aging Ment. Health* 4, 309–314. doi: 10.1080/13607860020010448
- Pachana, N. A., Byrne, G. J., Siddle, H., Koloski, N., Harley, E., and Arnold, E. (2007). Development and validation of the Geriatric Anxiety Inventory. *Int. Psychogeriatr.* 19, 103–114. doi: 10.1017/S1041610206003504
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research.
- Reis, S. P., Morizot, J., and Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res.* 16 (Suppl. 1), 19–31. doi: 10.1007/s11136-007-9183-7
- Reynolds, C. R., Richmond, B. O., and Lowe, P. A. (2003). *The Adult Manifest Anxiety Scale-Elderly Version (AMAS-E)*. Los Angeles, CA: Western Psychological Services.
- Rhemtulla, M., Brosseau-Liard, P., and Savalei, V. (2012). When can categorical variables be treated as continuous? a comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* 17, 354–373. doi: 10.1037/a0029315
- Rozzini, L., Chilovi, B. V., Peli, M., Conti, M., Rozzini, R., Trabucchi, M., et al. (2009). Anxiety symptoms in mild cognitive impairment. *Int. J. Geriatr. Psychiatry* 24, 300–305. doi: 10.1002/gps.2106
- Saggino, A., Balsamo, M., Da Fermo, G., Dettore, D., Marsigli, N., Petrucelli, I., et al. (2014). Standardizzazione e validazione della versione italiana dello Young Schema Questionnaire Versione L-3. *Psicoter. Cogn. Comput.* 20, 387–390.
- Saggino, A., Carlucci, L., Sergi, M. R., D'Ambrosio, I., Fairfield, B., and Balsamo, M. (2017). A validation study of the psychometric properties of the other as shamer scale-2. *SAGE Open* 7, 1–10. doi: 10.1177/2158244017704241
- Saris, W. E., Satorra, A., and Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. *Soc. Methodol.* 17, 105–129.
- Saris, W. E., Satorra, A., and Van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications?. *Struct. Equation Model.* 16, 561–582. doi: 10.1080/10705510903203433
- Satorra, A., and Bentler, P. M. (1994). "Corrections to test statistics and standard errors in covariance structure analysis," in *Latent Variables Analysis: Applications For Developmental Research*, eds A. von Eye and C. C. Clogg (Thousand Oaks, CA: Sage), 399–419.
- Satorra, A., and Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika* 75, 243–248. doi: 10.1007/s11336-009-9135-y
- Segal, D. L., June, A., Payne, M., Coolidge, F. L., and Yochim, B. (2010). Development and initial validation of a self-report assessment tool for anxiety among older adults: the geriatric anxiety scale. *J. Anxiety Disord.* 24, 709–714. doi: 10.1016/j.janxdis.2010.05.002
- Seignourel, P. J., Kunik, M. E., Snow, L., Wilson, N., and Stanley, M. (2008). Anxiety in dementia: a critical review. *Clin. Psychol. Rev.* 28, 1071–1082. doi: 10.1016/j.cpr.2008.02.008
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Pers. Individ. Diff.* 42 (5), 893–98. doi: 10.1016/j.paid.2006.09.017
- Tabachnick, B. G., and Fidell, L. S. (2007). *Using Multivariate Statistics*. Boston, MA: Pearson/Allyn and Bacon.
- Therrien, Z., and Hunsley, J. (2012). Assessment of anxiety in older adults: a systematic review of commonly used measures. *Aging Ment. Health* 16, 1–16. doi: 10.1080/13607863.2011.602960

- Thompson, B. (1997). Editorial policies regarding statistical significance tests: Further comments. *Educ. Res.* 26, 29–32.
- Van Dam, N. T., Gros, D. F., Earleywine, M., and Antony, M. M. (2013). Establishing a trait anxiety threshold that signals likelihood of anxiety disorders. *Anxiety Stress Coping* 26, 70–86. doi: 10.1080/10615806.2011.631525
- van de Schoot, R., Lugtig, P., and Hox, J. (2012). A checklist for testing measurement invariance. *Eur. J. Dev. Psychol.* 9, 486–492. doi: 10.1080/17405629.2012.686740
- West, S. G., Finch, J. F., and Curran, P. J. (1995). “Structural equation models with non normal variables: problems and remedies,” in *Structural Equation Modeling: Concepts, Issues, and Applications* ed R.H. Hoyle (Thousand Oaks, CA: Sage Publications, Inc.), 56–75.
- Wolitzky-Taylor, K. B., Castriotta, N., Lenze, E. J., Stanley, M. A., and Craske, M. G. (2010). Anxiety disorders in older adults: a comprehensive review. *Depress. Anxiety* 27, 190–211. doi: 10.1002/da.20653
- Yochim, B. P., Mueller, A. E., June, A., and Segal, D. L. (2011). Psychometric properties of the geriatric anxiety scale: comparison to the beck anxiety inventory and geriatric anxiety inventory. *Clin. Gerontol.* 34, 21–33. doi: 10.1080/07317115.2011.524600
- Yochim, B. P., Mueller, A. E., and Segal, D. L. (2013). Late life anxiety is associated with decreased memory and executive functioning in community dwelling older adults. *J. Anxiety Disord.* 27, 567–575. doi: 10.1016/j.janxdis.2012.10.010
- Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Unpublished doctoral dissertation, University of California, Los Angeles, CA.
- Zinbarg, R. E., Revelle, W., Yovel, I., and Li, W. (2005). Cronbach’s α , Revelle’s b , and McDonald’s ω^2 : their relations with each other and two alternative conceptualizations of reliability. *Psychometrika* 70, 123–133. doi: 10.1007/s11336-003-0974-7
- Zung, W. W. (1971). A rating instrument for anxiety disorders. *Psychosomatics* 12, 371–379. doi: 10.1016/S0033-3182(71)71479-0

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

Copyright © 2018 Picconi, Balsamo, Palumbo and Fairfield. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Dimensions of Anxiety, Age, and Gender: Assessing Dimensionality and Measurement Invariance of the State-Trait for Cognitive and Somatic Anxiety (STICSA) in an Italian Sample

Leonardo Carlucci^{1*}, Marley W. Watkins², Maria Rita Sergi¹, Fedele Cataldi¹, Aristide Saggino¹ and Michela Balsamo¹

¹ School of Medicine and Health Sciences, G. d'Annunzio University of Chieti–Pescara, Chieti, Italy, ² Department of Educational Psychology, Baylor University, Waco, TX, United States

OPEN ACCESS

Edited by:

Elisa Pedrolí,
Istituto Auxologico Italiano (IRCCS),
Italy

Reviewed by:

Melissa Ree,
The Marian Centre, Australia
María C. Fuentes,
University of Valencia, Spain

*Correspondence:

Leonardo Carlucci
l.carlucci@unich.it

Specialty section:

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

Received: 27 April 2018

Accepted: 08 November 2018

Published: 27 November 2018

Citation:

Carlucci L, Watkins MW,
Sergi MR, Cataldi F, Saggino A and
Balsamo M (2018) Dimensions
of Anxiety, Age, and Gender:
Assessing Dimensionality
and Measurement Invariance of the
State-Trait for Cognitive and Somatic
Anxiety (STICSA) in an Italian Sample.
Front. Psychol. 9:2345.
doi: 10.3389/fpsyg.2018.02345

The State–Trait Inventory for Cognitive and Somatic Anxiety (STICSA) is a widely used measure of state and trait anxiety that permits a specific assessment of cognitive and somatic anxiety. Previous research provided inconsistent findings about its factor structure in non-clinical samples (e.g., hierarchical or bi-factor structure). To date, no psychometric validation of the Italian version of the STICSA has been conducted. Our study aimed to determine the psychometric functioning of the Italian version of the STICSA, including its dimensionality, gender and age measurement equivalence, and convergent/divergent validity in a large sample of community-dwelling participants ($N = 2,938$; 55.9% female). Through confirmatory factor analysis, the multidimensional structure of both State and Trait STICSA scales, with each including Cognitive and Somatic dimensions, was supported. Factor structure invariance was tested and established at configural, metric, and scalar levels for males and females. Additionally, full factorial measurement invariance was supported for the State scale across young, middle age, and old adult groups whereas the Trait scale was partially invariant across age groups. The STICSA also showed good convergent validity with concurrent anxiety measures (State-Trait Anxiety Inventory and Beck Anxiety Inventory), and satisfactory internal discriminant validity with two depression measures (Teate Depression Inventory and Beck Depression Inventory-II). Results provided support for the multidimensionality of the STICSA, as well as the generalizability of the State and Trait scales as independent measures of Cognitive and Somatic symptomatology across gender in the general population. Implications for research and personality and clinical assessment are discussed.

Keywords: anxiety, depression, trait, state, invariance, multigroup confirmatory factor analysis

INTRODUCTION

Anxiety is an emotional state defined by cognitive as well as somatic symptomatology such as feelings of tension, worried thoughts, increased blood pressure, sweating, derealization, and the anticipation of a future danger or threat (American Psychiatric Association, 2013). Anxiety symptoms are the most common of mental disorders and affect nearly 33% of adults at some point

in their lives (Bandelow and Michaelis, 2015). In Italy, the prevalence in life of anxiety disorders is close to 11% (de Girolamo et al., 2006; Kessler et al., 2007).

The most widely documented results from psychiatric epidemiology are that anxiety symptoms develop from childhood and persist into adulthood if not detected and treated (Regier et al., 1990; Wittchen and Jacobi, 2005; Kessler et al., 2012; Bandelow and Michaelis, 2015), and that females are significantly more likely than males to develop anxiety disorders throughout the lifespan (ratio 1:2) (McLean and Anderson, 2009; McLean et al., 2011).

Historically, there has been considerable debate regarding the dimensionality of anxiety: it was considered unidimensional by Freud (1920) but characterized with both trait and state dimensions by contemporary researchers (Spielberger, 1985; Endler, 1997). The modern differentiation between trait and state anxiety has a long and controversial history (Allport and Odbert, 1936; Carr and Kingsbury, 1938; Zuckerman, 1960, 1983; Fridhandler, 1986; Endler and Kocovski, 2001; Heeren et al., 2018). A notable amount of research differentiates, for anxiety as well as other psychological states, between transitory emotion that varies in duration and is characterized by observable symptoms (i.e., state anxiety), and an individual's unobservable disposition to experience elevated anxiety in response to threat (i.e., trait anxiety) (Spielberger, 1983, 1985; Endler and Kocovski, 2001; Pacheco-Unguetti et al., 2010; Heeren et al., 2018).

Trait anxiety has been extensively conceptualized as a fundamental dimension along which people differ (Allport and Odbert, 1936; Cattell, 1946; Eysenck, 1953). According to several personality trait taxonomies, trait anxiety has been variously theorized as negative emotionality (Tellegen, 1985), neuroticism (Costa and MacCrae, 1992; Ashton et al., 2004), low emotional stability (Goldberg, 1992), a risk factor for the development of anxious symptomatology (Weems et al., 2007), and comparable to anxiety sensitivity (Lilienfeld et al., 1993). On the other hand, state anxiety has been viewed as an emotional state that varies in duration depending of the presence of the provocative stimulus. According to this distinction, individuals high on trait anxiety are more likely to experience episodes of state anxiety (in terms of intensity, frequency, duration) than those low on trait anxiety (Heeren et al., 2018).

Nevertheless, the state-trait distinction has been labeled as arbitrary and based on weak assumptions, such as the minor difference in instructions included in anxiety measures divided into state and trait scales (e.g., "last week" versus "generally") by several authors (Allen and Potkay, 1981; Zuckerman, 1983; Luthans et al., 2007). Therefore, state and trait anxiety could be considered as only interchangeable labels, representing two interconnected components. Ultimately, a trait index can be inferred from a state measurement (Allen and Potkay, 1981).

More recently, the distinction between cognitive and somatic anxiety symptoms has been explored (Steptoe and Kearsley, 1990; Ree et al., 2008; Waechter and Stolz, 2015). Clinical investigators have long considered the symptoms of anxiety to be phenomenologically heterogeneous and involving a wide array of physical, emotional, and cognitive components (Buss, 1962; Schalling et al., 1975; Steptoe and Kearsley, 1990).

For example, anxiety was seen to involve somatic symptoms such as hyperventilation, sweating, and trembling as well as cognitive symptoms such as worry, intrusive thoughts, and lack of concentration (Ree et al., 2008). This cognitive/somatic distinction might better encompass all aspects included in the construct of anxiety and might allow treatment to be tailored for the predominant modality of anxiety experienced (e.g., cognitively orientated meditation versus self-instructional training with physiologically orientated relaxation) (Steptoe and Kearsley, 1990). Another controversial issue regarding anxiety is the overlap between anxiety and depression (Flint, 2005; Wetherell and Gatz, 2005; Bryant et al., 2008).

A considerable amount of research has emphasized that anxiety and depression share a common component of general distress in addition to components specific to each disorder (Clark and Watson, 1991; Watson et al., 1995a,b; Smoller and Tsuang, 1998; Costello et al., 2003; Hasler et al., 2004; Godfrey et al., 2005; Shafer, 2006; Ree et al., 2008). This finding was not surprising given the high comorbidity between anxiety and depression mood disorders (Watson et al., 1995a,b; Costello et al., 2003; Godfrey et al., 2005). In line with the tripartite model (Clark and Watson, 1991), aversive emotional states (fear, anger, guilt) are associated with both anxiety and depression; the lack of positive affect (feeling tired) is associated with depression whereas physiological hyperarousal (trembling, dizziness, shaking) with anxiety (Beck et al., 1988; Clark and Watson, 1991; Watson et al., 1995a,b).

Assessment of Anxiety

Given the hypothetical multidimensional nature of anxiety and its manifold symptomatic manifestations, the assessment of anxiety represents a challenge for clinicians and researchers.

STAI

The most widely used self-rating measure for measuring anxiety in its trait and state components is Spielberger (1983) State-Trait Anxiety Inventory (STAI), but recent studies have raised doubts about the anxiety construct as measured by the STAI (Bieling et al., 1998; Caci et al., 2003; Bados et al., 2010; Balsamo et al., 2013c; Hill et al., 2013). According to these authors, the STAI can best be conceptualized as assessing negative affect, rather than a pure measure of anxiety. Indeed, the STAI has exhibited poor discriminatory power between anxiety and depression (Kabacoff et al., 1997; Bieling et al., 1998; Kennedy et al., 2001; Balsamo et al., 2013c; Bergua et al., 2016) and its scores have been more strongly correlated with a measure of depression than a measure of anxiety (Grös et al., 2007). In addition, its use appears to be particularly problematic among older adults, due to its length and format (McDonald and Spielberger, 1983; Dennis et al., 2007; Therrien and Hunsley, 2012; Balsamo et al., 2018).

STICSA

To overcome some of the issues associated with the use of the STAI, Ree et al. (2008) developed a new measure based on Spielberger's (1966) conceptualization of state and trait anxiety, named the State-Trait Inventory for Cognitive

and Somatic Anxiety (STICSA). The STICSA also contains subscales measuring somatic and cognitive symptom clusters. For example, the cognitive cluster aims to capture aspects of anxiety related to thoughts (e.g., difficulty concentrating, worry, intrusive thoughts), whereas the somatic cluster aims to capture features that directly relate to physical experiences (e.g., sweating, muscle tension, palpitations). Additionally, the use of balanced scales composed by separate groupings of cognitive and somatic anxiety items potentially facilitates the differentiation of anxiety from anxiety-like symptoms (i.e., symptoms caused by a medical condition). Therefore, the STICSA differs from most extant measures of anxiety which contain an overrepresentation of cognitive (like the State-Trait Anxiety Inventory) or somatic symptoms (like the Beck Anxiety Inventory), which makes it difficult to distinguish between anxiety, mood, and medical symptoms (Ree et al., 2008; Elwood et al., 2012; Deacy et al., 2016). The inclusion of both trait-state and somatic-cognitive clusters might allow the STICSA to better capture the heterogeneity of symptoms associated with anxiety disorders (Watson et al., 2005).

Previous research has demonstrated that the STICSA exhibited strong psychometric properties in both clinical and non-clinical samples of adults (Grös et al., 2007, 2010; Ree et al., 2008; Van Dam et al., 2013; Balsamo et al., 2015b; Roberts et al., 2016) and children (Deacy et al., 2016), as well as across African and European American samples (Lancaster et al., 2015).

Specifically, the STICSA has demonstrated sufficient to excellent values of internal consistency for both the State ($\alpha = 0.74\text{--}0.95$) and Trait ($\alpha = 0.75\text{--}0.95$) scales and test–retest reliability for the Trait scale ($r = 0.60\text{--}0.66$) among young and older adults, students, and clinical groups (Grös et al., 2007, 2010; Ree et al., 2008; Van Dam et al., 2013; Balsamo et al., 2015b; Deacy et al., 2016; Roberts et al., 2016). Concerning convergent and divergent validity, several studies revealed that STICSA, both at scale and dimension level, correlated at medium to high levels with other anxiety measures (i.e., Mood and Anxiety Symptom Questionnaire, Cognitive–Somatic Anxiety Questionnaire, avoidance measure, worry, and social anxiety) and at medium levels with depression self-report measures (i.e., Depression Anxiety Stress Scales, Beck Depression Inventory-II) (Grös et al., 2007; Ree et al., 2008). These studies highlighted the discriminant power of the STICSA, which allowed differentiation of anxiety from depression better than other anxiety measures, avoiding misdiagnosis, a fairly frequent problem in clinical practice (Therrien and Hunsley, 2012).

The STICSA was designed to tap two correlated subscales (State and Trait), each composed of two interrelated dimensions (Cognitive and Somatic). Accordingly, it produces four scores: State Cognitive, State Somatic, Trait Cognitive, and Trait Somatic. Additionally, all four scores might be combined to produce a total anxiety score, the two state scores could be combined to produce a state anxiety score, the two trait scores could be combined to produce a trait anxiety score, the two cognitive scores could be combined to produce a cognitive anxiety score, and the two somatic scores could be combined to produce a somatic anxiety score.

Given these possible scoring schemes, the structure of the STICSA can be conceptualized in several different ways. Not all of these conceptualizations have been considered in extant research. Ree et al. (2008) performed confirmatory factor analyses (CFA) on the trait and state scales separately among Australian students and adults, finding two correlated factors within each scale (i.e., cognitive and somatic). Grös et al. (2010) only examined the trait scale among American students and also found correlated cognitive and somatic factors. Grös et al. (2007) included both state and trait STICA items in their analysis of responses from Canadian psychiatric patients and U.S. college students. Given that trait and state items are identical (response instructions differentiate trait items as experienced “in general” and state items as experienced “right now”), Grös et al. (2007) allowed item error terms to correlate and found that four correlated factors [Somatic State (SS), Somatic Trait (ST), Cognitive State (CS), and Cognitive Trait (CT)] best fit their data. However, neither higher-order nor bifactor models were tested. Balsamo et al. (2015b) found a similar oblique four-factor structure among older Italian adults but did not allow correlated item errors and did not include higher-order or bifactor models. Roberts et al. (2016) also analyzed both trait and state STICSA items (among Canadian college students) and found support for a correlated four-factor model as well as a higher-order model with a global anxiety factor and four first-order factors. However, their models did not include correlated error terms across the state-trait items and they did not consider bifactor models. In contrast, Lancaster et al. (2015) did not find support for the oblique four-factor model among African American and European American university students. Unfortunately, Lancaster et al. (2015) failed to test other potential structural models.

Given the lack of clarity about the factor structure of the STICSA, the current study aimed to address evidence for the dimensionality of the STICSA on a large cross-age sample. Through a confirmatory methodology, we tested all the STICSA factor structure models found in the literature (hierarchical, bifactor, four-factor, and two factors models), to evaluate which model best represent the anxiety construct as conceptualized by this instrument.

Another unaddressed issue associated with the psychometric functioning of the STICSA is its measurement invariance across gender and age. Although studies of gender differences in anxiety have provided support for the higher prevalence rates of anxiety symptomatology and disorders among females across the life span, in both community and clinical samples (Lewinsohn et al., 1998; Egger et al., 2003; Bruce et al., 2005; McLean and Anderson, 2009; McLean et al., 2011), no studies have investigated the impact of gender differences on the measurement of anxiety with the STICSA. Additionally, studies on age differences in anxiety have provided support for quantitative and qualitative differences of presentation of anxiety symptomatology in younger and older adults (Blazer et al., 1991; Christensen et al., 1999; Balsamo et al., 2018), but no studies had investigated the impact of age differences on the measurement of anxiety by the STICSA. Accordingly, the second aim of this study was to provide evidence of measurement invariance of the STICSA across age and gender. Lastly, convergent and divergent validity of the STICSA was

addressed to provide further evidence for the ability of STICSA scores to differentiate anxious from depressive symptoms.

MATERIALS AND METHODS

Participants

Participants were 2,983 Italian adults, including 1,667 females (55.9%) and 1,316 males (44.1%), of whom 1,780 (59.7%) were undergraduate students. The sample's mean age was 36.26 years ($SD = 20.25$ years). The mean age for men was 37.94 years ($SD = 20.55$ years), and 34.94 years ($SD = 19.93$ years) for women. The mean level of education was 11.87 ($SD = 3.67$) years. In order to address measurement invariance across age of the STICSA State and Trait scales, the sample was split into three age groups: 18–25 ($N_{total} = 1,556$; $N_{male} = 624$, $N_{female} = 932$), 26–50 ($N_{total} = 675$; $N_{male} = 319$, $N_{female} = 356$), and 51–99 ($N_{total} = 743$; $N_{male} = 366$, $N_{female} = 377$) years. A statistically significant association between Gender and Age groups was found [$\chi(2) = 20.84$, $p < 0.001$], suggesting how differences between groups potentially could be influenced by the proportion of males and females across the age groups, rather than chance.

Procedure

The sample was recruited through advertisements (flyers, newspapers, and online ads) posted for established community groups (e.g., youth centers, church groups, university student associations) in Italian cities located in northern, central, and southern sections of the country. Part of the sample used here, took part in a study of anxiety, co-rumination, shame, young schema theory, personality, and eating disorders, described elsewhere (Saggino et al., 2017a; Picconi et al., 2018).

A battery of tests, randomly sequenced, was administered by a team of psychologists and researchers. Socio-demographic variables including age, gender, and education were also collected in the present study to provide a comprehensive framework of the participants' characteristics. Given the high prevalence of individual differences in anxiety disorders, we considered gender and age variables in the following analyses (e.g., McLean et al., 2011). Participants who did not complete any of the STICSA items were excluded a-priori from all analyses. Inclusion criteria were: ages from 18 to 99 years and the ability to complete self-administered questionnaires. Exclusion criteria included marked cognitive impairment, a drug abuse disorder, diagnoses of psychotic disorders, and major disorders of the central nervous system (e.g., Alzheimer's disease, Parkinson's disease, epilepsy). For invariance analyses, pairwise deletion was used to deal with the missing data in the age or gender variables. For all other analyses, only complete questionnaire data were used.

Study participants contributed voluntarily and anonymously, and no honorarium was given for completing the assessments. Written informed consent was obtained from all participants before starting the administration, according to the Declaration of Helsinki. The ethics committee of the Department of Psychological Sciences, Health and Territory, University of Chieti, Italy, approved the study.

Measures

State–Trait Inventory for Cognitive and Somatic Anxiety (STICSA)

The STICSA (Ree et al., 2008; for the Italian version see Balsamo et al., 2015b, 2016b) is a 21-item measure designed to assess cognitive (e.g., “I feel agonized over my problems,” “I think that others won't approve of me”) and somatic (e.g., “My heart beats fast,” “My muscles are tense”) symptoms, both on Trait and State variations. In the Trait Anxiety subscale, the individual rates how often a statement is true in general (on a four-point Likert-type scale from 1 = *almost never at all* to 4 = *almost always*), whereas in the State Anxiety subscale, the examinee rates how she or he feels at the moment of assessment (on a four-point Likert-type scale from 1 = *not at all* to 4 = *very much*). In total, the overall scale is made up of four subscales: State–Somatic (SS), Trait–Somatic (TS), State–Cognitive (SC), and Trait–Cognitive (TC).

State-Trait Anxiety Inventory-Form Y (STAI-Y)

The STAI-Y (Spielberger and Gorsuch, 1983) is a self-report anxiety behavioral instrument composed of two separate 20-item subscales that measure trait (baseline) and state (situational) anxiety, resulting from a revision of the original Form X (Spielberger et al., 1970). The STAI trait subscale measures relatively stable individual differences in anxiety proneness; i.e., differences in the tendency to experience anxiety; and the STAI state subscale measures the transitory anxiety state; i.e., subjective feelings of apprehension, tension, and worry that vary in intensity and fluctuate based on the situation. Respondents are asked to rate each item on a 4-point Likert-type scale, ranging from 1 = *almost never* to 4 = *almost always*. The total score ranges from 20 to 80, with higher scores indicating greater anxiety. Internal consistencies of scores on the STAI-Y ranged from good to excellent in non-clinical and clinical samples (Stanley et al., 1996; Roberts et al., 2016; Balsamo et al., 2018). Adequate test-retest reliabilities (Stanley et al., 1996; Dennis et al., 2007), and construct validity have emerged in several studies in older adult outpatients with a variety of psychiatric disorders (Kabacoff et al., 1997; Dennis et al., 2007). In this study, coefficient alphas were 0.94 (95% CI 0.932–0.948) and 0.91 (95% CI 0.896–0.921), respectively for the State and Trait subscales.

Beck Anxiety Inventory (BAI)

The BAI (Beck et al., 1988) is a self-report inventory of 21 items with a focus on somatic symptoms of anxiety (i.e., nervousness, inability to relax) that was developed as a measure adept at discriminating between anxiety and depression. Respondents are asked to assess the degree of distress caused by these symptoms over the previous 7 days on a 4-point Likert-type scale, ranging from 0 = *not at all* to 3 = *severely*. The total score ranges from 0 to 63, with higher scores indicating greater anxiety. The BAI showed good internal and test-retest reliability as well as acceptable discriminative validity in samples of anxiety patients and non-clinical older adults (Beck et al., 1988; de Beurs et al., 1997; Diefenbach et al., 2009; Balsamo et al., 2018). Coefficient alpha for this study was 0.95 (95% CI 0.952–0.957).

Teate Depression Inventory (TDI)

The TDI is a 21-item self-report instrument designed to assess depressive symptoms (Balsamo and Saggino, 2013, 2014; Balsamo et al., 2014), as specified for major depressive disorder by the latest editions of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR, DSM-5; American Psychiatric Association). It was developed via Rasch logistic analysis of responses in order to overcome inherent psychometric weaknesses of existing measures of depression (Balsamo and Saggino, 2007). Each item is rated on a five-point Likert-type scale, ranging from 0 = *always* to 4 = *never*. The TDI has exhibited strong psychometric properties in both clinical and non-clinical samples (Balsamo et al., 2013a,c, 2014, 2015a,c, 2016a; Innamorati et al., 2013; Saggino et al., 2017b, 2018; Carlucci et al., 2018; Contardi et al., 2018). In the present sample, Cronbach's alpha was 0.91 (95% CI 0.907–0.917).

Beck Depression Inventory–II (BDI–II)

The BDI–II is a 21-item self-report inventory designed to assess the presence and severity of depressive symptoms, according to DSM-IV criteria (Beck et al., 1996). Each item is rated on a 4-point Likert-type scale ranging from 0 to 3, based on the severity of depressive symptoms over the last 2 weeks. Each item is a list of four statements arranged in increasing severity about a particular symptom of depression. The total score ranges from 0 to 63, with higher scores indicating more severe depressive symptoms. Several studies revealed high overall internal and test-retest reliability and validity for the BDI–II in undergraduates, psychiatric, and normal older adults (Gallagher et al., 1983; Beck et al., 1996; Dozois et al., 1998; Sprinkle et al., 2002; Titov et al., 2011). Coefficient alpha for this study was 0.83 (95% CI 0.817–0.873).

Data Analysis

We conducted CFAs in our sample to test the eight structural models underlying items of the STICSA that have been employed in prior studies (see Models 1–8 in **Supplementary Materials**)¹.

Model 1 – a one-factor model, in which all items were forced to load on a single higher order factor (Grös et al., 2007);

Model 2 – a two factor oblique model (CS-SS), in which items in the State scale loaded on either Cognitive and Somatic factors (Ree et al., 2008);

Model 3 – a two factor oblique model (CT-ST), in which items in the Trait scale loaded on either Cognitive and Somatic factors (Ree et al., 2008);

Model 4 – a two factor oblique model (S-T), in which items loaded on either State or Trait factors (Grös et al., 2007);

Model 5 – a two factor oblique model (C-S), in which items loaded on Cognitive or Somatic factors (Grös et al., 2007);

Model 6 – a four factor oblique model, in which the CT, ST, CS, SS subscales were directly modeled (Grös et al., 2007; Roberts et al., 2016);

Model 7 – a bifactor model, in which all STICSA items were forced to load both on a global anxiety factor and on 4 specific factors (CT, ST, CS, SS), corresponding to the STICSA subscales (Roberts et al., 2016);

Model 8 – a hierarchical model, with one higher order factor and four first-order factors, the CT, ST, CS, SS.

The robust weighted least squares (WLSMV) method using a diagonal weight matrix and robust standard errors and a mean- and variance adjusted χ^2 test statistic (Muthén, 1998; Muthén and Asparouhov, 2002; Muthén and Muthén, 2012b) was used to estimate parameters. The WLSMV is a robust estimator which does not assume normally distributed data (Brown, 2014) and seems to work well under a variety of conditions if sample size is 200 or better (Flora and Curran, 2004; Rhemtulla et al., 2012). Following Grös et al. (2007), in models 1 and 5, the error terms associated with corresponding items in the STICSA State and Trait were correlated. In these measurement models, the correlated error terms reflected a method effect (e.g., reversed/similarly worded items, acquiescence, or social desirability) (Marsh, 1996; Brown, 2014).

Model fit was assessed with the: (a) robust WLSMV chi-square (χ^2) statistic and its degrees of freedom; (b) Tucker Lewis Index (TLI); (c) comparative fit index (CFI); and (d) root mean square error of approximation (RMSEA) and its 90% confidence interval (90% CI). Due to the large sample size, interpretation of the robust WLSMV chi-square as a measure of fit was eschewed. An adequate fit between the target model and the observed data would produce TLI and CFI values of 0.90 and above, while values of 0.95 and above were considered to indicate excellent fit. RMSEA values of 0.08 or less were considered to reflect an adequate fit, while values of 0.05 or less were considered to reflect good fit (Schermelele-Engel et al., 2003; Brown, 2014).

To examine factor structure invariance (*measurement invariance*) across gender and age, multigroup CFAs were performed according to Muthén and Muthén (2012b), using the WLSMV method and theta parameterization. Configural invariance is established when factor loadings and thresholds are free across groups, residual variances fixed at one in all groups, and factor means fixed at zero in all groups. In the metric invariance model, factor loadings are constrained to be equal across groups, residual variances fixed at one in one group and free in the other groups, and factor means fixed at zero in one group and free in the other groups. Scalar invariance models had factor loadings and thresholds constrained to be equal across groups, residual variances fixed at one in one group and free in the other groups, and factor means fixed at one in one group and free in the other groups. Given the large sample size, chi-square difference tests would be overly sensitive to even trivial differences (Little et al., 2007). Therefore, evaluation of invariance was based on the difference (Δ) of CFI and RMSEA indexes (Chen, 2007). A change of CFI ≥ -0.010 between consecutive models and a change of RMSEA ≥ 0.015 was considered as non-invariance (Chen, 2007). To investigate

¹A series of additional models were also tested (i.e., Orthogonal two factor: State-Trait model; higher-order and bi-factor versions of the two-factor models). These additional models were under-identified or did not reach convergence. Therefore, they were considered unreliable and not informative, so they were not included in the present study.

concurrent validity of test score interpretations, Pearson correlations were calculated between scores on the STICSA and scores on the STAI-Y, BAI (for the convergent validity), TDI, and BDI-II (for the discriminant validity). We also compared the STICSA and STAI pairs of correlation coefficients in the analysis of discriminant validity following Meng et al. (1992). This procedure involves performing a Fisher Z transformation on the correlation coefficients so that they can be compared via a *t*-test.

MPLUS v7 (Muthén and Muthén, 2012a) was used for the confirmatory factor analyses, SPSS V.22 (Corp, 2013) was used for all descriptives, correlations, and alpha reliability coefficients. Also, R Statistic for hierarchical McDonald omega ($h\omega$) was used to estimate the reliability of the state and trait STICSA scales, since it is more accurate than coefficient alpha in multidimensional measures (Zinbarg et al., 2006; McDonald, 2013).

RESULTS

Confirmatory Factor Analysis

As expected, due to the large sample size, the chi-squared index was found to be significant for all models. However, only models 2 and 3 exhibited acceptable fit to the data, suggesting that the State and Trait scale of the STICSA, with each including Cognitive and Somatic dimensions well represented our STICSA Italian adaptation structure (see **Table 1**). The degree of relationship (standardized λ weights) for each item with its correspondent first-order factors were all significant ($p < 0.001$) in these two models (see **Supplementary Table 1**).

In Model 2, the STICSA State scale item loadings on the SS-CS factors ranged from 0.55 to 0.88, with an average standardized factor loading of 0.73. Squared multiple correlations ranged from 0.30 to 0.78, with an average SMC of 0.54 indicating that, on average, 29% of the variance in observed variables was accounted for by latent factors. The latent factor correlations were high (0.73). In Model 3, the standardized factor loadings of the STICSA Trait items ranged from 0.49 to 0.79 for the CT-ST factors, with an average standardized factor loading of 0.67. Squared multiple correlations ranged from 0.25 to 0.64, with

an average SMC of 0.46 indicating that, on average, 21% of the variance in observed variables was accounted for by latent factors. The latent factor correlations were high (75). In terms of local misfit, a careful inspection of the modification index did not suggested a respecification of either Model 2 or Model 3.

Multigroup CFA

Tests of measurement invariance across gender and age were examined through a multiple-group confirmatory factor analysis. Based on the previous findings, Models 2 and 3 were used as baseline models and tested for the data fit across: (a) male versus female groups for the first comparison and (b) age groups (18–25; 26–50; 51–99 years) for the second comparison. Following the (Muthén and Muthén, 2012b, p. 545) sequential procedures in each comparison, the fit of Models 2 and 3 were first tested separately in groups. Then, restrictive models were used to test for equal form (configural invariance), equal factor loadings (metric invariance), and equal indicator thresholds and residual variances (scalar invariance). Results of these measurement invariance analyses are presented in **Table 2**.

Configural, metric, and scalar invariance was demonstrated across male and female groups. As seen in **Table 2**, the Δ CFI were lower than |0.010| and RMSEA were lower than |0.015| for all the comparisons, therefore the assumption of equal factor loadings and indicator thresholds in males and females group were confirmed for Models 2 and 3. However, the χ^2 difference between all models tested was found to be significant ($p < 0.001$), both at State (Model 2) and at Trait (Model 3) scale of STICSA.

For age groups, measurement invariance was found for Models 2 and 3 in each of the three groups, separately. Subsequently, the adequacy of the same models was examined through the amount of configural, metric, and scalar invariance simultaneously in the three age groups (18–25; 26–50; 51–99 years old). Fit indices in general supported an adequate model fit for configural, metric, and scalar invariance across age for Model 2 (State scale of STICSA, with Cognitive and Somatic subscales). Configural invariance was also established for Model 3 across the three age groups. Fit indices showed that significant differences across the age groups were found on factor loadings, item thresholds, and residual levels for Model 3. All Δ CFI were greater than |0.01| cut-off criteria;

TABLE 1 | Fit indices for the structural models tested ($N = 2,983$).

Model	χ^2	df	TLI	CFI	RMSEA	90% CI
Model 1: One factor with correlated error	17461.732*	799	0.81	0.82	0.084	0.083–0.085
Model 2: Two state factors: SS and CS	2419.063*	188	0.95	0.95	0.063	0.061–0.065
Model 3: Two trait factors: ST and CT	2589.648*	188	0.94	0.94	0.065	0.063–0.068
Model 4: Two factors: State and trait	28848.444*	818	0.69	0.71	0.107	0.106–0.108
Model 5: Two factors: Somatic and cognitive with correlated error	10112.356*	797	0.89	0.90	0.063	0.062–0.064
Model 6: Four factors (ST, CT, SS, CS)	18244.376*	813	0.81	0.82	0.085	0.084–0.085
Model 7: Bifactor with four group factors (ST, CT, SS, CS) and one general factor	25349.554*	778	0.71	0.74	0.103	0.102–0.104
Model 8: Hierarchical structure	24872.698*	815	0.74	0.75	0.099	0.098–0.101

* $p < 0.001$. Model 7-8: (Roberts et al., 2016); Model 2-3: (Ree et al., 2008); Model 4-5-1: (Grös et al., 2007).

ST, STICSA somatic trait; CT, STICSA cognitive trait; SS, STICSA somatic state; CS, STICSA cognitive state; df, degrees of freedom; TLI, tucker lewis index; CFI, comparative fit index; RMSEA, root-mean-square error of approximation; 90% CI, 90% confidence interval of RMSEA.

TABLE 2 | Tests of measurement invariance across gender and age.

Model	χ^2	df	Δdf	TLI	CFI	ΔCFI	RMSEA	90% CI	$\Delta RMSEA$
GENDER†									
STATE									
Single-group solutions									
Male (n = 1,316)	1011.724*	188		0.957	0.962		0.058	0.054–0.061	
Female (n = 1,667)	1468.362*	188		0.941	0.947		0.064	0.061–0.067	
Configural	2497.532*	376		0.948	0.953		0.062	0.059–0.064	
Metric	2560.230*	394	18	0.949	0.953	0.000	0.061	0.058–0.063	–0.001
Scalar	2380.610*	439	45	0.959	0.957	–0.004	0.054	0.052–0.057	–0.007
TRAIT									
Single-group solutions									
Male (n = 1,316)	1133.253*	188		0.943	0.949		0.062	0.058–0.065	
Female (n = 1,667)	1573.278*	188		0.927	0.934		0.067	0.064–0.070	
Configural	2719.534*	376		0.933	0.940		0.065	0.062–0.067	
Metric	2841.328*	394	18	0.934	0.938	–0.002	0.065	0.062–0.067	0.000
Scalar	2789.981*	439	45	0.943	0.940	0.002	0.060	0.058–0.062	0.005
AGE‡									
STATE									
Single-group solutions									
Age 18–25 (n = 1,556)	1306.368*	188		0.949	0.954		0.062	0.059–0.065	
Age 26–50 (n = 675)	600.368*	188		0.956	0.961		0.057	0.052–0.062	
Age 51–99 (n = 742)	713.986*	188		0.951	0.956		0.061	0.057–0.066	
Configural	2606.486*	564		0.949	0.954		0.060	0.058–0.063	
Metric	2904.808*	664	100	0.952	0.950	0.004	0.058	0.056–0.061	0.002
Scalar	2674.628*	711	47	0.961	0.956	–0.006	0.053	0.051–0.055	–0.005
TRAIT									
Single-group solutions									
Age 18–25 (n = 1,556)	1557.729*	188		0.924	0.932		0.068	0.065–0.072	
Age 26–50 (n = 675)	589.854*	188		0.949	0.954		0.056	0.051–0.061	
Age 51–99 (n = 742)	604.609*	188		0.958	0.962		0.055	0.050–0.060	
Configural	2745.246*	564		0.937	0.944		0.062	0.060–0.065	
Metric	3377.579*	664	100	0.934	0.930	0.014	0.064	0.062–0.066	0.002
Scalar	4015.815*	711	47	0.925	0.915	0.015	0.068	0.066–0.071	0.004

df, degrees of freedom; TLI, tucker lewis index; CFI, comparative fit index; RMSEA, root-mean-square error of approximation; 90% CI = 90% confidence interval of RMSEA.

* $p < 0.001$.

†N = 2,983.

‡N = 2,974, from the total sample we excluded nine cases with missing values in age variable.

therefore, metric and scalar invariance between the age groups was not confirmed for model 3. A careful inspection of modification index (MI) revealed that the factor loadings of items 10–11 and 20, respectively, for the first (18–25 years) and third age group (51–99 years), differed significantly between groups.

Descriptives and Concurrent Validity

Means, standard deviations, correlations, and internal consistency are reported in Table 3. The reliability estimates for the STICSA subscale were high, with ω coefficients of 0.96 and 0.94, respectively, for State and Trait total scores. In order to investigate the concurrent and discriminant validity of the STICSA, one-tailed correlations among the STICSA dimensions with other measure of anxiety and measures of depression were computed (Table 3).

Convergent Validity

As expected, all the STICSA scales were highly inter-correlated (ranged from $r = 0.916$ to $r = 0.465$, $p < 0.001$, respectively, for STICSA Trait Somatic and STICSA Trait Cognitive). Additionally, STICSA Trait and State scale dimensions were medium to highly correlated with the STAI-Y scales (from $r = 0.502$ to $r = 0.699$, $p < 0.001$, and from $r = 0.483$ to $r = 0.735$, $p < 0.001$, respectively, for STAI-State and STAI-Trait), but moderately correlated with BAI total scores (from $r = 0.386$, to $r = 0.417$, $p < 0.001$, respectively, for STICSA State Cognitive and STICSA Trait Cognitive).

Discriminant Validity

All anxiety dimensions used in this study correlated moderately with depression measures. Notably, the Somatic subscale of the Trait STICSA was correlated with depression measures

TABLE 3 | Descriptives, correlations and reliabilities.

	Mean	SD	α (ω)	1	2	3	4	5	6	7	8	9
1. STICSA-Trait, somatic	18.62	5.69	(0.89)		0.644**	0.715**	0.554**	0.451**	0.502**	0.571**	0.422**	0.354**
2. STICSA-Trait, cognitive	19.10	6.08	(0.90)			0.465**	0.780**	0.417**	0.598**	0.735**	0.577**	0.603**
3. STICSA-State, somatic	16.21	5.61	(0.92)				0.626**	0.421**	0.542**	0.483**	0.354**	0.248**
4. STICSA-State, cognitive	17.17	6.28	(0.91)					0.386**	0.699**	0.634**	0.538**	0.592**
5. BAI [†]	12.65	12.64	0.95						0.489**	0.583**	0.442**	0.387**
6. STAI-Y State [‡]	37.38	11.51	0.94							0.751**	0.548**	.b
7. STAI-Y Trait [‡]	39.95	10.73	0.91								0.706**	.b
8. TDI	27.95	13.09	0.91									0.670**
9. BDI-II [§]	11.29	8.57	0.83									

[†]N = 722.

[‡]N = 444.

[§]N = 278.

.b Cannot be computed because there are not enough subjects who completed the BDI-II scale.

** $p < 0.001$ (one-tailed).

STICSA, state-trait inventory for cognitive and somatic anxiety; BAI, beck anxiety inventory; STAI-Y, state-trait anxiety inventory – form Y; TDI, teate depression inventory; BDI-II, beck depression inventory-II.

($r_{TDI} = 0.422$, $r_{BDI-II} = 0.354$; $p < 0.001$); as well as the Somatic subscale of the State scale ($r_{TDI} = 0.354$, $r_{BDI-II} = 0.248$; $p < 0.001$). No correlations were computed between the STAI-Y and the BDI-II, given the low number of participants who completed the BDI-II.

Subsequently, correlation coefficients between the STICSA State and Trait dimensions and STAI Trait and State anxiety and the TDI depression were statistically compared (Meng et al., 1992). Comparisons revealed that the STAI Trait correlated more highly with TDI scores than with STICSA Trait somatic, and cognitive subscales [$t_{(441)} = 9.09$, $p < 0.01$, $Z = 8.38$; $t_{(441)} = 5.29$, $p < 0.01$, $Z = 5.12$, respectively]. Similarly, the STAI State scale correlated more highly with TDI scores than with STICSA-State somatic subscale scores [$t_{(441)} = 5.10$, $p < 0.01$, $Z = 4.93$]. No differences were found between the STAI State and STICSA State cognitive with the TDI [$t_{(441)} = 0.33$, $p = 0.37$, $Z = 0.33$, respectively]. These results, in line with previous research (Grös et al., 2007), indicated that the STICSA State somatic, Trait scale, and cognitive and somatic subscales were better measures of anxiety than depression.

DISCUSSION

The STICSA was developed to overcome the psychometric weakness of existing instruments of anxiety based on the distinction between trait and state anxiety (i.e., the STAI), such as their extensive overlap with depression (Caci et al., 2003; Balsamo et al., 2013c; Roberts et al., 2016). Even though the STICSA State and Trait scale and subscales have exhibited high internal consistency reliability, as well as construct consistent correlations in patients, controls, and community groups (Grös et al., 2007; Ree et al., 2008; Van Dam et al., 2013), no consensus was found in the literature about the factor structure of the STICSA (Ree et al., 2008; Lancaster et al., 2015; Roberts et al., 2016).

The present study, firstly, provided further evidence that scores from the Italian adaptation of the STICSA were reliable and valid measures of multidimensional (cognitive and somatic)

anxiety in a non-clinical population. Consistent with some previous research, the confirmatory factor analysis confirmed the STICSA factor structure of the Trait and State scales as separate measures of anxiety (Ree et al., 2000, 2008; Deacy et al., 2016). Each of the State and Trait forms was composed of two latent and correlated factors, thereby lending support to the distinction between cognitive and somatic aspects of anxiety. No support was found for the hierarchical and bifactor model of the STICSA scores with a global anxiety factor and four specific factors corresponding to the four subscales of the STICSA (trait/state; cognitive/somatic) (Roberts et al., 2016), nor for an oblique four-factor model of STICSA scores with factors corresponding to the somatic and cognitive subscales of the state and trait versions of the STICSA previously found in an elderly population (Balsamo et al., 2015b). This result was not in accordance with the increasing number of studies which have supported a bifactor structure for psychopathological scales (Al-Turkait and Ohaeri, 2010; Kriston et al., 2012; Saggino et al., 2018; Wang et al., 2018).

The second aim of the study was to assess the measurement equivalence of the STICSA scores across males and females, and across young, middle age, and older adult samples in order to determine whether scores between these groups could be interpreted with confidence. For gender comparisons, results indicated that the STICSA State and Trait scale items showed the same consistency in factor structure across male and female respondents. Given the empirical evidence that females have demonstrated greater negative affectivity (such as trait anxiety) and higher rates of anxiety disorders and symptomatology (Kessler et al., 1994; Breslau et al., 2000) than men across the life span (Lewinsohn et al., 1998; McLean et al., 2011), this finding appears to be interesting since the STICSA factor structure it was found invariant across gender in this study. Full measurement invariance across gender suggested that the proposed factor structure, pattern of factor loadings, and thresholds of STICSA State and Trait scales were similar for males and female respondents in this study. Therefore, STICSA State and Trait scores appear

to reflect true gender differences in anxiety constructs and can be used interchangeably in males and females (Brown, 2014).

Concerning age, only the STICSA State scale was found to be invariant at the configural, metric, and scalar levels. Metric and scalar measurement equivalence was not found across age groups for the STICSA Trait scale. There is a general consensus in the literature on the impact of age in developing anxious symptomatology (Christensen et al., 1999; Balsamo et al., 2018). The nature of the anxiety experienced by older individuals may differ qualitatively from younger ones. For instance, older people reported greater ability to control their emotions (Lawton et al., 1993) and greater level of worry about health; whereas younger adults experienced worry about finances and family and tended to report more negative affect.

In our sample, younger, middle-aged, and older adult groups interpreted and responded to the Trait scale of the STICSA with significant variability between them. They differed significantly about how much of the latent trait was required to endorse an item. Great age-variability was found across items that assess somatic conditions (“*Butterflies in the stomach*”), and cognitive process (“*Can’t get thoughts out of mind*” and “*Trouble remembering things*”). Given these age differences, clinicians might misrepresent cognitive and somatic symptomatology or over/underestimate the magnitude of state anxiety reactions under stressful circumstances across gender groups (Ree et al., 2008). Therefore, future research should examine in detail the capacity of these items in discriminating trait anxiety across age.

In line with previous research, all the STICSA Trait and State scores of the Cognitive and Somatic scales were highly inter-correlated (Ree et al., 2000, 2008; Grös et al., 2007; Balsamo et al., 2015b; Roberts et al., 2016). The STICSA showed good convergent validity with the STAI, moderate convergent validity with the BAI, and satisfactory discriminant validity with the TDI and the BDI-II. In line with the Clark and Watson (1991) tripartite model, our results suggested that anxiety and depression shared a non-specific component of generalized distress (negative affect). In addition, STICSA State and Trait measures of Cognitive and Somatic symptoms were more specific to anxiety (i.e., physiological hyperarousal) than depression compared to the STAI. Similarly, the moderate to strong correlations between STICSA (and its subscale) scores and concurrent measures of anxiety in this study provided further evidence of STICSA scales as a pure measure of anxiety (Innamorati et al., 2013).

Limitations of the present study included the characteristics of our sample and the specific measures selected for the

validity analyses. The use of a convenience sample, composed of non-clinical participants (mostly undergraduate students), potentially limits the generalizability of this study (Peterson and Merunka, 2014). Additionally, the inclusion of student data in research might have introduced uncontrolled systematic variance components (Balsamo, 2010, 2013; Balsamo et al., 2013b; Innamorati et al., 2014). As the specific measures selected for the validity analyses the STICSA was investigated exclusively in relation to a measure of general anxiety (i.e., BAI), neglecting measures of specific anxiety disorder (i.e., Panic Attack and Anticipatory Anxiety Scale; the Anxiety Sensitivity Index). Another limitation was reliance on unbalance samples to perform the correlation analyses between the measures. This is, partially, due to the sample recruitment strategy. Further studies could address the issue of comorbidity in clinical samples, controlling the STICSA State and Trait scale scores for depression (i.e., MIMIC models) or including depression as a covariate in regression models.

Regardless of these limitations, the current study demonstrated that STICSA scores are psychometrically reliable and valid measures that discriminated anxiety from depression in a non-clinical Italian population. The ability of STICSA to distinguish State and Trait dimensions of Cognitive and Somatic anxiety could provide a helpful opportunity for clinicians to: (a) perform an accurate differential diagnosis (e.g., discriminating anxiety from somatic symptomatology in oncology and geriatrics as well as in medical conditions); (b) promote recognition and effective treatment of anxiety disorders and comorbid disorders; (c) prove the efficacy of certain treatments in reducing specific anxiety symptoms. Future research should examine this discriminant power in association with specific symptoms of anxiety.

AUTHOR CONTRIBUTIONS

LC and MB designed the study and conducted the statistical analyses. LC, MW, AS, and MB interpreted the data. LC, MB, and MW drafted the manuscript. MS and FC recruited the sample and collaborated in editing the final manuscript. All authors contributed toward data analysis, drafting and revising the paper, and agreed to be accountable for all aspects of the work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02345/full#supplementary-material>

REFERENCES

- Allen, B. P., and Potkay, C. R. (1981). On the arbitrary distinction between states and traits. *J. Pers. Soc. Psychol.* 41, 916–928. doi: 10.1037/0022-3514.41.5.916
- Allport, G. W., and Odbert, H. S. (1936). Trait-names: a psycho-lexical study. *Psychol. Monogr.* 47:i-171. doi: 10.1037/h0093360
- Al-Turkait, F. A., and Ohaeri, J. U. (2010). Dimensional and hierarchical models of depression using the Beck Depression Inventory-II in an Arab college student sample. *BMC Psychiatry* 10:60. doi: 10.1186/1471-244X-10-60
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Washington, DC: American Psychiatric Pub. doi: 10.1176/appi.books.9780890425596

- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., et al. (2004). A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages. *J. Pers. Soc. Psychol.* 86, 356–366. doi: 10.1037/0022-3514.86.2.356
- Bados, A., Gómez-Benito, J., and Balaguer, G. (2010). The state-trait anxiety inventory, trait version: does it really measure anxiety? *J. Pers. Assess.* 92, 560–567. doi: 10.1080/00223891.2010.513295
- Balsamo, M. (2010). Anger and depression: evidence of a possible mediating role for rumination. *Psychol. Rep.* 106, 3–12. doi: 10.2466/PRO.106.1.3-12
- Balsamo, M. (2013). Personality and depression: evidence of a possible mediating role for anger trait in the relationship between cooperativeness and depression. *Compr. Psychiatry* 54, 46–52. doi: 10.1016/j.comppsy.2012.05.007
- Balsamo, M., Carlucci, L., Sergi, M., and Saggino, A. (2016a). Validazione della versione italiana del Co-Rumination Questionnaire [Validation of the Italian version of the Co-Rumination Questionnaire]. *Psicoter. Cognit. Comport. Ital. J. Cogn. Behav. Psychot.* 22, 13–34.
- Balsamo, M., Carlucci, L., Sergi, M. R., Romanelli, R., D'Ambrosio, I., Fairfield, B., et al. (2016b). A new measure for trait and state anxiety: the state trait inventory of cognitive and somatic anxiety (STICSA). standardization in an Italian population. *Psicot. Cognit. Comport. Ital. J. Cogn. Behav. Psychot.* 22, 229–232.
- Balsamo, M., Carlucci, L., Sergi, M. R., Murdock, K. K., and Saggino, A. (2015a). The mediating role of early maladaptive schemas in the relation between co-rumination and depression in young adults. *PLoS One* 10:e0140177. doi: 10.1371/journal.pone.0140177
- Balsamo, M., Innamorati, M., Van Dam, N. T., Carlucci, L., and Saggino, A. (2015b). Measuring anxiety in the elderly: psychometric properties of the state trait inventory of cognitive and somatic anxiety (STICSA) in an elderly Italian sample. *Int. Psychogeriatr.* 27, 999–1008. doi: 10.1017/S1041610214002634
- Balsamo, M., Macchia, A., Carlucci, L., Picconi, L., Tommasi, M., Gilbert, P., et al. (2015c). Measurement of external shame: an inside view. *J. Pers. Assess.* 97, 81–89. doi: 10.1080/00223891.2014.947650
- Balsamo, M., Cataldi, F., Carlucci, L., and Fairfield, B. (2018). Assessment of anxiety in older adults: a review of self-report measures. *Clin. Interv. Aging* 2018, 573–593. doi: 10.2147/CIA.S114100
- Balsamo, M., Giampaglia, G., and Saggino, A. (2014). Building a new rasch-based self-report inventory of depression. *Neuropsychiatr. Dis. Treat.* 10, 153–165. doi: 10.2147/NDT.S53425
- Balsamo, M., Imperatori, C., Sergi, M. R., Belvederi Murri, M., Continisio, M., Tamburello, A., et al. (2013a). Cognitive vulnerabilities and depression in young adults: an ROC curves analysis. *Depress Res. Treat.* 2013, 1–8. doi: 10.1155/2013/407602
- Balsamo, M., Lauriola, M., and Saggino, A. (2013b). Work values and college major choice. *Learn. Individ. Differ.* 24, 110–116. doi: 10.1016/j.lindif.2012.12.022
- Balsamo, M., Romanelli, R., Innamorati, M., Ciccacese, G., Carlucci, L., and Saggino, A. (2013c). The state-trait anxiety inventory: shadows and lights on its construct validity. *J. Psychopat. Behav.* 35, 475–486. doi: 10.1007/s10862-013-9354-5
- Balsamo, M., and Saggino, A. (2007). Test per l'assessment della depressione nel contesto italiano: un'analisi critica. *Psicoter. Cognit. Comport. Ital. J. Cogn. Behav. Psychot.* 13, 167–199.
- Balsamo, M., and Saggino, A. (2013). *Teate Depression Inventory: Manuale*. Göttingen: Hogrefe.
- Balsamo, M., and Saggino, A. (2014). Determining a diagnostic cut-off on the teate depression inventory. *Neuropsychiatr. Dis. Treat.* 10, 987–995. doi: 10.2147/NDT.S55706
- Bandelow, B., and Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialog. Clin. Neurosci.* 17, 327–335.
- Beck, A., Steer, R., and Brown, G. (1996). *Beck Depression Inventory-II (BDI-II)*. San Antonio, TX: Psychological Corporation.
- Beck, A. T., Epstein, N., Brown, G., and Steer, R. A. (1988). An inventory for measuring clinical anxiety: psychometric properties. *J. Consult. Clin. Psychol.* 56, 893–897. doi: 10.1037/0022-006X.56.6.893
- Bergua, V., Meillon, C., Potvin, O., Ritchie, K., Tzourio, C., Bouisson, J., et al. (2016). Short stai-y anxiety scales: validation and normative data for elderly subjects. *Aging Ment. Health* 20, 987–995. doi: 10.1080/13607863.2015.1051511
- Bieling, P. J., Antony, M. M., and Swinson, R. P. (1998). The state-trait anxiety inventory, trait version: structure and content re-examined. *Behav. Res. Ther.* 36, 777–788. doi: 10.1016/S0005-7967(98)00023-0
- Blazer, D., George, L.K., and Hughes, D. (1991). "The epidemiology of anxiety disorders: an age comparison," in *Anxiety in the Elderly: Treatment and Research*, eds C. Salzman and B. D. Lebowitz (New York, NY: Pringer Publishing Co.), 17–30.
- Breslau, N., Chilcoat, H. D., Peterson, E. L., and Schultz, L. R. (2000). "Gender differences in major depression," in *Gender and Its Effects on Psychopathology*, ed. E. Frank (Washington, DC: American Psychiatric Pub).
- Brown, T. A. (2014). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Publications.
- Bruce, S. E., Yonkers, K. A., Otto, M. W., Eisen, J. L., Weisberg, R. B., Pagano, M., et al. (2005). Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: a 12-year prospective study. *Am. J. Psychiatry* 162, 1179–1187. doi: 10.1176/appi.ajp.162.6.1179
- Bryant, C., Jackson, H., and Ames, D. (2008). The prevalence of anxiety in older adults: methodological issues and a review of the literature. *J. Affect. Disord.* 109, 233–250. doi: 10.1016/j.jad.2007.11.008
- Buss, A. H. (1962). Two anxiety factors in psychiatric patients. *J. Abnorm. Soc. Psychol.* 65, 426–427. doi: 10.1037/h0040152
- Caci, H., Baylé, F. J., Dossios, C., Robert, P., and Boyer, P. (2003). The Spielberger trait anxiety inventory measures more than anxiety. *Eur. Psychiatry* 18, 394–400. doi: 10.1016/j.eurpsy.2003.05.003
- Carlucci, L., D'Ambrosio, I., Innamorati, M., Saggino, A., and Balsamo, M. (2018). Co-rumination, anxiety, and maladaptive cognitive schemas: when friendship can hurt. *Psychol. Res. Behav. Manage.* 2018, 133–144. doi: 10.2147/PRBM.S144907
- Carr, H., and Kingsbury, F. (1938). The concept of traits. *Psychol. Rev.* 45, 497–524. doi: 10.1037/h0063048
- Cattell, R. B. (1946). *Description and Measurement of Personality*. Oxford: World Book.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equat. Model.* 14, 464–504. doi: 10.1080/10705510701301834
- Christensen, H., Jorm, A., Mackinnon, A., Korten, A., Jacomb, P., Henderson, A., et al. (1999). Age differences in depression and anxiety symptoms: a structural equation modelling analysis of data from a general population sample. *Psychol. Med.* 29, 325–339. doi: 10.1017/S0033291798008150
- Clark, L. A., and Watson, D. (1991). Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *J. Abnorm. Psychol.* 100, 316–336. doi: 10.1037/0021-843X.100.3.316
- Contardi, A., Imperatori, C., Amati, I., Balsamo, M., and Innamorati, M. (2018). Assessment of affect lability: psychometric properties of the ALS-18. *Front. Psychol.* 9:427. doi: 10.3389/fpsyg.2018.00427
- Corp, I. (2013). *IBM SPSS Statistics for Windows, Version 22.0*. Armonk, NY: IBM Corp.
- Costa, P. T., and MacCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional Manual*. Psychological Assessment Resources, Incorporated. Odessa, FL: Psychological Assessment Resources.
- Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., and Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *AMA Arch. Gen. Psychiatry* 60, 837–844. doi: 10.1001/archpsyc.60.8.837
- de Beurs, E., Wilson, K. A., Chambless, D. L., Goldstein, A. J., and Feske, U. (1997). Convergent and divergent validity of the Beck Anxiety Inventory for patients with panic disorder and agoraphobia. *Depress. Anxiety* 6, 140–146. doi: 10.1002/(SICI)1520-6394(1997)6:4<140::AID-DA2>3.0.CO;2-G
- de Girolamo, G., Polidori, G., Morosini, P., Scarpino, V., Reda, V., Serra, G., et al. (2006). Prevalence of common mental disorders in Italy. *Soc. Psychiatry Psychiatry Epidemiol.* 41, 853–861. doi: 10.1007/s00127-006-0097-4
- Deacy, A. D., Gayes, L. A., De Lurgio, S., and Wallace, D. P. (2016). Adaptation of the state-trait inventory for cognitive and somatic anxiety for use in children: a preliminary analysis. *J. Pediatr. Psychol.* 41, 1033–1043. doi: 10.1093/jpepsy/jsw030

- Dennis, R., Boddington, S., and Funnell, N. (2007). Self-report measures of anxiety: are they suitable for older adults? *Aging Ment Health* 11, 668–677. doi: 10.1080/13607860701529916
- Diefenbach, G. J., Tolin, D. F., Meunier, S. A., and Gilliam, C. M. (2009). Assessment of anxiety in older home care recipients. *Gerontologist* 49, 141–153. doi: 10.1093/geront/gnp019
- Dozois, D. J., Dobson, K. S., and Ahnberg, J. L. (1998). A psychometric evaluation of the beck depression inventory-II. *Psychol. Assess.* 10, 83–89. doi: 10.1037/1040-3590.10.2.83
- Egger, H. L., Costello, J. E., and Angold, A. (2003). School refusal and psychiatric disorders: a community study. *J. Am. Acad. Child Adolesc. Psychiatry* 42, 797–807. doi: 10.1097/01.CHL.0000046865.56865.79
- Elwood, L. S., Wolitzky-Taylor, K., and Olatunji, B. O. (2012). Measurement of anxious traits: a contemporary review and synthesis. *Anxiety Stress Copin* 25, 647–666. doi: 10.1080/10615806.2011.582949
- Endler, N. S. (1997). Stress, Anxiety and coping: the multidimensional interaction model. *Can. Psychol.* 38, 136–153. doi: 10.1037/0708-5591.38.3.136
- Endler, N. S., and Kocovski, N. L. (2001). State and trait anxiety revisited. *J. Anxiety Disord.* 15, 231–245. doi: 10.1016/S0887-6185(01)00060-3
- Eysenck, H. (1953). The scientific study of personality. *Br. J. Math. Stat. Psychol.* 6, 44–52. doi: 10.1111/j.2044-8317.1953.tb00132.x
- Flint, A. J. (2005). Anxiety and its disorders in late life: moving the field forward. *Am. J. Geriatr. Psychiatry* 13, 3–6. doi: 10.1097/00019442-200501000-00002
- 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
- Freud, S. (1920). *A General Introduction to Psychoanalysis*. New York, NY: Boni and Liveright. doi: 10.1037/10667-000
- Fridhandler, B. M. (1986). Conceptual note on state, trait, and the state-trait distinction. *J. Pers. Soc. Psychol.* 50, 169–174. doi: 10.1037/0022-3514.50.1.169
- Gallagher, D., Breckenridge, J., Steinmetz, J., and Thompson, L. (1983). The beck depression inventory and research diagnostic criteria: congruence in an older population. *J. Consult. Clin. Psychol.* 51, 945–946. doi: 10.1037/0022-006X.51.6.945
- Godfrey, K., Yung, A., Killackey, E., Cosgrave, E., Pan Yuen, H., Stanford, C., et al. (2005). Patterns of current comorbidity in young help-seekers: implications for service planning and delivery. *Australas Psychiatry* 13, 379–383. doi: 10.1080/j.1440-1665.2005.02217.x
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychol. Assess.* 4, 26–42. doi: 10.1037/1040-3590.4.1.26
- Grös, D. F., Antony, M. M., Simms, L. J., and McCabe, R. E. (2007). Psychometric properties of the state-trait inventory for cognitive and somatic anxiety (STICSA): comparison to the state-trait anxiety inventory (STAI). *Psychol. Assess.* 19, 369–381. doi: 10.1037/1040-3590.19.4.369
- Grös, D. F., Simms, L. J., and Antony, M. M. (2010). Psychometric properties of the state-trait inventory for cognitive and somatic anxiety (STICSA) in friendship dyads. *Behav. Ther.* 41, 277–284. doi: 10.1016/j.beth.2009.07.001
- Hasler, G., Drevets, W. C., Manji, H. K., and Charney, D. S. (2004). Discovering endophenotypes for major depression. *Neuropsychopharmacology* 29, 1765–1781. doi: 10.1038/sj.npp.1300506
- Heeren, A., Bernstein, E. E., and McNally, R. J. (2018). Deconstructing trait anxiety: a network perspective. *Anxiety Stress Copin* 31, 262–276. doi: 10.1080/10615806.2018.1439263
- Hill, B. D., Musso, M., Jones, G. N., Pella, R. D., and Gouvier, W. D. (2013). A psychometric evaluation of the STAI-Y, BDI-II, and PAI using single and multifactorial models in young adults seeking psychoeducational evaluation. *J. Psychoeduc. Assess.* 31, 300–312. doi: 10.1177/0734282912462670
- Innamorati, M., Lester, D., Balsamo, M., Erbutto, D., Ricci, F., Amore, M., et al. (2014). Factor validity of the beck hopelessness scale in Italian medical patients. *J. Psychopat. Behav.* 36, 300–307. doi: 10.1007/s10862-013-9380-3
- Innamorati, M., Tamburello, S., Contardi, A., Imperatori, C., Tamburello, A., Saggino, A., et al. (2013). Psychometric properties of the attitudes toward self-revised in Italian young adults. *Depress Res. Treat.* 2013, 1–6. doi: 10.1155/2013/209216
- Kabacoff, R. I., Segal, D. L., Hersen, M., and Van Hasselt, V. B. (1997). Psychometric properties and diagnostic utility of the beck anxiety inventory and the state-trait anxiety inventory with older adult psychiatric outpatients. *J. Anxiety Disord.* 11, 33–47. doi: 10.1016/S0887-6185(96)00033-3
- Kennedy, B. L., Schwab, J. J., Morris, R. L., and Beldia, G. (2001). Assessment of state and trait anxiety in subjects with anxiety and depressive disorders. *Psychiatry* 72, 263–276. doi: 10.1023/A:1010305200087
- Kessler, R. C., Angermeyer, M., Anthony, J. C., De Graaf, R., Demyttenaere, K., Gasquet, I., et al. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry* 6, 168–176.
- Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., et al. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the national comorbidity survey. *Arch. Gen. Psychiatry* 51, 8–19. doi: 10.1001/archpsyc.1994.03950010080002
- Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., and Wittchen, H. U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *Int. J. Methods Psychiatr. Res.* 21, 169–184. doi: 10.1002/mpr.1359
- Kriston, L., Schäfer, J., von Wolff, A., Härter, M., and Hölzel, L. P. (2012). The latent factor structure of Young's early maladaptive schemas: are schemas organized into domains? *J. Clin. Psychol.* 68, 684–698. doi: 10.1002/jclp.21846
- Lancaster, S. L., Melka, S. E., Klein, K. P., and Rodriguez, B. F. (2015). Ethnicity and anxiety: a psychometric evaluation of the STICSA. *Measur. Eval. Counsel. Dev.* 48, 163–176. doi: 10.1177/0748175615578757
- Lawton, M. P., Kleban, M. H., and Dean, J. (1993). Affect and age: cross-sectional comparisons of structure and prevalence. *Psychol. Aging* 8, 165–175. doi: 10.1037/0882-7974.8.2.165
- Lewinsohn, P. M., Gotlib, I. H., Lewinsohn, M., Seeley, J. R., and Allen, N. B. (1998). Gender differences in anxiety disorders and anxiety symptoms in adolescents. *J. Abnorm. Psychol.* 107, 109–117. doi: 10.1037/0021-843X.107.1.109
- Lilienfeld, S. O., Turner, S. M., and Jacob, R. G. (1993). Anxiety sensitivity: an examination of theoretical and methodological issues. *Adv. Behav. Res. Ther.* 15, 147–183. doi: 10.1016/0146-6402(93)90019-X
- Little, T. D., Card, N. A., Slegers, D. W., and Ledford, E. C. (2007). "Representing contextual effects in multiple-group MACS models," in *Modeling Contextual Effects in Longitudinal Studies*, eds T. D. Little, J. A. Bovaird, and N. A. Card (Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers), 121–147.
- Luthans, F., Avolio, B. J., Avey, J. B., and Norman, S. M. (2007). Positive psychological capital: measurement and relationship with performance and satisfaction. *Pers. Psychol.* 60, 541–572. doi: 10.1111/j.1744-6570.2007.00083.x
- Marsh, H. W. (1996). Positive and negative global self-esteem: a substantively meaningful distinction or artifacts? *J. Pers. Soc. Psychol.* 70, 810–819. doi: 10.1037/0022-3514.70.4.810
- McDonald, R. J., and Spielberger, C. D. (1983). Measuring anxiety in hospitalized geriatric patients. *Ser. Clin. Commun. Psychol. Stress Anxiety* 2, 135–143.
- McDonald, R. P. (2013). *Test Theory: A Unified Treatment*. London: Psychology Press. doi: 10.4324/9781410601087
- McLean, C. P., and Anderson, E. R. (2009). Brave men and timid women? A review of the gender differences in fear and anxiety. *Clin. Psychol. Rev.* 29, 496–505. doi: 10.1016/j.cpr.2009.05.003
- McLean, C. P., Asnaani, A., Litz, B. T., and Hofmann, S. G. (2011). Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. *J. Psychiatr. Res.* 45, 1027–1035. doi: 10.1016/j.jpsychires.2011.03.006
- Meng, X.-L., Rosenthal, R., and Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychol. Bull.* 111, 172–175. doi: 10.1037/0033-2909.111.1.172
- Muthén, B., and Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Notes* 4, 1–22.
- Muthén, B. O. (1998). *Mplus Technical Appendices*. Los Angeles, CA: Muthén & Muthén 2004.
- Muthén, L., and Muthén, B. (2012a). *Mplus Statistical Modeling Software (Version 7)*. Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., and Muthén, B. O. (2012b). *Mplus Version 7 User's Guide*. Los Angeles, CA: Muthén & Muthén.
- Pacheco-Unguetti, A. P., Acosta, A., Callejas, A., and Lupiáñez, J. (2010). Attention and anxiety: different attentional functioning under state and trait anxiety. *Psychol. Sci.* 21, 298–304. doi: 10.1177/0956797609359624

- Peterson, R. A., and Merunka, D. R. (2014). Convenience samples of college students and research reproducibility. *J. Bus. Res.* 67, 1035–1041. doi: 10.1016/j.jbusres.2013.08.010
- Picconi, L., Jackson, C. J., Balsamo, M., Tommasi, M., and Saggino, A. (2018). Factor structure and measurement invariance across groups of the Italian Eysenck Personality Questionnaire-Short form (EPP-S). *Pers. Individ. Differ.* 123, 76–80. doi: 10.1016/j.paid.2017.11.013
- Ree, M., MacLeod, C., French, D., and Locke, V. (2000). The state-trait inventory for cognitive and somatic anxiety: development and validation. *Poster session presented at the annual meeting of the Association for the Advancement of Behavior Therapy*, New Orleans, LA.
- Ree, M. J., French, D., MacLeod, C., and Locke, V. (2008). Distinguishing cognitive and somatic dimensions of state and trait anxiety: development and validation of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA). *Behav. Cogn. Psychother.* 36, 313–332. doi: 10.1017/S1352465808004232
- Regier, D. A., Narrow, W. E., and Rae, D. S. (1990). The epidemiology of anxiety disorders: the epidemiologic catchment area (ECA) experience. *J. Psychiatr. Res.* 24, 3–14. doi: 10.1016/0022-3956(90)90031-K
- Rhemtulla, M., Brosseau-Liard, P. É., and Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol. Methods* 17, 354–373. doi: 10.1037/a0029315
- Roberts, K. E., Hart, T. A., and Eastwood, J. D. (2016). Factor structure and validity of the state-trait inventory for cognitive and somatic anxiety. *Psychol. Assess.* 28, 134–146. doi: 10.1037/pas0000155
- Saggino, A., Balsamo, M., Carlucci, L., Cavalletti, V., Sergi, M. R., da Fermo, G., et al. (2018). Psychometric properties of the Italian version of the young schema questionnaire I-3: preliminary results. *Front. Psychol.* 9:312. doi: 10.3389/fpsyg.2018.00312
- Saggino, A., Balsamo, M., Carlucci, L., Sergi, M. R., Da Fermo, G., Déttore, D., et al. (2017a). Analysis of the factor structure of the Italian version of the Young Schema Questionnaire L-3 in an Italian clinical and nonclinical sample: preliminary results of a multicenter study. *Giornale Ital. Psicol.* 44, 445–466.
- Saggino, A., Carlucci, L., Sergi, M. R., D'Ambrosio, I., Fairfield, B., Cera, N., et al. (2017b). A validation study of the psychometric properties of the other as shamer scale-2. *SAGE Open* 7:2158244017704241. doi: 10.1177/2158244017704241
- Schalling, D., Cronholm, B., and Asberg, M. (1975). "Components of state and trait anxiety as related to personality and arousal," in *Emotions: Their Parameters and Measurement*, ed. L. Levi (New York, NY: Raven Press), 603–617.
- 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. Online* 8, 23–74.
- Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: beck, CES-D, Hamilton, and Zung. *J. Clin. Psychol.* 62, 123–146. doi: 10.1002/jclp.20213
- Smoller, J. W., and Tsuang, M. T. (1998). Panic and phobic anxiety: defining phenotypes for genetic studies. *Am. J. Psychiatry* 155, 1152–1162. doi: 10.1176/ajp.155.9.1152
- Spielberger, C. D. (1966). "Theory and research on anxiety," in *Anxiety and Behavior*, ed. C. D. Spielberger (New York, NY: Academic Press), 3–19. doi: 10.1016/B978-1-4832-3131-0.50006-8
- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory STAI (form Y) ("Self-Evaluation Questionnaire")*. Available at: <https://ubir.buffalo.edu/xmlui/handle/10477/1873?show=full>
- Spielberger, C. D. (1985). Assessment of state and trait anxiety: conceptual and methodological issues. *Southern Psychol.* 2, 6–16.
- Spielberger, C. D., and Gorsuch, R. L. (1983). *Manual for the State-Trait Anxiety Inventory (form Y): ("Self-Evaluation Questionnaire")*. Sunnyvale, CA: Consulting Psychologists Press, Incorporated.
- Spielberger, C. D., Gorsuch, R. L., and Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Sprinkle, S. D., Lurie, D., Insko, S. L., Atkinson, G., Jones, G. L., Logan, A. R., et al. (2002). Criterion validity, severity cut scores, and test-retest reliability of the Beck Depression Inventory-II in a university counseling center sample. *J. Couns. Psychol.* 49, 381–385. doi: 10.1037/0022-0167.49.3.381
- Stanley, M. A., Beck, J. G., and Zebb, B. J. (1996). Psychometric properties of four anxiety measures in older adults. *Behav. Res. Ther.* 34, 827–838. doi: 10.1016/0005-7967(96)00064-2
- Stephens, A., and Kearsley, N. (1990). Cognitive and somatic anxiety. *Behav. Res. Ther.* 28, 75–81. doi: 10.1016/0005-7967(90)90057-P
- Tellegen, A. (1985). "Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report," in *Anxiety and the Anxiety Disorders*, eds A. H. Tuma and J. D. Maser. (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), 681–706.
- Therrien, Z., and Hunsley, J. (2012). Assessment of anxiety in older adults: a systematic review of commonly used measures. *Aging Ment. Health* 16, 1–16. doi: 10.1080/13607863.2011.602960
- Titov, N., Dear, B. F., McMillan, D., Anderson, T., Zou, J., and Sunderland, M. (2011). Psychometric comparison of the PHQ-9 and BDI-II for measuring response during treatment of depression. *Cogn. Behav. Ther.* 40, 126–136. doi: 10.1080/16506073.2010.550059
- Van Dam, N. T., Gros, D. F., Earleywine, M., and Antony, M. M. (2013). Establishing a trait anxiety threshold that signals likelihood of anxiety disorders. *Anxiety Stress Copin* 26, 70–86. doi: 10.1080/10615806.2011.631525
- Waechter, S., and Stolz, J. A. (2015). Trait anxiety, state anxiety, and attentional bias to threat: assessing the psychometric properties of response time measures. *Cognit. Behav. Ther.* 39, 441–458. doi: 10.1007/s10608-015-9670-z
- Wang, M., Shou, Y., and Wendeng, Y. (2018). Psychometric properties and measurement invariance of the brief symptom inventory-18 among chinese insurance employees. *Front. Psychol.* 9:519. doi: 10.3389/fpsyg.2018.00519
- Watson, D., Clark, L. A., Weber, K., Assenheimer, J. S., Strauss, M. E., and McCormick, R. A. (1995a). Testing a tripartite model: II. Exploring the symptom structure of anxiety and depression in student, adult, and patient samples. *J. Abnorm. Psychol.* 104:15. doi: 10.1037/0021-843X.104.1.15
- Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., and McCormick, R. A. (1995b). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *J. Abnorm. Psychol.* 104, 3–14. doi: 10.1037/0021-843X.104.1.3
- Watson, D., Gamez, W., and Simms, L. J. (2005). Basic dimensions of temperament and their relation to anxiety and depression: a symptom-based perspective. *J. Res. Pers.* 39, 46–66. doi: 10.1016/j.jrp.2004.09.006
- Weems, C. F., Pina, A. A., Costa, N. M., Watts, S. E., Taylor, L. K., and Cannon, M. F. (2007). Predisaster trait anxiety and negative affect predict posttraumatic stress in youths after Hurricane Katrina. *J. Consult. Clin. Psychol.* 75, 154–159. doi: 10.1037/0022-006X.75.1.154
- Wetherell, J. L., and Gatz, M. (2005). The beck anxiety inventory in older adults with generalized anxiety disorder. *J. Psychopathol. Behav.* 27, 17–24. doi: 10.1007/s10862-005-3261-3
- Wittchen, H.-U., and Jacobi, F. (2005). Size and burden of mental disorders in Europe—a critical review and appraisal of 27 studies. *Eur. Neuropsychopharmacol.* 15, 357–376. doi: 10.1016/j.euroneuro.2005.04.012
- Zinbarg, R. E., Yovel, I., Revelle, W., and McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: a comparison of estimators for ω_h . *Appl. Psychol. Measur.* 30, 121–144. doi: 10.1177/0146621605278814
- Zuckerman, M. (1960). The development of an affect adjective check list for the measurement of anxiety. *J. Consult Psychol.* 24, 457–462. doi: 10.1037/h0042713
- Zuckerman, M. (1983). *Biological Bases of Sensation Seeking, Impulsivity, and Anxiety*. Lawrence: Erlbaum Assoc Inc.

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

The MR declared a past co-authorship with several of the authors LC, MS, AS, MB to the handling Editor.

Copyright © 2018 Carlucci, Watkins, Sergi, Cataldi, Saggino and Balsamo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Psychometric Properties and Measurement Invariance of the Brief Symptom Inventory-18 Among Chinese Insurance Employees

Mingshu Li^{1,2}, Meng-Cheng Wang^{1,2,3*}, Yiyun Shou⁴, Chuxian Zhong^{1,2}, Fen Ren⁵, Xintong Zhang^{1,2} and Wendeng Yang^{1,3*}

¹ Department of Psychology, Guangzhou University, Guangzhou, China, ² The Center for Psychometrics and Latent Variable Modeling, Guangzhou University, Guangzhou, China, ³ The Key Laboratory for Juveniles Mental Health and Educational Neuroscience in Guangdong Province, Guangzhou University, Guangzhou, China, ⁴ Research School of Psychology, The Australian National University, Canberra, ACT, Australia, ⁵ School of Education and Psychology, University of Jinan, Jinan, China

OPEN ACCESS

Edited by:

Marco Innamorati,
Università Europea di Roma, Italy

Reviewed by:

Leonardo Carlucci,
Università degli Studi "G. d'Annunzio"
Chieti-Pescara, Italy
Marco Lauriola,
Sapienza Università di Roma, Italy

*Correspondence:

Meng-Cheng Wang
wmcheng2006@126.com
Wendeng Yang
yangwendeng@163.com

Specialty section:

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

Received: 25 January 2018

Accepted: 27 March 2018

Published: 18 April 2018

Citation:

Li M, Wang M-C, Shou Y, Zhong C, Ren F, Zhang X and Yang W (2018) Psychometric Properties and Measurement Invariance of the Brief Symptom Inventory-18 Among Chinese Insurance Employees. *Front. Psychol.* 9:519. doi: 10.3389/fpsyg.2018.00519

This study aimed to examine the psychometric properties and factorial invariance of the Brief Symptom Inventory-18 (BSI-18). Confirmatory factor analyses (CFAs) were performed to verify the BSI-18's factor structure in a large sample of Chinese insurance professionals ($N = 2363$, 62.7% women; age range = 19–70). Multigroup CFA were performed to test the measurement invariance of the model with the best fit across genders. In addition, structural equation modeling was conducted to test the correlations between the BSI-18 and two covariates – social support perception and grit trait. Results indicated that the bi-factor model best fit the data and was also equivalent across genders. The BSI-18's general factor, and somatization and depression dimensions were significantly related to social support perception and grit trait, whereas the anxiety dimension was not. Overall, our findings suggested that the BSI-18's can be a promising tool in assessing general psychological distress in Chinese employees.

Keywords: psychometric properties, Brief Symptom Inventory-18, bi-factor model, measurement invariance, Chinese insurance professionals

INTRODUCTION

The Brief Symptom Inventory-18 (BSI-18; Derogatis, 2001) is an 18-item self-report checklist, a common screening tool for psychological symptoms adapted from the Symptom Checklist-90-Revised (SCL-90-R; Derogatis, 1977) and BSI-53 (Zabora et al., 2001). Previous studies have found that the BSI-18 was highly correlated with its parent instruments—the SCL-90-R and BSI-53. Although the SCL-90-R and BSI-53 have been used extensively in clinical and community samples, both have complicated structural dimensions and large numbers of items. The BSI-18 with 18 items only was developed to more effectively obtain the most critical information about psychiatric symptoms.

The BSI-18's brief items improved test efficiency to some extent; however, the previous findings regarding the factor structure of the BSI-18 were inconsistent. Using a Latina-speaking sample, Prelew et al. (2005) found that a single-factor model resulted from the exploratory factor analysis

(EFA) was the best and most concise model. However, the authors also found that a hypothetical three-factor model fit the data reasonably well when performing confirmatory factor analyses (CFAs) using cross-validation subsamples (Prelow et al., 2005). In the three-factor model, the BSI-18 items were equally distributed to represent the three-factors of depression, anxiety, and somatization (Wiesner et al., 2010). On the other hand, Andreu et al. (2008) found a four-factor structure in a non-clinical sample of 1134 subjects (Andreu et al., 2008). Two of these four-factors (I and II) contained the same items of somatization and depression dimensions. The other two-factors had items of from the initial anxiety factor. One included a group of items assessing distress and widespread nervousness, and another included three items assessing panic symptoms.

In recent years, the bi-factor model has been increasingly popular in mapping the constructs of psychopathological scales, for instance, the Psychopathy Checklist-Revised (Flores-Mendoza et al., 2008) and the Beck Depression Inventory (Al-Turkait and Ohaeri, 2010). The bi-factor measurement structure can be an effective method for modeling multidimensional measurement tools (Reise, 2012). In the process of measuring psychological symptoms, the bi-factor model not only measures the overall situation but also places a secondary load on special symptoms represented by specific dimensions. It has been an increasingly popular view that a bi-factor structure exists between psychiatric symptoms and disorders, where both common and specific components play an important role (Watson, 2005; Thomas, 2012). It has been found that SCL-90-R and BSI-53 had bi-factor model structures (Vassend and Skrandal, 1999; Urbán et al., 2014). The bi-factor model, however, has not been tested for the BSI-18.

Another question surrounding the BSI-18 is whether the instrument has universal applicability among various non-clinical samples, ethnicities, and genders. Preceding studies demonstrated that the BSI-18 is a widely adopted measure with high internal consistency and test-retest reliability in clinical research areas (Wang et al., 2013). However, only one study tested and supported a three-factor model (somatization, depression, and anxiety) of the BSI-18 in Chinese-speaking population using a clinical sample of substance users (Wang et al., 2013). It is unclear whether the BSI-18 self-report version is similarly applicable to non-clinical samples in China.

Previous cross-cultural studies have focused on samples between different ethnicities. To determine the factorial structure and measurement invariance across races/ethnicities, Prelow et al. (2005) emphasized the need for strict invariance testing of the BSI-18 through multigroup CFA. Wiesner et al. (2010) applied a multigroup CFA to evaluate factorial invariance of the BSI-18 in women across multiple ethnicities and the three-factor model only achieved partial metric invariance. From a cross-cultural standpoint, psychological symptoms are sometimes manifested and expressed differently across populations (Wiesner et al., 2010). Each cultural group manifests its specific expression under influences of a typical language format, traditional culture, and educational background. Evaluating the measurement invariance of the BSI-18's self-report version in a Chinese non-clinical sample can be helpful for cross-cultural research.

Many epidemiological investigations have demonstrated that women's incidence of emotional disorders, anxiety disorders, and affective psychosis is higher than men's (Urbán et al., 2014). In a study of insurance employee samples, Dai (2003) found that male insurers were more serious than women in terms of obsessive-compulsive and psychoticism mental health problems. Differences between men and women make it particularly important to verify the measurement model's gender invariance. To date, the BSI-18's measurement invariance across gender is scarcely known, especially among non-Western people.

Mental Health in Chinese Insurance Employees

Insurance industry in mainland China faced with the pressure of external competition and self-development due to the disadvantages of being late starters (Yang, 2013). Huge organization, numerous labor, complex personnel, and high performance requirement along with high working pressure lead to high employee turnover rate and various psychological problems (Yang, 2013). Existing evidence has showed that these psychological problems correlated with social factors and personal traits, especially in insurance staff samples (Dai, 2003). It has been found that psychological distress was influenced by both perceived social support and personality traits among various professional groups (Williams et al., 2002). A lack of social support and sense of belonging have been associated with a person's vulnerability to depression (Williams et al., 2002).

For example, grit is one important trait that may be associated with the success in insurance employees (Ling et al., 2001). Grit trait refers to firm and persistent for a long time with unswerving determination (Duckworth et al., 2007). Whether an insurance employee can tolerate high frustration and work efficiently under pressure all the time will determine his or her success as well as psychological conditions (Ling et al., 2001).

With regards to social factors, previous study shows that the level of social support of insurance employee significantly influenced their psychological well-being (Dai, 2003). For instance, severity of depressive symptoms and frequency of suicidal ideation showed negative significant correlations with low levels of social support (Zhang et al., 2010). These studies predicted, to a certain extent, correlation between specific dimensions (depression and suicide) of psychiatric symptoms and external criteria.

Objective of the Study

This study aimed to examine the BSI-18's psychometric properties in a large sample of Chinese insurance employees. The first goal was to examine the BSI-18's factor structure. CFA were used to examine five hypothetical models: the original single-factor model, three-factor model, four-factor model, and bi-factor model (i.e., the three and four-factor model with one general factor). The second goal was to test measurement invariance of the BSI-18's best-fitting model across genders using the multigroup CFA. Finally, we want to explore the manner in which the general factor (can be considered as an overall mental health status) and dimensions (specific psychiatric symptoms) were related to the social and personality covariates. Specifically,

criterion validity of the BSI-18 bi-factor model was examined using structural equation modeling (SEM).

MATERIALS AND METHODS

Participants

Participants were 2,363 insurance employees from 39 insurance companies in Guangdong Province, China. Their mean age was 35.14 ($SD = 8.985$; age range = 19–70), and 62.7% of the participants were women. Approximately 60.3% of the participants were married, and 65.4% of the participants had attained higher education (see **Table 1** for more information).

Measures

Brief Symptom Inventory-18

The BSI-18 (Derogatis, 2001), a brief self-report version of the 53-item BSI (Derogatis, 1993), was developed to assess general psychological distress in clinical and community populations. The BSI-18 requires participants to evaluate the extent of distress or annoyance they had experienced. Responses were rated on a five-point Likert-type scale, ranging from 1 (not at all) to 5 (very much). The questionnaire's global score summed up all the 18 items. Internal consistency reliability for the present sample was good (Cronbach's alpha = 0.947, 0.867, 0.859, and 0.907 for BSI total, Somatization, Depression, and Anxiety respectively).

Grit-8

The Grit-8 (Duckworth and Quinn, 2009), was an eight-item self-report measure that comprises eight items over two-factors, i.e., consistency of interests and perseverance of effort. These eight

items were rated on a five-point Likert scale, ranging from 1 (not at all like me) to 5 (very much like me). Items 1, 3, 5, and 6 scored negatively; items 2, 4, 7, and 8 scored positively. In the current sample, the Cronbach's alpha was 0.738 for the total scores, indicating good reliability.

Perceived Social Support Scale

The Perceived Social Support Scale (PSSS; Zimet et al., 1990) was a self-report instrument that measures how an individual comprehends various sources of social support, such as family and friends; the total score reflected the total degree of social support that individuals received. The PSSS comprises eight items rated on a seven-point Likert scale, ranging from 1 (not at all true) to 7 (definitely true). For this study, we selected eight items from the family and friend support dimensions. In this study, internal consistency was good for the PSSS total ($\alpha = 0.900$) and two subscales ($\alpha = 0.875$ for family support and $\alpha = 0.870$ for friend support).

Procedures

Participants completed aforementioned self-report questionnaires during their company's morning meeting (administration time was approximately 30 min). The survey was administered by a trained research assistant (RA). The RA provided a general instruction of the survey before the participants started the survey. Participants could ask the RA for clarification if they did not understand any parts of the questionnaire. This study was carried out in accordance with the recommendations of Human Subjects Review Committee at Guangzhou University. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

Data Analysis Strategy

The CFA was performed separately to test five-factor structures, including the single-factor model, three-factor model, four-factor model, and two bi-factor models. Items were treated as categorical variables; thus, robust weighted least squares with mean and variance adjustment (WLSMV) was used in model estimation (Flora and Curran, 2004). Additionally, robust maximum likelihood estimator was employed to obtain the Bayesian information criterion (BIC) value for comparing the non-nested models. Model fits were evaluated using chi-squares, root-mean-square error of approximation (RMSEA), the Tucker-Lewis Index (TLI), the comparative fit index (CFI), and the BIC. Conventional guidelines indicate that an RMSEA value ≤ 0.08 indicates acceptable model fit and a value ≤ 0.05 indicates good model fit. Moreover, CFI and TLI ≥ 0.90 indicate adequate model fit (Kline, 2010). In addition, the ΔBIC value of the two models was greater than 10, indicating that the model with a smaller BIC showed a better model fit (Kuha, 2004).

To further evaluate the bi-factor models, coefficient omega hierarchical (ω_H), the hierarchical omega subscales (ω_{HS}) and the explained common variances (ECVs) were calculated to examine whether the specific factors provide utility beyond the general factor (based on the factor loadings) using the "psych" package (version 1.7.8; Revelle, 2017) in R statistical software (R Core Team, 2017). The proportion of variance in total scores estimated

TABLE 1 | Sample demographic statistics.

	Category	Frequency	Percentage
Gender	Male	865	36.6
	Female	1481	62.7
	Unreported	17	0.7
Nation	Han nationality	2275	96.3
	Other	62	2.6
	Unreported	26	1.1
Age	19–30	820	34.7
	31–40	836	35.4
	41–60	624	26.4
	>60	5	0.2
	Unreported	78	3.3
Education background	\leq High school	800	33.9
	Junior college	962	40.7
	Master degree and above	583	24.7
	Unreported	18	0.8
Marital status	Unmarried	760	32.2
	Married	1425	60.3
	Other	140	5.9
	Unreported	38	1.6
Total		2363	100

by ω_H can be attributed to a single general factor (e.g., Zinbarg et al., 2006), while the reliability of a subscale (or factor) score was reflected by ω_{HS} after controlling for the variance due to the general factor (Reise et al., 2013). When the coefficient ω_H is higher than 0.80, total scores can be regarded as unidimensional because of the most reliable is due to a single common factor (Rodriguez et al., 2016). Meanwhile, the large coefficient ω_H (>0.80) indicates that the vast majority of reliable variance imputing to a specific factor rather than a general factor (Reise et al., 2013).

To test the measurement invariance, the best-fit model resulted from the CFA was initially assessed in both male and female groups separately. Configural invariance can be indicated by that the model fits both genders equally well. Next, metric invariance and scalar invariance were tested by constraining factor loadings and thresholds of the factor models. A DIFFTEST was used to compare improvement in fit between nested models. Notably, the chi-square test was easily affected by sample size so that with increased sample size, even small differences resulted in significant differences. Thus, this research adopted the CFI (ΔCFI) difference numerical model fit indexed to evaluate measurement invariance (Cheung and Rensvold, 2002). According to Cheung and Rensvold (2002), the equivalent model is considered to be acceptable when $\Delta CFI \leq 0.010$ and $\Delta TLI \leq 0.010$.

Finally, the correlations among the factors of the BIS-18 and external criteria variables were examined using a SEM. This study used latent variables to compare observed variables and examined relations among constructs without measurement error (Oh et al., 2004). All models were performed by Mplus 7.4 (Muthén and Muthén, 1998–2010).

RESULTS

Descriptive Statistics

Descriptive statistics and skewness and kurtosis for all key variables were included in **Table 2**. Due to the large values of the skewness and kurtosis, it was necessary to treat the BSI-18 variables as categorical instead on interval. Thus, we used the WLSMV to estimate models.

Factor Structure

Table 3 exhibits fit indices of five competing models for the polychromic correlation matrix of the BSI-18 in the whole sample. As depicted in **Table 3**, all five hypothetical models exhibited good fit to the data (CFIs > 0.90 , TLIs > 0.90). Overall, the bi-factor model provided the best fit to this data among these five alternative models (WLSMV $\chi^2 = 957.934^*$, $df = 117$, CFI = 0.985, TLI = 0.980, RMSEA = 0.055, BIC = 55923.251). In the model, the general factor and three dimensions containing somatization, depression, and anxiety factors were all considered. Because the fit statistics such as CFI were similarly good for the five models (CFI values all greater than or equal to 0.950), the BIC value was used for further verification. The ΔBIC value between the three-factor bi-factor model and four-factor model

TABLE 2 | Descriptive statistics and skewness and kurtosis for all scales included.

	M(SD)	Skewness	Kurtosis
BSI1	1.44 (0.745)	2.084	5.109
BSI2	1.65 (0.912)	1.477	1.875
BSI3	1.72 (0.986)	1.407	1.429
BSI4	1.27 (0.684)	3.064	10.042
BSI5	1.68 (1.038)	1.683	2.217
BSI6	1.88 (1.045)	1.261	1.087
BSI7	1.41 (0.816)	2.322	5.443
BSI8	1.62 (0.981)	1.789	2.745
BSI9	1.49 (0.924)	2.127	4.142
BSI10	1.23 (0.661)	3.487	13.252
BSI11	1.52 (0.914)	2.012	3.806
BSI12	1.37 (0.817)	2.655	7.149
BSI13	1.26 (0.679)	3.219	11.338
BSI14	1.54 (0.943)	1.989	3.610
BSI15	1.47 (0.862)	2.209	4.941
BSI16	1.51 (0.892)	2.041	4.033
BSI17	1.13 (0.575)	5.096	27.325
BSI18	1.34 (0.787)	2.824	8.332
BSI-18 total	26.15 (10.903)	2.290	6.307
GRIT total	29.88 (5.222)	-0.122	-0.280
PSSS total	39.78 (8.629)	-0.443	0.502

BSI-18 total, Brief Symptom Inventory-18 total; GRIT, grit total; PSSS, perceived social support total.

was 189.733, indicating that the smaller value, i.e., the three-factor bi-factor model, shows better fit (see **Table 3**). Due to the items 3 and 6 were not fully representative of anxiety dimension and item 2 of the depression dimension, we tried to re-specify a new model in which items 2, 3, and 6 don't loading on specific factor. Difference testing result indicated a worse model fit [$\Delta \chi^2 = 47.255$, $\Delta df = 3$ ($p < 0.001$)]. Thus, we still used the original three-factor bi-factor model as the best model.

The ω_H for the general factor was 0.87, and the ω_{HS} for somatization factor was 0.28, for anxiety factor was 0.17, and for depression factor was 0.04. In addition, the ECV was 80%. The bi-factor model's standardized factor loadings were presented in **Table 4**.

Measurement Invariance

To ensure that the three-factor bi-factor model provided adequate fit in each group, we first examined it separately for males and females. Results indicated that the bi-factor model fit the two groups well (see **Table 5**). Then the metric invariance model, in which item factor loadings were constrained to be equal, was tested. Results indicated negligible gender differences in model fits ($\Delta CFI \leq 0.01$). Finally, the scalar invariance was tested by further constraining the thresholds to be equal across the two gender groups. The scalar invariance was achieved with a $\Delta CFI = +0.003$.

Criterion Validity

The SEM exhibited mediocre fit to the data (CFI = 0.847, TLI = 0.807). The general BSI factor negatively correlated

TABLE 3 | Goodness-of-fit indices for the five tested models of the BSI-18.

Model	WLSMV χ^2	df	$\Delta\chi^2$	Δdf	TLI	CFI	RMSEA (90% CI)	AIC	BIC	ΔBIC
Single-factor	2678.274*	135	–	–	0.949	0.955	0.089 (0.086, 0.092)	56378.054	56896.955	–
Three-factor	1631.443*	132	1046.831	3	0.969	0.973	0.069 (0.066, 0.072)	55674.591	56210.789	686.166
Four-factor	1538.652*	129	92.791	3	0.970	0.975	0.068 (0.065, 0.071)	55559.49	56112.984	97.805
Three-factor bi-factor	957.934*	117	580.718	12	0.980	0.985	0.055 (0.052, 0.058)	55181.919	55923.251	189.733
Four factor bi-factor	968.816*	117	10.882	0	0.980	0.985	0.056 (0.052, 0.059)	55195.282	55967.684	44.433

WLSMV, weighted least squares with mean and variance adjustment; df, degrees of freedom; TLI, Tucker-Lewis Index; CFI, comparative fit index; $\Delta\chi^2$, change in χ^2 relative to the preceding model; Δdf , change in degrees of freedom relative to the preceding model; ΔCFI , change in comparative fit index relative to the preceding model; ΔTLI , change in Tucker-Lewis Index relative to the preceding model; RMSEA, root-mean-square error of approximation; AIC, Akaike Information Criterion; ΔAIC , change in Akaike information criterion relative to the preceding model; BIC, Bayesian information criterion; ΔBIC , change in Bayesian information criterion relative to the preceding model. * $p < 0.05$. The best fitting model was in bold.

TABLE 4 | The standardized factor loadings for the BSI-18 bi-factor model.

		18 items			
		Somatization	Depression	Anxiety	General
Item 1	Faintness	0.390***			0.633***
Item 4	Pains in chest	0.537***			0.701***
Item 7	Nausea	0.449***			0.672***
Item 10	Trouble getting breath	0.499***			0.767***
Item 13	Numbness	0.427***			0.743***
Item 16	Feeling weak	0.193***			0.797***
Item 2	Feeling no interest in things		0.056*		0.729***
Item 5	Feeling lonely		0.077**		0.801***
Item 8	Feeling blue		0.048*		0.875***
Item 11	Feeling of worthlessness		0.473***		0.751***
Item 14	Feeling hopeless about future		0.468***		0.763***
Item 17	Suicidal thoughts		0.161***		0.812***
Item 3	Nervousness			–0.251***	0.853***
Item 6	Feeling tense			–0.064*	0.825***
Item 15	Feeling restless			0.056**	0.883***
Item 9	Suddenly scared			0.202***	0.888***
Item 12	Spells of panic			0.311***	0.895***
Item 18	Feeling fearful			0.210***	0.872***

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

TABLE 5 | Fit indices for measurement invariance.

Model	WLSMV χ^2	df	$\Delta\chi^2$	Δdf	TLI	CFI	ΔTLI	ΔCFI	RMSEA (90% CI)
Bi-factor boys	447.379*	117	–	–	0.982	0.986	–	–	0.057 (0.052, 0.063)
Bi-factor girls	573.146*	117	–	–	0.982	0.986	–	–	0.051 (0.047, 0.056)
A-Metric invariance	956.841*	266	–	–	0.986	0.988	–	–	0.047 (0.044, 0.050)
B-Scalar invariance	846.772*	320	108.271	+54	0.991	0.991	+0.005	+0.003	0.038 (0.034, 0.041)

WLSMV, weighted least squares with mean and variance adjustment; df, degrees of freedom; TLI, Tucker-Lewis Index; CFI, comparative fit index; RMSEA, root-mean-square error of approximation; $\Delta\chi^2$, change in χ^2 relative to the preceding model; Δdf , change in degrees of freedom relative to the preceding model; ΔCFI , change in comparative fit index relative to the preceding model; ΔTLI , change in Tucker-Lewis Index relative to the preceding model. * $p < 0.05$. Chi-square difference test with WLSMV estimation is different from the conventional chi-square difference test.

with all factors of the Grit, the correlation coefficients ranged from -0.239 to -0.374 ($p < 0.001$; see **Table 6** for details). On the other hand, the somatization factor was positively and significantly related to all factors of GRIT. The depression subscale had negative correlations with Grit total and Grit Effort factor (see **Table 6**). No significant

correlations were found between the anxiety factor and the Grit.

For PSSS, both the general BSI factor and the depression factor showed the strongest negative relation with all PSSS factors, the correlations coefficients ranged from -0.293 to -0.331 ($p < 0.001$). On the other hand, somatization was positively

TABLE 6 | Correlations between Grit, PSSS, and the BSI-18 general and dimension factors.

Factor	GRIT			PSSS		
	GRIT total	Effort	Interest	PSSST	Family	Friend
General BSI	−0.374***	−0.239***	−0.356***	−0.331***	−0.293***	−0.298***
Somatization	0.082***	0.068**	0.057**	0.096***	0.075**	0.089***
Depression	−0.039*	−0.020	−0.044**	−0.031	−0.024	−0.037*
Anxiety	0.011	0.006	0.011	0.003	0.004	0.004

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; GRIT, grit total; PSSST, perceived social support total.

related to all PSSS factors (see **Table 6**). Anxiety factor was not significantly related to PSSS.

DISCUSSION

This study aimed to test the BSI-18's factor structure and measurement invariance in a large sample of Chinese employees. The BSI-18's bi-factor model best fit the present data. The MI tests indicated that the BSI-18 was equivalent for males and females. The results also reveal significant correlations between BSI-18 scores and grit trait and social support.

Although the one-, three-, and four-factor models achieved satisfactory fit, the bi-factor models outperformed the other three models. Moreover, the three-factor bi-factor model better fit the data than four-factor bi-factor model considering the model conciseness, and three-factor bi-factor model was chosen in the follow research. The bi-factor model consists of a general factor (General BSI) that accounted for covariation among all indicators of the comprehensive mental health level and three specific factors (somatization, depression, and anxiety) accounting for variance beyond the general factor in covariation among specific factor indicators (Ward et al., 2015). The current results for the bi-factor model supported the bi-factor structure of psychiatric symptoms, providing general and specific areas of composition. The bi-factor model considers the general mental health status (General BSI) while accounting for the three specific symptoms. Future studies may consider cross-cultural MI tests to clarify the cultural differences in the factor models. From another point of view, the three domain-specific components of the bi-factor model and discriminant validity demonstrated the panic factor was a product of over-extraction (Recklitis et al., 2006); this is in accordance with the result of Derogatis (2001) and concludes that panic may be associated with broader anxiety symptoms.

The current finding also provided evidence for the measurement invariance of the BIS-18 across the male and female samples. The three levels of measurement invariance – configural, metric, and scalar invariance were all achieved in the present study, indicating that the BIS-18 may measure the constructs equally across the two genders.

Finally, we tested the potential covariates that may contribute to the mental conditions measured by the BSI-18. Moderate but significantly negative relations of the general BSI with grit and perceived social support were observed. This finding is consistent with the literature that the grit trait and social support were

important for individuals' mental health conditions (Williams et al., 2002; Dai, 2003). For the three dimensions beyond the general factor, the depression dimension showed modest negative correlation with grit trait and perceived social support. This is in line with the previous finding that suggested low levels of social support may result in higher level of depressive symptoms (Williams et al., 2002) and frequency of suicidal ideation (Zhang et al., 2010).

In contrast, the somatization dimension had positive correlations with the two covariates. The positive correlations of somatization with grit trait and social support may be explained by the cultural influences. Traditional Chinese culture seems to discourage people from expressing their feelings directly; thus, somatization is an alternative way to express emotional disorders (Kleinman, 1982; Cheung, 1995). When there is sufficient social support and grit characteristics, people may fear expressing their psychological distress overtly. This leads to the positive correlations of somatization with social support and grit trait. Another likely reason is that the use of the bi-factor model requires the consideration of the common differences between dimensions (Reise, 2012), and these differences can lead to cross suppression effect (Patrick et al., 2007). The correlations between anxiety dimension and grit trait and social support in this sample were not significant. In other words, all factors are included in the structural equation model, and the direction of the relationship can be reversed when these factors are tested separately.

Clinical Significance

Cultural factors, including ethnic identity and cultural values, influence an individual's idiomatic expression of psychological distress, conceptualization of psychological problems' etiology, and subsequent help-seeking behavior (Torres et al., 2013). The findings in the present study have suggested that Chinese may express psychological problems via somatic symptoms. This is important in for clinical research that aims to measure mental health conditions, and for clinical practice in which how clinicians better assess patients' problems. Clinicians may encourage Chinese patients to isolate the influence of cultural beliefs to be aware of and identify emotional problems, which may also facilitate the patients' help-seeking behaviors. Second, as we observed, both social support and grit are associated with general mental health among the insurance employees. This implied that specific treatment options may be developed

and used for the insurance practitioners. For instance, group career guidance will be affected during the morning session to alleviate employees' occupational stress. Finally, focusing on various special symptoms, the concurrency of various adverse symptoms and the individual's overall psychological state should be simultaneously considered during clinical research and intervention.

Limitations

Some limitations need to be acknowledged. Because expression of psychopathology may be restricted by specific cultural backgrounds, the research's lack of foreign samples as reference groups may be problematic. From this perspective, assessing psychological symptoms of a specific cultural norm to strengthen cross-cultural research is essential. In terms of relationships with external criteria, more other social and behavioral manifestations can be investigated, such as family-to-work conflicts, sources of social pressure, social desirability, days out of work, and so forth. Moreover, the general factor of bi-factor model maybe represent a statistical artifact factor, only in theory, but the factor loadings on general factor were large enough in our data, so it is couldn't be caused by artifact effect or method effects.

In sum, this study suggested that the BSI-18 is a reliable and valid general psychological distress measurement instrument that can be extended to Chinese insurance employees. The bi-factor model better represented the BSI-18's underlying structure. Meanwhile, Chinese men and women shared a common understanding of psychological distress as measured by the BSI-18. Furthermore, this study highlighted the importance of assessing the general factor and viewed the mental health of insurance practitioners as a holistic approach rather than focusing on individual dimensions while excluding the artifact effect or method effects.

REFERENCES

- Al-Turkait, F. A., and Ohaeri, J. U. (2010). Dimensional and hierarchical models of depression using the beck depression inventory-ii in an Arab college student sample. *BMC Psychiatry* 10:60. doi: 10.1186/1471-244X-10-60
- Andreu, Y., Galdón, M. J., Dura, E., Ferrando, M., Murgui, S., and García, A. (2008). Psychometric properties of the Brief Symptoms Inventory-18 (BSI-18) in a Spanish sample of outpatients with psychiatric disorders. *Psicothema* 20, 844–850.
- Cheung, F. M. (1995). "Facts and myths about somatization among the Chinese," in *Chinese Societies and Mental Health*, eds T.-Y. Lin, W. S. Tseng, and E. K. Yeh (New York, NY: Oxford University Press), 156–180.
- Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Modeling* 9, 233–255. doi: 10.1097/NNR.0b013e3182544750
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna: R Core Team.
- Dai, F. (2003). *Research on the Mental Health Status and Related Factors of Insurance Salesmen*. M.A. thesis, Jiangxi Normal University, Jiangxi.
- Derogatis, L. (1993). *Brief Symptom Inventory (BSI): Administration, Scoring, and Procedures Manual*, 3rd Edn. Minneapolis, MN: NCS Pearson.
- Derogatis, L. (2001). *Brief Symptoms Inventory 18: Administration, Scoring, and Procedures Manual*. Minneapolis, MN: NCS Pearson.

AUTHOR CONTRIBUTIONS

ML and CZ made substantial contribution to the analysis and to the interpretation of the data, drafted the manuscript, provided final approval for the version to be published, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. M-CW made substantial contributions to the conception and the design of the study, drafted the manuscript, provided final approval for the version to be published, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. FR, XZ, YS, and WY helped out in the interpretation of data for the work, revised the manuscript critically for important intellectual content, provided final approval for the version to be published, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This study was funded by the National Natural Science Foundation of China (Grant No. 31400904), Xinmiao Project of Guangzhou University (2014-27), Foundation of Guangzhou Shishu Gaoxiao Project (Grant No. 1201431330), and Guangzhou University's 2017 training program for young top-notch personnels (BJ201715).

ACKNOWLEDGMENTS

We would like to thank Guangdong Insurance Intermediary Association for their assistance in collecting data.

- Derogatis, L. R. (1977). *The SCL-90 Manual: Scoring, Administration and Procedures for the SCL-90*. Baltimore, MD: John Hopkins University School of Medicine.
- Duckworth, A. L., Peterson, C., Matthews, M. D., and Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *J. Pers. Soc. Psychol.* 92, 1087–1101. doi: 10.1037/0022-3514.92.6.1087
- Duckworth, A. L., and Quinn, P. D. (2009). Development and validation of the short grit scale (grit-s). *J. Pers. Assess.* 91, 166–174. doi: 10.1080/00223890802634290
- 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
- Flores-Mendoza, C. E., Herrero, O., and Abad, F. J. (2008). Factor structure and behavioural correlates of the psychopathy checklist-revised [PCL-R] in a Brazilian prisoner sample. *Pers. Individ. Dif.* 45, 584–590. doi: 10.1016/j.paid.2008.06.016
- Kleinman, A. (1982). Neurasthenia and depression: a study of somatization and culture in china. *Cult. Med. Psychiatry* 6, 117–190. doi: 10.1007/BF0051427
- Kline, R. B. (2010). *Principles and Practice of Structural Equation Modeling*, 3rd Edn. New York, NY: Guilford Press.
- Kuha, J. (2004). AIC and BIC: comparisons of Assumptions and Performance. *Sociol. Methods Res.* 33, 188–229. doi: 10.1177/0049124103262065

- Ling, W. Q., Zhang, D. K., and Fang, L. L. (2001). The analysis of the structure for self-efficacy of insurance salesmen. *Acta Psychol. Sin.* 33, 63–67.
- Muthén, L. K., and Muthén, B. O. (1998–2010). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.
- Oh, H.-J., Glutting, J. J., Watkins, M. W., Youngstrom, E. A., and McDermott, P. A. (2004). Correct interpretation of latent versus observed abilities: implications from structural equation modeling applied to the WISC-III and WIAT linking sample. *J. Spec. Educ.* 38, 159–173. doi: 10.1177/00224669040380030301
- Patrick, C. J., Hicks, B. M., Nichol, P. E., and Krueger, R. F. (2007). A bifactor approach to modeling the structure of the psychopathy check-list-revised. *J. Pers. Disord.* 21, 118–141. doi: 10.1521/pedi.2007.21.2.118
- Prelow, H. M., Weaver, S. R., Swenson, R. R., and Bowman, M. A. (2005). A preliminary investigation of the validity and reliability of the brief-symptom inventory-18 in economically disadvantaged Latina American mothers. *J. Commun. Psychol.* 33, 139–155. doi: 10.1002/jcop.20041
- Recklitis, C. J., Parsons, S. K., Shih, M. C., Mertens, A., Robison, L. L., and Zeltzer, L. (2006). Factor structure of the brief symptom inventory-18 in adult survivors of childhood cancer: results from the childhood cancer survivor study. *Psychol. Assess.* 18, 22–32. doi: 10.1037/1040-3590.18.1.22
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behav. Res.* 47, 667–696. doi: 10.1080/00273171.2012.715555
- Reise, S. P., Bonifay, W. E., and Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *J. Pers. Assess.* 95, 129–140. doi: 10.1080/00223891.2012.725437
- Revelle, W. (2017). *Psych: Procedures for Personality and Psychological Research*. Evanston, IL: Northwestern University.
- Rodriguez, A., Reise, S. P., and Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *J. Pers. Assess.* 98, 223–237. doi: 10.1080/00223891.2015.1089249
- Thomas, M. L. (2012). Rewards of bridging the divide between measurement and clinical theory: demonstration of a bifactor model for the brief symptom inventory. *Psychol. Assess.* 24, 101–113. doi: 10.1037/a0024712
- Torres, L., Miller, M. J., and Moore, K. M. (2013). Factorial invariance of the brief symptom inventory-18 (bsi-18) for adults of Mexican descent across nativity status, language format, and gender. *Psychol. Assess.* 25, 300–305. doi: 10.1037/a0030436
- Urbán, R., Kun, B., Farkas, J., Paksi, B., Kökönyei, G., and Unoka, Z. (2014). Bifactor structural model of symptom checklists: scl-90-r and brief symptom inventory (BSI) in a non-clinical community sample. *Psychiatry Res.* 216, 146–154. doi: 10.1016/j.psychres.2014.01.027
- Vassend, O., and Skrondal, A. (1999). The problem of structural indeterminacy in multidimensional symptom report instruments the case of SCL-90-R. *Behav. Res. Ther.* 37, 685–701. doi: 10.1016/S0005-7967(98)00182-X
- Wang, J., Kelly, B. C., Liu, T., Zhang, G., and Hao, W. (2013). Factorial structure of the brief symptom inventory (BSI)-18 among Chinese drug users. *Drug Alcohol Depend.* 133, 368–375. doi: 10.1016/j.drugalcdep.2013.06.017
- Ward, J. T., Nobles, M. R., and Fox, K. A. (2015). Disentangling self-control from its elements: a bifactor analysis. *J. Quant. Criminol.* 31, 595–627. doi: 10.1007/s10940-014-9241-6
- Watson, D. (2005). Rethinking the mood and anxiety disorders: a quantitative hierarchical model for DSM-V. *J. Abnorm. Psychol.* 114, 522–536. doi: 10.1037/0021-843X.114.4.522
- Wiesner, M., Chen, V., Windle, M., Elliott, M. N., Grunbaum, J. A., and Kanouse, D. E. (2010). Factor structure and psychometric properties of the brief symptom inventory-18 in women: a macs approach to testing for invariance across racial/ethnic groups. *Psychol. Assess.* 22, 912–922. doi: 10.1037/a0020704
- Williams, R. A., Hagerty, B. M., Yousha, S. M., Hoyle, K. S., and Oe, H. (2002). Factors associated with depression in navy recruits. *J. Clin. Psychol.* 58, 323–337. doi: 10.1002/jclp.1146
- Yang, X. Y. (2013). *Research on Mental Health to Staff for Insurance Marketing in T Branch of HX Life-Insurance Company*. M.A. thesis, University Of International Business and Economics, Beijing.
- Zabora, J., Brintzenhofesoc, K., Jacobsen, P., Curbow, B., Piantadosi, S., and Hooker, C. (2001). A new psychosocial screening instrument for use with cancer patients. *Psychosomatics* 42, 241–246. doi: 10.1176/appi.psy.42.3.241
- Zhang, J., Grabiner, V. E., Zhou, Y., and Li, N. (2010). Suicidal ideation and its correlates in prisoners: a comparative study in china. *Crisis* 31, 335–342. doi: 10.1027/0227-5910/a000055
- Zimet, G. D., Powell, S. S., Farley, G. K., Werkman, S., and Berkoff, K. A. (1990). Psychometric characteristics of the multidimensional scale of perceived social support. *J. Pers. Assess.* 55, 610–617. doi: 10.1080/00223891.1990.9674095
- Zinbarg, R., Yovel, I., Revelle, W., and McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: a comparison of estimators for h. *Appl. Psychol. Meas.* 30, 121–144. doi: 10.1177/0146621605278814

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

Copyright © 2018 Li, Wang, Shou, Zhong, Ren, Zhang and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Is Parent–Child Disagreement on Child Anxiety Explained by Differences in Measurement Properties? An Examination of Measurement Invariance Across Informants and Time

Thomas M. Olin^{1*}, Megan Finsaas², Lea R. Dougherty³ and Daniel N. Klein²

OPEN ACCESS

Edited by:

Marco Innamorati,
Università Europea di Roma, Italy

Reviewed by:

Marco Tommasi,
Università degli Studi G. d'Annunzio
Chieti e Pescara, Italy
Daiana Colledani,
Università degli Studi di Padova, Italy
Jesús M. Alvarado,
Complutense University of Madrid,
Spain

Daniel Ondé,
Complutense University of Madrid,
Spain, in collaboration with
reviewer JA.

*Correspondence:

Thomas M. Olin
thomas.olino@temple.edu;
thomas.olino@gmail.com

Specialty section:

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

Received: 23 March 2018

Accepted: 05 July 2018

Published: 31 July 2018

Citation:

Olin TM, Finsaas M, Dougherty LR
and Klein DN (2018) Is Parent–Child
Disagreement on Child Anxiety
Explained by Differences
in Measurement Properties? An
Examination of Measurement
Invariance Across Informants
and Time. *Front. Psychol.* 9:1295.
doi: 10.3389/fpsyg.2018.01295

¹ Department of Psychology, Temple University, Philadelphia, PA, United States, ² Department of Psychology, Stony Brook University, Stony Brook, NY, United States, ³ Department of Psychology, University of Maryland College Park, College Park, MD, United States

There are numerous empirical studies demonstrating that agreement between parent-reports of youth and youth self-reports of internalizing behavior problems is modest at best. This has spurred much research on factors that influence the magnitude of associations between informants, including individual difference characteristics of the informants and contexts through which individuals interact with the child. There is also tremendous interest in understanding symptom trajectories longitudinally. However, each of these lines of work are predicated on the assumptions that the psychometric construct that is being assessed from each informant and at each measurement occasion is the same. This study examined measurement invariance between maternal and child reports and longitudinally across ages 9 and 12 on five dimensions of anxiety using the Screen for Child Anxiety and Related Disorders (SCARED; Birmaher et al., 1999). No cross-informant models for anxiety dimensions achieved acceptable fit and at least partial metric and scalar invariance. Moreover, few longitudinal models demonstrated acceptable fit and at least partial metric and scalar invariance. Thus, using the SCARED as an example, these results show that inter-informant agreement may be compromised by different item functioning, and highlight the need for testing invariance before using measures for longitudinal tracking of symptoms.

Keywords: measurement invariance, anxiety, development, parent–child agreement, assessment

INTRODUCTION

There has been extensive research on agreement and disagreement between raters of symptoms of behavior problems in children and adolescents. These studies have examined multiple constellations of raters, including parents of the same target child, a parental caregiver and teachers, and parents and their child. Overall, there is modest agreement between parents and children and parents and teachers, but moderate agreement between parents (De Los Reyes et al., 2015). Attempts to understand factors that influence agreement between raters and also within raters over time have not provided complete explanations for lack of agreement. However, there have

been no studies that test whether the underlying constructs reported by different informants, particularly primary caregivers and their children, are equivalent. There are few studies examining parallel issues over time. Without such evidence, it is difficult to interpret associations across informants as reflecting agreement on the same construct and how to evaluate longitudinal changes in the constructs. Thus, the present study examines whether measurement differences are present between parent- and child self-reports of anxiety that may partially explain lack of agreement across raters and across development.

The overall pattern of inter-informant agreement on child mental health symptoms have been extensively examined and summarized in two meta-analyses spanning a 28-year period. In the first, Achenbach et al. (1987) examined the associations between youth, parent, and teacher reports of internalizing and externalizing problems. In their work, there was stronger agreement among individuals with the same relationship to the target child (e.g., inter-parental agreement, average $r = 0.61$ across informant types), but more modest associations across different informant types (average $r = 0.29$ across all informants). Inter-informant agreement for overcontrolled and undercontrolled behavior problems, similar to internalizing and externalizing problems, respectively, were in the small-moderate range ($r_s = 0.32$ and 0.41 , respectively). More recently, De Los Reyes et al. (2015) conducted an updated analysis of studies since the Achenbach et al. (1987) paper. In this work, the authors found that the magnitude of interparental agreement (mean $r = 0.59$) was similar to that of other informant pairs with the same relationship to the target (i.e., teachers, mental health workers; average $r = 0.58$). However, agreement between raters with different relationships to the target was markedly lower (average $r = 0.29$). Overall inter-informant agreement was modest for both internalizing ($r = 0.25$) and externalizing problems ($r = 0.30$). The convergent findings from the two meta-analyses indicate that individuals with greater similarity in information will have a higher degree of similarity in their ratings of behavior. This has served as the foundation for the Operations Triad Model (De Los Reyes et al., 2013, 2015), which emphasizes context as an important factor in understanding reports of child behavior problems and assessing the incremental value of information from disparate sources.

Numerous studies have examined factors that explain the modest levels of convergence between informants on youth internalizing and externalizing behavior problems. These studies have considered moderating factors such as parent-child relationship functioning (Treutler and Epkins, 2003), parent symptoms (Youngstrom et al., 2000; Treutler and Epkins, 2003; Rothen et al., 2009), parental stress (Youngstrom et al., 2000; Langberg et al., 2010), child race (Youngstrom et al., 2000), child sex (Rothen et al., 2009), and characteristics of the symptoms themselves (e.g., observability, salience; Frank et al., 2000; Karver, 2006). However, these findings lack coherence and are sparsely replicated across samples.

There have been numerous studies examining the developmental course of anxiety disorders and symptoms with studies focusing on different age spans (Feng et al., 2008; Van Oort et al., 2009; Olino et al., 2010b, 2014). These

studies have focused on risk factors predicting course as well as course predicting outcomes. However, there has been a paucity of attention to longitudinal MI for youth anxiety. This precludes understanding whether observed mean-level changes are reflecting true score changes, or if these changes are influenced by changes in measurement properties. In one study (Mathyssek et al., 2013), the authors found evidence supporting MI for individual dimensions of anxiety from the Revised Child Anxiety and Depression Scale (RCADS; Chorpita et al., 2000). However, this study examined this issue using only youth-reports for a single assessment measure. Thus, comparisons between youth and parent reports across time are novel.

A key challenge in examining inter-informant agreement and assessing stability over time concerns the psychometric functioning of the measures used to assess the constructs. De Los Reyes et al. (2015) identified several sources of measurement error that may lead to attenuation of associations. Some of these are factors such as parental psychopathology or personality that may lead to distorted reports of youth behavior (Kagan, 1997; Najman et al., 2001; Hayden et al., 2010). Random error, such as imperfect test-retest reliability, could also limit the magnitude of associations across raters. Finally, the authors identify systematic error across informants as a potential explanation for the limited inter-informant associations.

Systematic error in ratings can come from several sources. De Los Reyes et al. (2015) focus on studies demonstrating differences in item response scaling as a possible, but unlikely, contributor to low inter-informant agreement. However, there are additional considerations that have not yet been explored in this area. For example, systematic error may be introduced because the constructs that individual informants are reporting on have different psychometric properties. Estimation of reliability is frequently indexed by Cronbach's alpha (Cronbach, 1951). However, alpha is more correctly interpreted as a measure of internal consistency (Sijtsma, 2009). It does not provide information about the specific measurement structure of the items comprising a test/scale.

To evaluate this possibility, more sophisticated analytic tools are necessary. For example, confirmatory factor analysis (CFA) can evaluate measurement properties such as how items relate to constructs. Extensions of CFA have been developed to test whether measurement properties of constructs are consistent across informants (Olino and Klein, 2015) and assessment waves (Widaman et al., 2010). These methods have been termed measurement invariance (MI; Meredith, 1993).

There are multiple levels of MI that reflect increasingly strict model properties, and address different psychometric questions (Widaman et al., 2010; Millsap, 2011). A fundamental requirement is that the same items are associated with the same construct across units (e.g., informants and time). Simply stated, do the same items load on the same factors when assessed in the different units. This is referred to as configural invariance. If the items assessing what are purportedly the same constructs differ across groups, the items have different meanings within each group. Next, it is important that the magnitude of the associations between the items and the underlying construct is the same across groups (i.e., are the factor loadings for each factor comparable

when assessed within the different groups?). This is referred to as metric invariance. Finally, the probability of item endorsement should be the same across groups (Reise et al., 1993; Vandenberg and Lance, 2000). This is referred to as scalar invariance. When configural, metric, and scalar invariance are established for a particular measure across groups, scale scores can be considered to reflect the same psychometric quantities among the groups. Thus, it is critical to evaluate whether lack of MI is contributing to reduced associations between parents and children. However, complete MI imposes highly rigorous assumptions (i.e., equality of all factor loadings and item thresholds across informants). Consequently, there has been increasing attention to the presence of partial MI that specifies invariance on parameters for some, but not all, items (Byrne et al., 1989). This approach has gained prominence and has permitted meaningful comparisons when full MI fails (Steinmetz, 2013).

In the present study, we examine MI across maternal- and child-reports of youth anxiety symptoms when children are ages 9 and 12. Thus, we are able to describe differences in MI across this 3-year developmental span. We also present analyses examining MI across time for maternal- and child-reports separately.

In light of the consistently modest agreement between maternal and child reports of symptomatology, we expect to find a lack of MI across informants at both assessment waves. We do not posit whether this is due to differences in factor loadings or thresholds. However, we expect there to be stronger support for MI across time within informants as there is evidence for longitudinal stability of youth anxiety (Prenoveau et al., 2011). In instances when full MI fails, we examine partial MI that permits some flexibility in the models.

MATERIALS AND METHODS

Participants and Procedure

Participants were from a larger sample of 559 children and their families living in a suburban community who were participating in the Stony Brook Temperament Study, a longitudinal study of temperament and psychopathology, which began when children were 3 years old (Olino et al., 2010a). Potential participants were identified using a commercial mailing list and screened by telephone. Families with a 3-year-old child who lived with an English-speaking biological parent within 20 contiguous miles of Stony Brook, New York and did not have significant medical conditions or developmental disabilities were included. Of the 815 identified eligible families, 68.5% entered the study. No significant differences were found between families who did and did not participate on child sex and race/ethnicity, and parental marital status and education. Informed and written consent was obtained from the parent prior to participation. The study was approved by the institutional review board at Stony Brook University, and families were compensated for their participation. At the second wave of the study, 3 years later, 50 additional minority families were recruited to increase racial/ethnic diversity (total $N = 609$; Bufferd et al., 2012).

At the age 9 visit, 487 mothers (80.0%) and 481 youth (79.0%) completed the measures of youth anxiety symptoms used in this

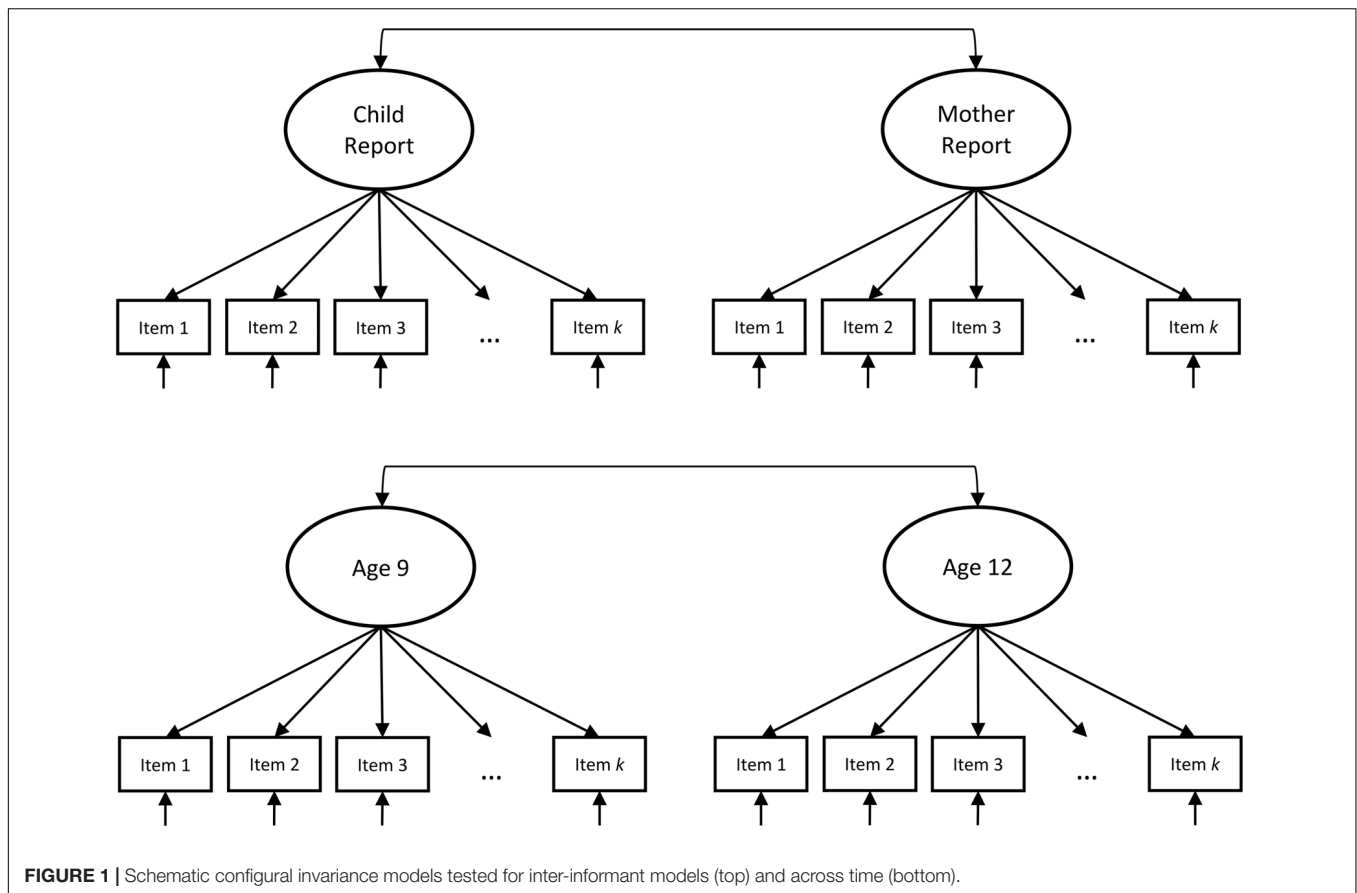
study; a mother or child from 492 families (80.8%) participated. At the age 12 visit, 468 mothers (76.8%) and 470 youth (77.2%) completed these measures; a mother or child from 479 families (78.7%) participated. The mean age of the children was 9.18 years ($SD = 0.40$) at the 9-year assessment and 12.66 ($SD = 0.46$) at the 12-year assessment. Approximately half the children were female (9-year visit: 226, 45.9%; 12-year visit: 225, 47.0%) and the majority were White/non-Hispanic (9-year visit: 390, 79.3%; 12-year visit: 381, 79.5%). At the time of the 12-year visit, most mothers were married (373, 77.9%) and approximately half had graduated from college (279; 58.2%), and the median income bracket was \$100,000–\$119,999. Youth who participated at age 9 did not differ from those participating at age 3 on child sex, race, or total or externalizing behavior problems, as assessed by maternal reports on the Child Behavior Checklist (Achenbach and Rescorla, 2001; all $ps > 0.05$). However, youth who did not continue with the study at age 9 had higher levels of internalizing problems at age 3 than those who continued with the study, though the effect is small [$t(547) = 4.69, P < 0.05, d = 0.09$].

Measures

Children and their parents completed the 41-item youth self-report and parent-report versions, respectively, of the Screen for Childhood Anxiety Related Disorders (SCARED; Birmaher et al., 1997, 1999). Children and their parents are asked to rate the presence of anxiety symptoms in the child over the past 3 months on a three-point scale (0 = not true or hardly ever true; 1 = somewhat true or sometimes true; 2 = very true or often true). The SCARED is made up of five factor-analytically derived subscales: panic/somatic, general anxiety, separation anxiety, social phobia, and school phobia. These subscales reflect anxiety disorder symptoms as conceptualized in the DSM-IV-TR. Each factor has been shown to have good internal consistency and test-retest reliability (range of α : 0.78–0.87; Birmaher et al., 1999; intraclass correlation across time for each scale ranged from 0.70–0.80; Birmaher et al., 1997).

Statistical Analyses

In line with a model building approach and to identify whether one-factor models were appropriate for testing, we estimated a series of initial single-factor CFAs separately for youth self- and parent-reports at the ages 9 and 12 waves. Items from the panic/somatic, general anxiety, and social phobia subscales were included in models reflecting each of these constructs, respectively. Next, models were fit sequentially to evaluate MI and we continued testing for MI only when there was evidence that a one-factor model for each was an acceptable fit to the data. We followed the same logical progression of testing MI across informants as is used in examinations of longitudinal invariance (Widaman et al., 2010) with minor modifications. We tested first for configural invariance (schematic models for configural invariance models are displayed in **Figure 1**), or whether the pattern of significant (i.e., non-zero) factor loadings is similar across youth and parent-reports. We estimated models for each of the subscales including a single factor for youth and a single factor maternal-reports simultaneously while permitting the factors to be correlated. These models were specified freely



estimating all factor loadings and fixing the latent variable variance at 1 for purposes of model identification. Next, we tested for metric invariance, or whether factor loadings for each item are equal across informants. In these models, we freely estimated the variance of the maternal-report latent factor as fixing factor loadings to be equal across informants permits this constraint to be relaxed for one informant. Finally, we tested for scalar invariance, or whether the probability of item endorsement is similar across informants, by constraining the thresholds across informants to be equal. In these models, we freely estimated the mean of the maternal-report latent factor as fixing thresholds to be equal across informants permits this constraint to be relaxed for one informant. If all three types of invariance hold, this indicates that the scales measure the same constructs across reporters on the same scale. Thus, differences in mean trait levels can be interpreted as true score differences, as opposed to differences in measurement.

For models that did not achieve full MI, we tested partial MI, which identifies whether some, but not all, items are invariant across informants and/or time. We examined the presence of comparable factor loadings using the MODEL CONSTRAINT command in Mplus to assess differences in configural invariance. When factor loadings were identified that did not significantly differ at $P < 0.05$, a partial metric invariant model was estimated that included equality constraints on those factor loadings. In

this partial metric invariance model, we used the MODEL CONSTRAINT command that tests whether the difference between specified parameters significantly differ, to examine the presence of comparable item thresholds. When item thresholds were identified that did not significantly differ at $P < 0.05$, a partial scalar invariant model was estimated that included equality constraints on those item thresholds.

All models were estimated in Mplus version 8 (Muthén and Muthén, 1998–2017) using the weighted least squares estimator (WLSMV; Flora and Curran, 2004), which is a robust estimator suited for modeling binary data. There were low rates of responses in the highest response category (i.e., “very true or often true”) on many items. Specifically, for 34 (82.9%) items at both ages 9 and 12, 5% or fewer of parents endorsed the highest category. Similarly, for 7 (17.1%) items at age 9, and 23 items (56.1%) at age 12, 5% or fewer of children endorsed the most severe response option. Consequently, the top two item response categories were collapsed, making all items binary. We evaluated models on two goodness of fit indices. Specifically, we used the comparative fit index (CFI; Bentler, 1990) and Root Mean Square Error of Approximation (RMSEA; Steiger, 1990). Although cut-offs are somewhat arbitrary (Marsh et al., 2004), current conventions suggest that excellent model fit is indicated by CFI values ≥ 0.95 (Hu and Bentler, 1999) and RMSEA values ≤ 0.05 (MacCallum et al., 2006); good fit is indicated by CFI greater than 0.90 and a RMSEA between 0.05 and 0.10.

We estimated configural (similar pattern of factor loadings across groups), metric (equality of factor loadings across groups), and scalar (equality of thresholds across groups) for comparisons between maternal- and child-reports. In addition to testing MI across informants, we also tested the same sequence of models for evaluating longitudinal MI in each informant, separately. Model fit comparisons were evaluated by investigating change in both CFI and RMSEA using Chen's (2007) guidelines. Chen (2007) recommended interpreting reductions in CFI of 0.01 and RMSEA of 0.015 as indicating non-invariance (i.e., failure to demonstrate MI). When the RMSEA and CFI changes led to different conclusions, we relied on the more conservative index to inform interpretations.

RESULTS

Measurement Models for Informant and Age

Initial models estimated one-factor models for each of the SCARED subscales for child self- and maternal-reports at ages 9 and 12. These models were estimated to identify scales that fit the data well enough to pursue tests of MI. **Table 1** displays overall fit for each of the models tested. For age 9 data, one-factor models demonstrated excellent fit for child-reported generalized anxiety disorder (GAD), panic, and social phobia and demonstrated a good fit for maternal-reported GAD, panic, and separation anxiety. For age 12 data, one-factor models demonstrated excellent fit for child-reported panic and good fit for GAD and social phobia, and demonstrated excellent fit for maternal-reported panic and good fit for GAD, separation anxiety, and social phobia. One-factor models for child-reported separation anxiety were poor fits to the data at each time point. Model fit for school avoidance was also less than adequate. For child reports at age 12 and mother reports at age 9, the CFI was acceptable, but the RMSEA was greater than 0.10. In addition, the model for maternal-report of school avoidance at age 12 failed to provide an admissible solution. Owing to the brevity of the school phobia scale, the school avoidance models included only four observed indicators, which may have led to model instability.

As child-report separation anxiety provided poor fit to the data at ages 9 and 12, we did not assess MI for the youth reports on this subscale. However, as maternal reports of separation anxiety demonstrated good fit, we examined longitudinal invariance for mothers' reports on this subscale. Due to the problematic fit of the school avoidance models, we did not conduct any MI analyses on this subscale. All model parameters are available in the Supplementary Materials.

Tests of MI: Child- and Maternal-Reports at Age 9

The configural invariance model for GAD across youth self- and maternal-reports was a good fit to the data (**Table 2**). Likewise, the fit of the metric invariance model was good, and imposing constraints on the factor loadings did not markedly diminish model fit. However, when imposing

constraints on the item thresholds across informants, model fit diminished substantially. Comparisons identified three item thresholds that did not significantly differ across informants. Estimating a partial scalar invariant model that constrained those three item thresholds to equality yielded good model fit. Thus, this model supports partial scalar MI.

The configural invariance models for panic disorder across youth self- and maternal-reports were a poor fit to the data. Thus, further tests of metric and scalar invariance were not pursued.

The fit for the configural invariance model for social phobia across youth self- and maternal-reports was good. Likewise, the metric invariance model was a good fit to the data, and imposing constraints on the factor loadings did not markedly diminish model fit. Similarly, imposing constraints on the item thresholds across informants did not substantially diminish model fit, supporting full-scalar MI.

Tests of MI: Child- and Maternal-Reports at Age 12

The configural invariance model for GAD across youth self- and maternal-reports at age 12 was a good fit to the data (**Table 3**). Likewise, the fit of the metric invariance model was good, and imposing constraints on the factor loadings did not markedly diminish model fit. However, when imposing constraints on the item thresholds across informants, model fit diminished substantially, failing to support scalar invariance. Comparisons identified only one item threshold that did not significantly differ across informants. Thus, this model also failed to support partial scalar MI.

The configural invariance model for panic disorder demonstrated adequate fit. Including constraints on factor loadings across informants to test metric invariance yielded a model with an adequate fit to the data and did not markedly differ from the configural invariance model. However, when including constraints on item thresholds to test for scalar invariance, model fit was poor and was reduced relative to the metric invariance model. Moreover, all item thresholds significantly differed across informants, hence there was no basis for evaluating partial scalar invariance.

The configural invariance model for social phobia across youth self- and maternal-reports was a good fit to the data. Likewise, the fit of the metric invariance model was good, and imposing constraints on the factor loadings did not markedly diminish model fit. Finally, after imposing constraints on the item thresholds across informants, model fit was not substantially diminished. Thus, this model supports full-scalar MI.

Tests of MI: Child-Reports Across Ages 9 and 12

The fit for the configural invariance model for GAD for youth self-reports across ages 9 and 12 was excellent (**Table 4**). Likewise, the metric invariance model was an excellent fit to the data as imposing constraints on the factor loadings did not markedly diminish model fit. When imposing constraints on the item thresholds across informants to test scalar invariance, overall

TABLE 1 | Initial model fit for child self- and maternal-report of SCARED subscales at ages 9 and 12.

	Age 9					Age 12				
	Child self-report									
	χ^2	<i>df</i>	<i>P</i>	CFI	RMSEA	χ^2	<i>df</i>	<i>P</i>	CFI	RMSEA
GAD	37.93	27	0.08	0.987	0.029 (0–0.049)	98.09	27	0.00	0.941	0.075 (0.059–0.091)
Panic	64.05	65	0.51	1.000	0.000 (0.000–0.026)	79.32	65	0.11	0.982	0.022 (0.000–0.037)
School	11.70	2	0.00	0.941	0.100 (0.05–0.159)	35.51	2	0.00	0.914	0.189 (0.137–0.246)
Separation anxiety	134.76	20	0.00	0.868	0.109 (0.092–0.127)	121.30	20	0.00	0.827	0.104 (0.086–0.122)
Social anxiety	28.31	14	0.01	0.978	0.046 (0.021–0.071)	51.44	14	0.00	0.961	0.075 (0.054–0.098)
	Maternal-report									
GAD	138.02	27	0.00	0.943	0.092 (0.077–0.107)	73.55	27	0.00	0.974	0.061 (0.044–0.078)
Panic	161.52	65	0.00	0.916	0.055 (0.045–0.066)	93.32	65	0.01	0.961	0.031 (0.015–0.044)
School	23.69	2	0.00	0.944	0.149 (0.099–0.206)	NA	NA	NA	NA	NA
Separation anxiety	109.92	20	0.00	0.910	0.096 (0.079–0.114)	58.93	20	0.00	0.911	0.064 (0.046–0.084)
Social anxiety	84.18	14	0.00	0.985	0.101 (0.081–0.123)	43.47	14	0.00	0.987	0.067 (0.045–0.09)

GAD, generalized anxiety disorder symptoms; panic, panic disorder symptoms; school, school phobia symptoms; separation anxiety, separation anxiety disorder symptoms; and social anxiety, social anxiety symptoms.

TABLE 2 | Tests of MI between child self- and maternal-reports at age 9.

	χ^2	<i>df</i>	<i>P</i>	CFI	RMSEA	Δ CFI	Δ RMSEA
GAD							
Configural	252.433	134	<0.001	0.951	0.042 (0.034–0.050)		
Full metric	287.766	142	<0.001	0.940	0.046 (0.038–0.053)	–0.011	0.004
Full scalar	411.928	150	<0.001	0.892	0.060 (0.053–0.066)	–0.048	0.014
Partial scalar ^a	292.481	144	<0.001	0.939	0.046 (0.038–0.053)	–0.001	0.000
Panic							
Configural	526.00	298	<0.001	0.864	0.039 (0.034–0.045)		
Social							
Configural	157.74	76	<0.001	0.984	0.047 (0.036–0.057)		
Full metric	168.81	82	<0.001	0.983	0.046 (0.036–0.056)	–0.001	–0.001
Full scalar	191.03	88	<0.001	0.980	0.049 (0.039–0.058)	–0.003	0.003

GAD, generalized anxiety disorder symptoms; panic, panic disorder symptoms; and social anxiety, social anxiety symptoms. Changes in CFI and RMSEA are calculated as differences between the metric invariance model relative to the configural invariance model and between the scalar invariance model relative to the metric invariance model. ^aIn this model, three of nine threshold parameters were constrained to be equal. This model is compared with the full-metric model.

model fit was still good; however, model fit was diminished relative to the metric invariance model. Comparisons identified only two item thresholds that did not significantly differ across informants. This partial scalar invariance model yielded excellent model fit. However, with only two invariance item intercepts, this model failed to sufficiently support partial scalar MI.

The fit for the configural invariance model for panic disorder for youth self-reports across ages 9 and 12 was excellent (Table 4). The metric invariance model was also an excellent fit to the data. However, there was a substantial reduction in model fit as indexed by the CFI and a more modest reduction in fit according to the RMSEA. Comparisons identified three factor loadings that differed across age. Model fit for the partial metric invariance model was an excellent fit to the data. As only partial metric invariance was supported, when estimating scalar invariance, thresholds for items that did not evince equal factor loadings across time were freely estimated. After imposing

constraints on the other item thresholds across time, overall model fit was still good; however, model fit was diminished relative to the partial metric invariance model. Comparisons identified four item thresholds that did not significantly differ across time. This partial scalar invariance model yielded excellent model fit.

The fit for the configural invariance model for social phobia for youth self-reports across ages 9 and 12 was an excellent fit to the data. The fit of the metric invariance model was also good. However, there was a substantial reduction in model fit as indexed by the CFI, and a modest reduction in the RMSEA. Comparisons identified three factor loadings that did not statistically differ across age. Model fit for the partial metric invariance model was an excellent fit to the data. As only partial metric invariance was supported, when estimating scalar invariance, item thresholds for items that did not evince equal factor loadings across time were freely estimated. Three item thresholds were constrained

TABLE 3 | Tests of MI between child self- and maternal-reports at age 12.

	χ^2	<i>df</i>	<i>P</i>	CFI	RMSEA	Δ CFI	Δ RMSEA
GAD							
Configural	285.64	134	<0.001	0.944	0.049 (0.041–0.056)		
Full metric	315.25	142	<0.001	0.936	0.050 (0.043–0.058)	–0.008	0.001
Full scalar	447.81	150	<0.001	0.89	0.064 (0.058–0.071)	–0.046	0.014
Partial scalar ^a	315.25	142	<0.001	0.936	0.050 (0.043–0.058)	0.046	–0.014
Panic							
Configural	395.49	298	<0.001	0.91	0.026 (0.019–0.033)		
Full metric	414.35	310	<0.001	0.903	0.027 (0.019–0.033)	–0.007	0.001
Full scalar	907.12	335	<0.001	0.47	0.060 (0.055–0.064)	–0.433	0.033
Social							
Configural	257.64	76	<0.001	0.941	0.071 (0.061–0.080)		
Full metric	253.94	79	<0.001	0.943	0.068 (0.059–0.077)	0.002	–0.003
Full scalar	297.77	88	<0.001	0.932	0.071 (0.062–0.079)	–0.011	0.003

GAD, generalized anxiety disorder symptoms; panic, panic disorder symptoms; and social anxiety, social anxiety symptoms. Changes in CFI and RMSEA are calculated as differences between the metric invariance model relative to the configural invariance model and between the scalar invariance model relative to the metric invariance model. The model with best statistical fit is highlighted in bold. ^aIn this model, one of nine threshold parameters were constrained to be equal. This model is compared to the full-metric model.

TABLE 4 | Tests of MI for child self-reports across ages 9 and 12.

	χ^2	<i>df</i>	<i>P</i>	CFI	RMSEA	Δ CFI	Δ RMSEA
GAD							
Configural	201.47	134	<0.001	0.963	0.031 (0.022–0.04)		
Full metric	219.00	142	<0.001	0.958	0.033 (0.024–0.041)	–0.005	0.002
Full scalar	273.10	150	<0.001	0.933	0.040 (0.032–0.048)	–0.025	0.007
Partial scalar ^a	219.41	143	<0.001	0.958	0.032 (0.024–0.041)	0.025	–0.008
Panic							
Configural	324.73	298	<0.001	0.983	0.013 (0.000–0.022)		
Full metric	364.34	310	<0.001	0.964	0.019 (0.008–0.026)	–0.019	0.006
Partial metric ^b	328.83	307	<0.001	0.986	0.012 (0.000–0.021)	0.022	–0.007
Scalar ^c	442.99	316	<0.001	0.917	0.028 (0.022–0.034)	–0.069	0.016
Partial scalar ^d	334.14	310	<0.001	0.984	0.012 (0.000–0.021)	–0.002	0.000
Social							
Configural	143.92	76	<0.001	0.955	0.042 (0.031–0.052)		
Full metric	189.32	82	<0.001	0.929	0.051 (0.041–0.060)	–0.026	0.009
Partial metric ^e	142.38	78	<0.001	0.957	0.040 (0.030–0.051)	0.002	–0.002
Scalar ^f	150.62	80	<0.001	0.953	0.042 (0.031–0.052)	–0.004	0.002

GAD, generalized anxiety disorder symptoms; panic, panic disorder symptoms; and social anxiety, social anxiety symptoms. Changes in CFI and RMSEA are calculated as differences between the metric invariance model relative to the configural invariance model and between the scalar invariance model relative to the metric invariance model. The model with best statistical fit is highlighted in bold. ^aIn this model, two of nine threshold parameters were constrained to be equal. This model is compared to the full-metric model. ^bIn this model, 10 of 13 factor loading parameters were constrained to be equal. This model is compared to the configural invariance model. ^cIn this model, 3 of 13 threshold parameters were freely estimated across time. This model is compared to the partial metric model. ^dIn this model, 4 of 13 threshold parameters were constrained to be equal. This model is compared to the partial metric model. ^eIn this model, three of seven factor loading parameters were constrained to be equal. This model is compared to the configural invariance model. ^fIn this model, three of seven threshold parameters are constrained across time. This model is compared to the partial metric model.

across time. After imposing constraints on the item thresholds across informants to test for scalar invariance, model fit was not substantially diminished, supporting partial scalar MI.

Tests of MI: Maternal-Reports Across Ages 9 and 12

The configural invariance model for GAD for mother-reports across ages 9 and 12 was an excellent fit to the data (Table 5). The fit of the metric invariance model was good, and

imposing constraints on the factor loadings did not markedly diminish model fit, supporting metric invariance. After imposing constraints on the item thresholds across informants, overall model fit was still good and showed a minor reduction in model fit as indexed by the CFI and a trivial reduction in the RMSEA. Thus, scalar MI was supported.

The configural invariance model for panic disorder was an adequate fit to the data. However, there were problems in estimating the metric and scalar invariance models due to low endorsement rates of item response options across multiple items

TABLE 5 | Tests of MI for maternal-reports across ages 9 and 12.

	χ^2	<i>df</i>	<i>P</i>	CFI	RMSEA	Δ CFI	Δ RMSEA
GAD							
Configural	319.88	134	<0.001	0.950	0.052 (0.045–0.059)		
Full metric	333.82	142	<0.001	0.948	0.051 (0.044–0.058)	–0.002	–0.001
Full scalar	396.99	150	<0.001	0.933	0.057 (0.050–0.063)	–0.015	0.006
Panic							
Configural	501.40	298	<0.001	0.899	0.036 (0.031–0.042)		
Separation							
Configural	256.54	103	<0.001	0.900	0.054 (0.046–0.062)		
Full metric	270.92	110	<0.001	0.896	0.053 (0.045–0.061)	–0.004	–0.001
Partial metric ^a	252.09	109	<0.001	0.907	0.050 (0.042–0.059)	0.007	–0.004
Scalar ^b	307.39	117	<0.001	0.876	0.056 (0.049–0.064)	–0.031	0.006
Social							
Configural	234.44	76	<0.001	0.978	0.064 (0.054–0.073)		
Full metric	348.49	82	<0.001	0.963	0.079 (0.071–0.088)	–0.015	0.015
Partial metric ^c	211.41	81	<0.001	0.982	0.056 (0.047–0.065)	0.004	–0.008
Scalar ^d	219.18	85	<0.002	0.981	0.055 (0.046–0.064)	–0.001	–0.001

GAD, generalized anxiety disorder symptoms; panic, panic disorder symptoms; separation anxiety, separation anxiety disorder symptoms; social anxiety, social anxiety symptoms. Changes in CFI and RMSEA are calculated as differences between the metric invariance model relative to the configural invariance model and between the configural invariance model relative to the metric invariance model. The model with best statistical fit is highlighted in bold. ^aIn this model, seven of eight factor loading parameters were constrained to be equal. This model is compared to the configural invariance model. ^bIn this model, 3 of 13 threshold parameters were freely estimated across time. This model is compared to the partial metric model. ^cIn this model, six of seven factor loading parameters were constrained to be equal. This model is compared to the configural invariance model. ^dIn this model, one of seven threshold parameters were freely estimated across time. This model is compared to the partial metric model.

(i.e., empty cells in bivariate distributions). Thus, those models could not be adequately tested.

The configural invariance model for separation anxiety was good. The metric invariance model marginally reduced model fit, but it was enough to result in a less than adequate fit to the data. Comparisons of factor loadings identified one parameter that statistically differed across time. Model fit for the partial metric invariance model was good, supporting partial metric invariance. After adding constraints on item thresholds across time, model fit was reduced and demonstrated a poor fit to the data. Comparisons of item thresholds revealed that all parameters differed across time. Thus, there was no support for partial scalar invariance.

The fit for the configural invariance model for social phobia for maternal-reports across ages 9 and 12 was excellent. The metric invariance model was also an excellent fit to the data. However, there was a reduction in model fit as indexed by the CFI and the RMSEA. Comparisons of factor loadings identified six (of seven) factor loadings that did not statistically differ across age. Fit for the partial metric invariance model was excellent, supporting partial metric invariance. As only partial metric invariance was supported, when estimating scalar invariance, the item threshold for the item that did not evince equal factor loadings across time was freely estimated. After imposing constraints on the item thresholds across time to test for scalar invariance, overall model fit was excellent and the model did not demonstrate a substantial reduction in fit relative to the partial metric invariant model, supporting scalar invariance.

DISCUSSION

There has been much previous work examining factors and contexts that influence correspondence between parents' and their children's reports of psychopathology (Achenbach et al., 1987; De Los Reyes et al., 2015). However, there has been much less research examining measurement properties between informants that could influence the comparability of reports of youth behavior. Similarly, there has been little attention to examining MI across time, which is critical to understanding whether mean-level changes across time are contaminated by changes in measurement properties of items (Widaman et al., 2010). In the present study, we used the subscales from the SCARED to examine overall fit of each anxiety construct in each informant and at each assessment. Then we examined MI between mothers and their children at ages 9 and 12. Finally, we examined invariance for each rater from middle childhood to early adolescence. Overall, full MI was supported between children and their mothers for social anxiety at both ages 9 and 12, but not for any other SCARED subscale. We found support for partial metric invariance across mothers and children at age 9 for GAD. Longitudinally, full-scalar invariance was found for maternal reports of GAD over time and partial scalar invariance was supported for child reported panic and social anxiety and for maternal reported separation anxiety across the two waves.

Thus, we found support for full-scalar invariance across informants for only one SCARED subscale-social anxiety. This indicates that direct comparisons of mean levels of child and maternal reported anxiety symptoms are valid only for this scale of the SCARED.

To demonstrate “strong enough” measurement properties, there has to be consistent evidence supporting at least partial metric invariance across informants at both ages 9 and 12 (Marsh and Grayson, 1994). This indicates that a subset of items reflect the same target latent construct across mothers and their children. Thus, the construct reported on by each informant is conceptually similar in form and reflects rank-order associations among like-constructs. This suggests that for the scales demonstrating at least partial metric invariance inter-informant associations are meaningful. This condition was satisfied by the GAD scale at both ages 9 and 12. However, the lack of scalar invariance precludes comparing mean levels of generalized anxiety across informants (Millsap, 2011).

Panic, school avoidance, and separation anxiety showed the least evidence for MI. Although the panic symptom models demonstrated good fit to the data in our four preliminary models (i.e., separate informant and assessment; **Table 1**), tests of configural invariance across informant yielded poor fit to the data at age 9 and marginal fit to the data at age 12. Moreover, the fit of configural invariance models for school avoidance and separation anxiety was poor. Fit of these models may have been impacted by the developmental level of the children in the study. School avoidance and separation anxiety are typically observed at higher levels earlier in development. Thus, the coherence of the items in later childhood may be poorer than earlier in development (Hayward et al., 2000; Mathyssek et al., 2012). Moreover, incidence of panic continues to rise through adolescence (Beesdo et al., 2009) and item functioning may continue to change.

Examining the pattern of differences in factor loadings and thresholds between child and maternal reports, there is a consistent pattern of maternal reports having larger factor loadings and thresholds. Stronger factor loadings for maternal scores suggest that their ratings have greater precision and are better at discriminating between children with high and low levels of anxiety. Higher item thresholds for maternal than child-reported items suggest that symptoms need to be more severe for mothers to rate them as present relative to children. Taken together, these findings pose significant challenges to comparing levels of anxiety across mothers and youth. With only a few exceptions, these results argue against direct comparisons of mothers’ and youth’s anxiety ratings.

Our models testing longitudinal invariance demonstrated greater, albeit modest, support for MI over time for each informant taken separately. Maternal reports of youth GAD achieved full-scalar invariance, suggesting that scores from this scale are comparable from middle childhood to early adolescence. Child-reports of panic and social anxiety and maternal-reports of separation anxiety demonstrated a good fit to the data and partial scalar invariance. For these scales, there were some items that demonstrate invariance across time, permitting longitudinal comparisons of latent mean-level differences on the full set of items or examining mean-level differences on the subset of items. These comparisons should reflect true changes in the constructs, rather than being conflated with changes in item properties. Child-report of GAD and maternal-report of social anxiety each had a small number of items

with invariant factor loadings and threshold. Based on these results, there should be concern about relying on this set of items/scales to assess developmental changes on dimensions of anxiety symptoms, particularly when relying on child self-reports, and provide little basis for combining these ratings. However, our findings raise the question of whether these subscales evidence MI invariance over shorter periods of time and from pre- to post-test in evaluations of interventions. If psychometric functioning is changing over time, it may not be possible to distinguish intervention effects from measurement changes.

In our work, we focused on the primary, lower-order scales that demonstrated at least adequate fit for a one-factor model. In this evaluation, school phobia and some of the assessments of separation anxiety were not unitary factors. Thus, we did not evaluate these dimensions for MI. This suggests that more in-depth analysis of these dimensions is warranted, although there are only four items on the school phobia subscale, restricting alternative modeling strategies to yield better fit. Alternatively, because school phobia and separation anxiety are most common in early childhood, there may have been limited variability in responses for these dimensions at ages 9 and 12. Earlier assessments of school phobia and separation anxiety may have greater variability (Merikangas et al., 2010) and could lead to better fitting models. Examination of other instruments (e.g., the RCADS; Chorpita et al., 2000) across informants and time would provide leverage to determine whether this is a measure-specific or construct assessment challenge.

The present study employed an underutilized lens to better understand sources of discrepancy between child- and parent-reports of anxiety, as well as instability of anxiety symptoms from middle childhood to early adolescence. We employed a relatively large sample of mothers and youth who reported on multiple dimensions of anxiety symptomatology in middle childhood and early adolescence. However, our work has some limitations. First, our data came from a community sample with modest levels of symptomatology. Further, we had truncated ranges of item endorsement and collapsed our highest endorsement categories. We are unsure how this may have affected the findings. Second, we used only a single measure of anxiety, albeit one of the most frequently employed with children and adolescents. It is possible that other measures may demonstrate different levels of robustness across informants or longitudinal assessments. Third, we relied solely on comparisons between mothers and children. It is important to consider whether other caregivers (e.g., fathers) and teachers report on the same constructs of behavior problems in children. Fourth, we focused on individual subscales, rather than the total SCARED score. Thus, our work emphasizes these anxiety domains, but does not speak to the similarity in the overall structure of anxiety between informants and across time. Additional analyses would be necessary that focus on the broader dimensional model of the SCARED as a whole. Here, preliminary multidimensional models for the total SCARED produced good fit at age 9, but only a marginal fit at age 12. Thus, there is some evidence that the general structure may differ across time. Adequate testing of this more complex model would require a

larger sample with greater variability in anxiety severity. Fifth, there was some selection for continuing the study when youth had lower levels of internalizing problems at age 3. Though this difference was small.

In sum, our findings illustrate that it is critical to evaluate measurement properties of anxiety symptom rating scales using sophisticated measurement strategies. We found that associations across informants may be compromised by differences in the functioning of items on the scale being examined. In such cases, testing for differences between informants and combining ratings across informants to yield single indices of severity are both inappropriate. However, there was also evidence that measurement functioning for some anxiety dimensions remained consistent over time. Thus, a few of the dimensions of the SCARED are valid for assessing longitudinal change. As it may be difficult to know *a priori* which measures are appropriate for assessing change, there is a pressing need for a comprehensive effort to evaluate MI for the full range of scales commonly used to assess developmental trajectories and response to treatment in child and adolescent clinical psychology and psychiatry.

ETHICS STATEMENT

Informed consent was obtained prior to participation in accordance with the Declaration of Helsinki. The study was

approved by the institutional review board at Stony Brook University.

AUTHOR CONTRIBUTIONS

TO conceptualized the research questions, drafted the manuscript, and conducted analyses. MF provided assistance in conducting analyses and provided critical feedback on the manuscript. LD provided critical feedback on the manuscript. DK provided substantial contribution to the research design and critical feedback on the manuscript.

FUNDING

This work was partially supported by the National Institute of Mental Health Grants R01 MH069942 (PI: DK) and R01 MH107495 (PI: TO) and a National Science Foundation Graduate Research Fellowship (PI: MF).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01295/full#supplementary-material>

REFERENCES

- Achenbach, T. M., McConaughy, S. H., and Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychol. Bull.* 101, 213–232. doi: 10.1037/0033-2909.101.2.213
- Achenbach, T. M., and Rescorla, L. (2001). *Manual for the ASEBA School-age Forms & Profiles*. Burlington, VT: University of Vermont.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychol. Bull.* 107, 238–246. doi: 10.1037/0033-2909.107.2.238
- Beesdo, K., Knappe, S., and Pine, D. S. (2009). Anxiety and anxiety disorders in children and adolescents: developmental issues and implications for DSM-V. *Psychiatr. Clin. North Am.* 32, 483–524. doi: 10.1016/j.psc.2009.06.002
- Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., and Baugher, M. (1999). Psychometric properties of the screen for child anxiety related emotional disorders (SCARED): a replication study. *J. Am. Acad. Child Adolesc. Psychiatry* 38, 1230–1236. doi: 10.1097/00004583-199910000-00011
- Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J., et al. (1997). The screen for child anxiety related emotional disorders (SCARED): scale construction and psychometric characteristics. *J. Am. Acad. Child Adolesc. Psychiatry* 36, 545–553. doi: 10.1097/00004583-199704000-00018
- Bufferd, S. J., Dougherty, L. R., Carlson, G. A., Rose, S., and Klein, D. N. (2012). Psychiatric disorders in preschoolers: continuity from ages 3 to 6. *Am. J. Psychiatry* 169, 1157–1164. doi: 10.1176/appi.ajp.2012.12020268
- Byrne, B. M., Shavelson, R. J., and Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychol. Bull.* 105, 456–466. doi: 10.1177/10731911111419091
- 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
- Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., and Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: a revised child anxiety and depression scale. *Behav. Res. Ther.* 38, 835–855. doi: 10.1016/S0005-7967(99)00130-8
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334. doi: 10.1007/BF02310555
- De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A., Burgers, D. E., et al. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychol. Bull.* 141, 858–900. doi: 10.1037/a0038498
- De Los Reyes, A., Thomas, S. A., Goodman, K. L., and Kundery, S. M. (2013). Principles underlying the use of multiple informants' reports. *Annu. Rev. Clin. Psychol.* 9, 123–149. doi: 10.1146/annurev-clinpsy-050212-185617
- Feng, X., Shaw, D. S., and Silk, J. S. (2008). Developmental trajectories of anxiety symptoms among boys across early and middle childhood. *J. Abnorm. Psychol.* 117, 32–47. doi: 10.1037/0021-843X.117.1.32
- 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
- Frank, S. J., Van Egeren, L. A., Fortier, J. L., and Chase, P. (2000). Structural, relative, and absolute agreement between parents' and adolescent inpatients' reports of adolescent functional impairment. *J. Abnorm. Child Psychol.* 28, 395–402. doi: 10.1023/A:1005125211187
- Hayden, E. P., Durbin, C. E., Klein, D. N., and Olino, T. M. (2010). Maternal personality influences the relationship between maternal reports and laboratory measures of child temperament. *J. Pers. Assess.* 92, 586–593. doi: 10.1080/00223891.2010.513308
- Hayward, C., Killen, J. D., Kraemer, H. C., and Taylor, C. B. (2000). Predictors of panic attacks in adolescents. *J. Am. Acad. Child Adolesc. Psychiatry* 39, 207–214. doi: 10.1097/00004583-200002000-00021
- 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
- Kagan, J. (1997). Temperament and the reactions to unfamiliarity. *Child Dev.* 68, 139–143. doi: 10.2307/1131931

- Karver, M. S. (2006). Determinants of multiple informant agreement on child and adolescent behavior. *J. Abnorm. Child Psychol.* 34, 242–253. doi: 10.1007/s10802-005-9015-6
- Langberg, J. M., Epstein, J. N., Simon, J. O., Loren, R. E., Arnold, L. E., Hechtman, L., et al. (2010). Parent agreement on ratings of children's attention deficit/hyperactivity disorder and broadband externalizing behaviors. *J. Emot. Behav. Disord.* 18, 41–50. doi: 10.1177/1063426608330792
- MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (2006). Power analysis and determination of sample size for covariance structure modeling. *Psychol. Methods* 1, 130–149. doi: 10.1037/1082-989X.1.2.130
- Marsh, H. W., and Grayson, D. (1994). Longitudinal stability of latent means and individual differences: a unified approach. *Struct. Equ. Modeling* 1, 317–359. doi: 10.1080/10705519409539984
- Marsh, H. W., Hau, K. T., and Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Struct. Equ. Modeling* 11, 320–341. doi: 10.1207/s15328007sem1103_2
- Mathyssek, C. M., Olino, T. M., Hartman, C. A., Ormel, J., Verhulst, F. C., and Van Oort, F. V. (2013). Does the Revised Child Anxiety and Depression Scale (RCADS) measure anxiety symptoms consistently across adolescence? The TRAILS study. *Int. J. Methods Psychiatr. Res.* 22, 27–35. doi: 10.1002/mpr.1380
- Mathyssek, C. M., Olino, T. M., Verhulst, F. C., and van Oort, F. V. (2012). Childhood internalizing and externalizing problems predict the onset of clinical panic attacks over adolescence: the TRAILS study. *PLoS One* 7:e51564. doi: 10.1371/journal.pone.0051564
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58, 525–543. doi: 10.1007/BF02294825
- Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., et al. (2010). Lifetime prevalence of mental disorders in US adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *J. Am. Acad. Child Adolesc. Psychiatry* 49, 980–989. doi: 10.1016/j.jaac.2010.05.017
- Millsap, R. E. (2011). *Statistical Approaches to Measurement Invariance*. New York, NY: Taylor and Francis Group.
- Muthén, L. K., and Muthén, B. O. (1998–2017). *Mplus User's Guide*, 8th Edn. Los Angeles, CA: Muthén & Muthén.
- Najman, J. M., Williams, G. M., Nikles, J., Spence, S., Bor, W., O'Callaghan, M., et al. (2001). Bias influencing maternal reports of child behaviour and emotional state. *Soc. Psychiatry Psychiatr. Epidemiol.* 36, 186–194. doi: 10.1007/s001270170062
- Olino, T. M., and Klein, D. N. (2015). Psychometric comparison of self- and informant-reports of personality. *Assessment* 22, 655–664. doi: 10.1177/1073191114567942
- Olino, T. M., Klein, D. N., Dyson, M. W., Rose, S. A., and Durbin, C. E. (2010a). Temperamental emotionality in preschool-aged children and depressive disorders in parents: associations in a large community sample. *J. Abnorm. Psychol.* 119, 468–478. doi: 10.1037/a0020112
- Olino, T. M., Klein, D. N., Lewinsohn, P. M., Rohde, P., and Seeley, J. R. (2010b). Latent trajectory classes of depressive and anxiety disorders from adolescence to adulthood: descriptions of classes and associations with risk factors. *Compr. Psychiatry* 51, 224–235. doi: 10.1016/j.comppsy.2009.07.002
- Olino, T. M., Stepp, S. D., Keenan, K., Loeber, R., and Hipwell, A. (2014). Trajectories of depression and anxiety symptoms in adolescent girls: a comparison of parallel trajectory approaches. *J. Pers. Assess.* 96, 316–326. doi: 10.1080/00223891.2013.866570
- Prenoveau, J. M., Craske, M. G., Zinbarg, R. E., Mineka, S., Rose, R. D., and Griffith, J. W. (2011). Are anxiety and depression just as stable as personality during late adolescence? Results from a three-year longitudinal latent variable study. *J. Abnorm. Psychol.* 120, 832–843. doi: 10.1037/a0023939
- Reise, S. P., Widaman, K. F., and Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychol. Bull.* 114, 552–566. doi: 10.1037/0033-2909.114.3.552
- Rothen, S., Vandeleur, C. L., Lustenberger, Y., Jeanprêtre, N., Ayer, E., Gamma, F., et al. (2009). Parent–child agreement and prevalence estimates of diagnoses in childhood: direct interview versus family history method. *Int. J. Methods Psychiatr. Res.* 18, 96–109. doi: 10.1002/mpr.281
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika* 74, 107–120. doi: 10.1007/s11336-008-9101-0
- Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behav. Res.* 25, 173–180. doi: 10.1207/s15327906mbr2502_4
- Steinmetz, H. (2013). Analyzing observed composite differences across groups: is partial measurement invariance enough? *Methodology* 9, 1–12. doi: 10.1027/1614-2241/a000049
- Treutler, C. M., and Epkins, C. C. (2003). Are discrepancies among child, mother, and father reports on children's behavior related to parents' psychological symptoms and aspects of parent–child relationships? *J. Abnorm. Child Psychol.* 31, 13–27. doi: 10.1023/A:1021765114434
- Van Oort, F., Greaves-Lord, K., Verhulst, F., Ormel, J., and Huizink, A. (2009). The developmental course of anxiety symptoms during adolescence: the TRAILS study. *J. Child Psychol. Psychiatry* 50, 1209–1217. doi: 10.1111/j.1469-7610.2009.02092.x
- 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
- Widaman, K. F., Ferrer, E., and Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: measuring the same construct across time. *Child Dev. Perspect.* 4, 10–18. doi: 10.1111/j.1750-8606.2009.00110.x
- Youngstrom, E., Loeber, R., and Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *J. Consult. Clin. Psychol.* 68, 1038–1050. doi: 10.1037/0022-006X.68.6.1038

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

Copyright © 2018 Olino, Finsaas, Dougherty and Klein. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Assessment of Affect Lability: Psychometric Properties of the ALS-18

Anna Contardi^{1*}, Claudio Imperatori¹, Italia Amati², Michela Balsamo³ and Marco Innamorati¹

¹ Department of Human Sciences, European University of Rome, Rome, Italy, ² Dipartimento di Tecnologie, Comunicazione e Società, Università degli Studi Guglielmo Marconi, Rome, Italy, ³ Dipartimento di Scienze Psicologiche, della Salute e del Territorio, Università degli Studi "G. d'Annunzio" Chieti-Pescara, Chieti, Italy

OPEN ACCESS

Edited by:

Pietro Cipresso,
Istituto Auxologico Italiano (IRCCS),
Italy

Reviewed by:

Dejan Stevanovic,
Clinic for Neurology and Psychiatry
for Children and Youth, Serbia
Marco Lauriola,
Sapienza Università di Roma, Italy
Elisa Pedrolì,
Istituto Auxologico Italiano (IRCCS),
Italy

*Correspondence:

Anna Contardi
anna.contardi@unier.it

Specialty section:

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

Received: 02 February 2018

Accepted: 14 March 2018

Published: 29 March 2018

Citation:

Contardi A, Imperatori C, Amati I,
Balsamo M and Innamorati M (2018)
Assessment of Affect Lability:
Psychometric Properties of the
ALS-18. *Front. Psychol.* 9:427.
doi: 10.3389/fpsyg.2018.00427

Affect lability, an important aspect of emotion dysregulation, characterizes several psychiatric conditions. The short Affective Lability Scales (ALS-18) measures three aspects of changeability between euthymia and affect states (Anxiety/Depression, AD; Depression/Elation, DE; and Anger, Ang). The aim of our study was to investigate the psychometric characteristics of an Italian version of the ALS-18 in a sample of adults recruited from the general population. The sample was composed of 494 adults (343 women and 151 men) aged 18 and higher (mean age = 31.73 years, *SD* = 12.6). All participants were administered a checklist assessing socio-demographic variables, the ALS-18 and measures of depression and difficulties in emotion regulation. Confirmatory factor analyses indicated adequate fit of the three-factor model (RMSEA = 0.061, 95% CI = 0.054/0.069; CFI = 0.99; SRMR = 0.055), and the presence of a higher-order general factor. Internal consistency was satisfactory for all the lower-order dimensions and the general factor (ordinal $\alpha > 0.70$). The ALS-18 was significantly associated with concurrent measures of depression and difficulties in emotion regulation. These findings indicate that the ALS-18 is a valid and reliable instrument for measuring affect lability, although discriminant validity of subdimensions scores could be problematic.

Keywords: affect lability, emotion dysregulation, Affective Lability Scales (ALS-18), psychometric properties of ALS-18, validity and reliability of ALS-18

INTRODUCTION

Past studies suggested that emotion dysregulation could be associated with the development and maintenance of various psychiatric disorders and maladaptive behaviors (Amstadter, 2008; Aldao et al., 2010; Svaldi et al., 2012; American Psychiatric Association, 2013; Contardi et al., 2013). An important aspect of emotion dysregulation is affect lability intended as abnormally frequent, intense, and wide ranging changes in affective states (Thompson et al., 2011). Affect lability is present in several psychiatric conditions and is characteristic of the bipolar disorder and the borderline personality disorder (Henry et al., 2008; Aminoff et al., 2012; Reich et al., 2012). For example, affect lability has recently been found to mediate the relationship between childhood trauma and suicide attempts in bipolar patients (Aas et al., 2017).

The construct of affect lability has strong relationships with other personological constructs such as neuroticism (McCrae and Costa, 1987) and cyclothymia (Akiskal, 2001). Neuroticism (also known as emotional stability-instability or negative emotionality) is part of major models of normal personality structure (i.e., Eysenck's Three Factor model, and the Big Five model) and

it is an ubiquitous element of many personality measures (McCrae and Costa, 1987; Zuckerman et al., 1993). Negative emotionality is a central component in neuroticism, along with cognitive and behavioral facets (McCrae and Costa, 1987). For example, in the NEO Personality Inventory (NEO-PI-3) neuroticism is composed of six facets (i.e., Anxiety, Angry Hostility, Depression, Self-Consciousness, Impulsiveness, and Vulnerability) (McCrae et al., 2005). Questionnaires assessing neuroticism generally measure the frequency of negative emotional states and how easily they are experienced by the individual (e.g., “Get stressed out easily,” “Often feel blue,” “Lose my temper.”) (Maples et al., 2014). Conversely, measures of affect lability assess how frequently emotionality change between two specific polarized emotions (e.g., “At times I feel just as realized as everyone else and then within minutes I become so nervous that I feel light-headed and dizzy,” “I switch back and forth between being extremely energetic and having so little energy that it’s a huge effort just to get where I’m going.”). Thus, although affect lability could be considered a facet of neuroticism, it has a close relationship with the psychiatric construct of cyclothymia. Kraepelin described the cyclothymic disposition as one of the constitutional substrates of the manic-depressive illness (Akiskal, 2001). According to Akiskal (2001) cyclothymic individuals report short cycles of mood swings, characterized mainly by depression and hypomania but also by labile-angry-irritable moods. Mood swings are essentially biphasic, with lethargy alternating with eutonia, or unexplained tearfulness alternating with excessive punning and jocularity (Akiskal and Mallya, 1987).

To measure affect lability, Harvey et al. (1989) developed the Affective Lability Scales (ALS), a 58-item questionnaire measuring changeability among euthymia and four affect states (i.e., depression, elation, anger, and anxiety). The four studies presented in their research indicated satisfactory reliability (internal consistency and stability), discriminant validity with a measure of affect intensity, and concurrent validity with measures of depression (Harvey et al., 1989). Nevertheless, Oliver and Simons (2004) considered the ALS too lengthy and developed a 18-item short form (ALS-18) consisting of at least two items from each dimensions of the ALS. In a non-clinical sample of university students a confirmatory factor analysis supported the adequacy of both a three-factor structure (Anxiety/Depression, AD; Depression/Elation, DE, Anger, Ang) (Bentler–Bonnett Non-normed Fit Index [NNFI] = 0.90; Comparative Fit Index [CFI] = 0.92; Root Mean Square Error of Approximation [RMSEA] = 0.06), and a six-factor model reflecting the structure of the original 58-item version (NNFI = 0.94; CFI = 0.96; RMSEA = 0.05) (Oliver and Simons, 2004). However, the six-factor model included two dimensions composed of only 2 items (Elation and Hypomania), and internal consistency was found to be lower than for the three-factor model. Further studies investigated successfully the adequacy of the three-factor model in different clinical populations (e.g., personality disorders, bipolar disorder patients and relatives, and ADHD) (Look et al., 2010; Aas et al., 2015; Weibel et al., 2017). For example, Look et al. (2010) investigated factor structure and psychometric properties of the ALS-18 in patients with personality disorders

and individuals without any psychiatric conditions, and reported satisfactory reliability and good discriminant validity (i.e., people with DSM-IV Cluster B personality disorders reported higher scores than individuals with Cluster A and Cluster C disorders, and people without any psychiatric condition) (Look et al., 2010). Discriminant validity was also supported when differentiating ADHD patients from healthy controls (Weibel et al., 2017), or bipolar patients from relatives and healthy controls (Aas et al., 2015).

Based on the results presented above, and given that the psychometric characteristics of an Italian version of the ALS-18 (as well as the original 54-item version) have not already been investigated, the aim of our study was to investigate factor structure, validity and reliability of the Italian version of the ALS-18 in a non-clinical sample of adults from the general population, as a first step for a cross-cultural validation of the questionnaire. In line with previous studies (Look et al., 2010; Aas et al., 2015; Weibel et al., 2017), we tested the fit of a three-factor model and its superiority over a one-factor model. Considering that previous studies indicated that dimensions of the ALS-18 could be strongly correlated ($r \geq 0.64$) (Amstadter, 2008; Look et al., 2010), we also tested whether a hierarchical factor model, with three specific factors (AD, DE, and Ang) loading on a higher-order general factor, or a bi-factor model, with each items loading upon both a group factor (AD, DE, and Ang) and a general factor (AL) could represent well the factor structure of the ALS-18 (Figure 1).

MATERIALS AND METHODS

Participants and Procedure

The sample was composed of 494 adults (343 women and 151 men). Mean age of the participants was 31.73 years ($SD = 12.61$). Inclusion criteria were ages of 18 and higher. Exclusion criteria were the presence of any condition affecting the ability to take the assessment, including illiteracy or denial of informed consent. The sample was recruited through advertisements (flyers, newspaper and online ads) posted for established community groups, and directly from university communities ($n = 194$) of the authors of the present research. Individuals were approached by psychologists who informed them about the aim of the study and explained how to fill-in the questionnaire. They participated in the study voluntarily and received no payment. Each participant provided written, informed consent prior to data collection. The study protocol received ethics approval from the local research ethics review board. Sociodemographic characteristics of the sample are reported in Table 1.

Measures

At entry into the study, all participants were administered a checklist assessing socio-demographic variables (sex, age, marital status, job, and school attainment), and the Italian version of the ALS-18, the Teate Depression Inventory (TDI; Balsamo et al., 2014; Balsamo and Saggino, 2014), and the Difficulties in Emotion Regulation Scale (DERS) (Gratz and Roemer, 2004).

The original version of the ALS-18 is an 18-item self-report measure used to assess the affect lability. Items are rated on a

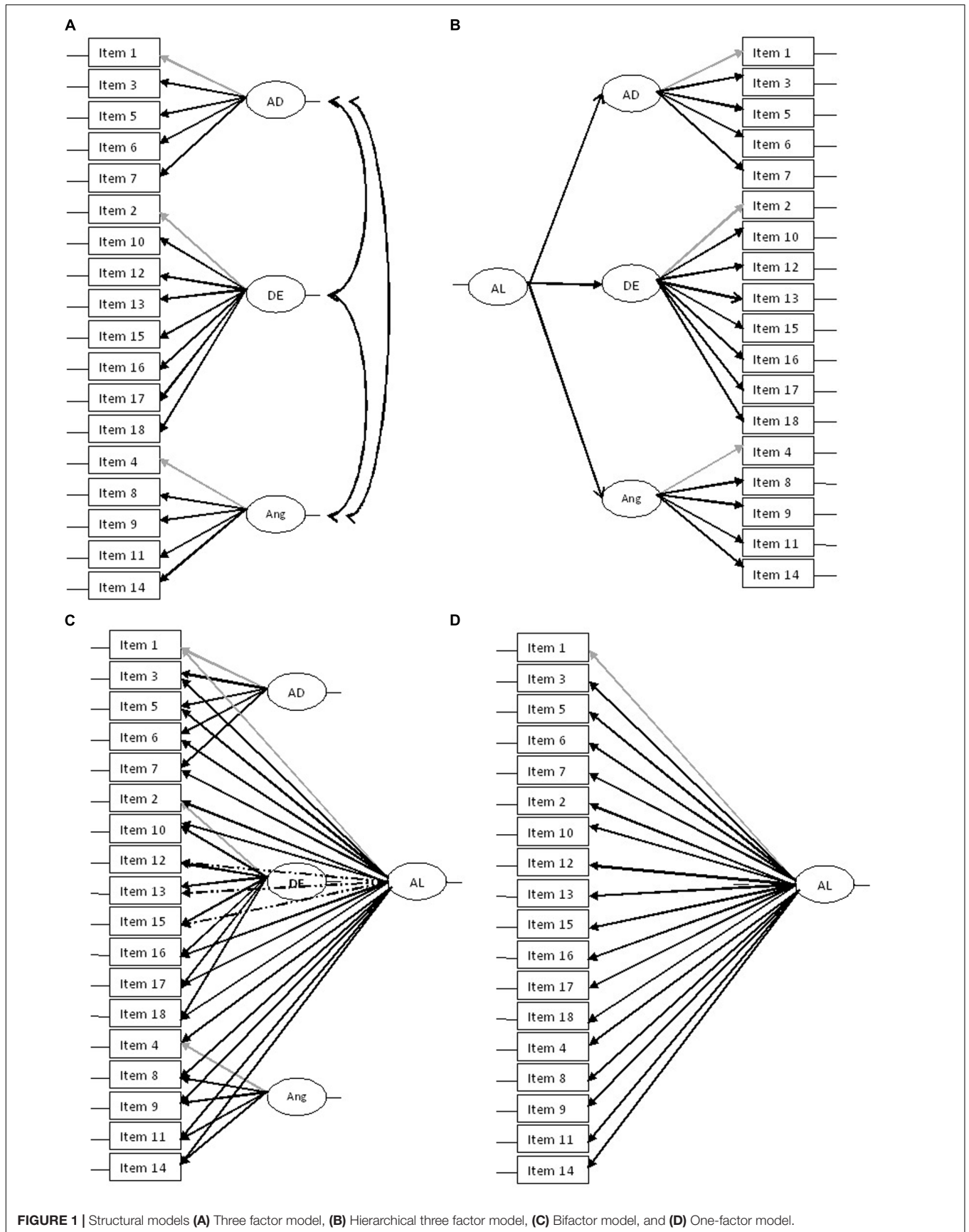


FIGURE 1 | Structural models (A) Three factor model, (B) Hierarchical three factor model, (C) Bifactor model, and (D) One-factor model.

TABLE 1 | Characteristics of the sample ($n = 494$).

Variables	Percentages (Mean \pm SD)
Sex	
Men	30.6%
Women	69.4%
Age – $M(SD)$	(31.73 \pm 12.61)
Marital status	
Not-married (including widowed and divorced)	79.7%
Married	20.3%
School attainment (number of years)	
≤ 8	7.1%
≤ 13	53.8%
≥ 16	39.1%
Job status	
Student	43.6%
Employed	47.2%
Unemployed	5.1%
Retired	4.1%

SD, standard deviation.

4-point Likert type scale (from 0 = very uncharacteristic of me to 3 = very characteristic of me). In the present study, we used an Italian adaptation of this scale. Two bilingual researchers adapted the present version of the questionnaire from the original English version using the back-translation procedure.

The TDI is a 21-item self-report instrument designed to assess major depressive disorder as specified by the latest editions of the Diagnostic and Statistical Manual of Mental Disorders (DSM; American Psychiatric Association, 2000, 2013), in order to overcome psychometric weaknesses of existing measures of depression (Balsamo and Saggino, 2007). Each item is rated on a five-point Likert-type scale, ranging from 0 (always) to 4 (never). The TDI demonstrated good psychometric properties (Balsamo et al., 2013, 2014, 2015a,b,c; Innamorati et al., 2013; Saggino et al., 2017). In the present sample, Cronbach's α was 0.93.

The DERS is a 36-item multidimensional self-report measure assessing the individual's characteristic patterns of emotion regulation. Items are rated on a 5-point Likert-type scale (from 1 = *almost never* to 5 = *almost always*) indicating the degree to which each statement describes the respondent's behavior. It contains the following six subscales: (1) *Non-acceptance* of emotional responses; (2) Difficulties engaging in goal-directed behavior when experiencing negative emotions (*Goals*); (3) Impulse control difficulties when experiencing negative emotions (*Impulse*); (4) *Lack of emotional awareness*; (5) *Limited access* to emotion regulation *strategies* that are perceived as effective;

and (6) *Lack of emotional clarity*. In the current sample internal consistency ranged between 0.76 for Awareness and 0.88 for Acceptance.

Statistical Analysis

All the analyses were performed with the Statistical Package for the Social Sciences (SPSS) 19.0 for Windows, and Lisrel 8.80 (Jöreskog and Sörbom, 2006).

Confirmatory factor analysis was performed using a Robust Diagonally Weighted Least Squares estimator (DWLSE) with a polychoric correlation matrix. Model fit was assessed using the following indices: (1) the Root Mean Square Error of Approximation (RMSEA), with values between 0.05 and 0.08 indicative of adequacy of the model, and values below 0.05 indicating evidence of good fit (Browne and Cudek, 1993; Hu and Bentler, 1999); (2) the Comparative Fit Index (CFI), with values greater than 0.95/0.96 indicating good fit of the model; (3) the Standardized Root Mean Square Residual (SRMR), with values of less than 0.08 indicating good fit (Hu and Bentler, 1999); and (4) the Satorra-Bentler scaled chi-square (χ^2) test and the normed χ^2 ($\chi^2/\text{degrees of freedom}$). P -values for the χ^2 test greater than 0.05 and a normed χ^2 less than 3 (Schreiber et al., 2006) indicate that the model is an adequate fit to the data, although the χ^2 test over-reject true models for large samples. The Expected Cross-validation Index (ECVI) was used to compare competing models (Browne and Cudek, 1989).

Although the aim of the present study was to compare the four competing factor models and select the one with the best fit, the proposed three-factor model and the hierarchical three-factor model are equivalent (i.e., each factor directly or indirectly is related to all the other latent variables) and yield the same fit to the data (MacCallum et al., 1993; Leone, 2009). Thus, the comparison of fit indices is inconclusive in demonstrating which of the two models is better.

As measures of reliability, we reported ordinal Cronbach's alpha (α) (Zumbo et al., 2007). Associations with sociodemographic variables and other measures were evaluated by means of a series of t -tests and Pearson's r indices of correlations.

RESULTS

Confirmatory Factor Analysis

The bi-factor model did not converge, and the statistical software issued a warning message indicating that Phi (i.e., the variance/covariance matrix between latent variables) was not positive definite. The other competing models all had significant

TABLE 2 | Fit indices for the competing factor models.

Model	χ^2 (df)	RMSEA (90% CI)	CFI	SRMR	ECVI (90% CI)
Original three-factor model	369.02* (132)	0.061 (0.054/0.069)	0.99	0.054	0.93 (0.81/1.05)
Three-factor hierarchical model	369.02* (132)	0.061 (0.054/0.069)	0.99	0.054	0.93 (0.81/1.05)
One-factor model	610.30* (135)	0.085 (0.079/0.092)	0.98	0.066	1.41 (1.26/1.58)

* $p < 0.001$.

χ^2 ($p < 0.001$), indicating potential misfit of the models (see **Table 2**). On the contrary other fit indices indicated the adequacy of both the one-factor model (RMSEA = 0.072, 95% CI = 0.064/0.079; CFI = 0.98; SRMR = 0.066), and the three factor model (RMSEA = 0.061, 95% CI = 0.054/0.069; CFI = 0.99; SRMR = 0.055). Nevertheless, the ECVI suggested the superiority of the three-factor model (0.93 vs. 1.41).

The latent dimensions of the three-factor model were highly correlated (r between 0.83 for AD/DE and 0.93 for AD/Ang), and when modeling a hierarchical factor model, factor loadings on the higher-order general factor were all significant (0.93 for AD, 0.89 for DE and 0.99 for Ang). Each item of the ALS-18 loaded significantly on its hypothesized dimension (**Table 3**).

Psychometric Properties of the ALS-18

Considering that fit indices suggested that the three-factor model could represent better the latent structure of the ALS-18, the following analyses will be based on this factor model. Internal consistency of the ALS-18 was satisfactory for all the lower-order dimensions (**Table 4**), and for the general factor (ordinal alpha = 0.95). Scores on the ALS-18 were not associated with sex ($p > 0.05$ for t -tests), or with age (r between -0.09 for DE and -0.14 for AD). The ALS-18 dimensions and the general factor were all significantly associated with concurrent measures of depression and difficulties in emotion regulation (**Table 4**). Correlations with the TDI were all significant and moderate ($r \geq 0.4$), ranging from 0.47 for Ang to 0.59 for AD.

DISCUSSION

In our sample, the three-factor model fitted the data well. This is in line with previous studies which evaluated the structure of other versions of the ALS-18 (Look et al., 2010; Aas et al., 2015; Weibel et al., 2017). Our results could also support the presence of a higher-order general factor suggesting the possibility to compute a total score as generally reported in the literature (Look et al., 2010; Aas et al., 2015; Weibel et al., 2017). Nevertheless, when comparing the three-factor and the hierarchical models

TABLE 3 | Standardized factor loadings for the ALS-18 items ($n = 494$).

Items	AD	Items	DE	Items	Ang
ALS1	0.83	ALS2	0.66	ALS4	0.80
ALS3	0.86	ALS10	0.77	ALS8	0.85
ALS5	0.84	ALS12	0.86	ALS9	0.87
ALS6	0.72	ALS13	0.62	ALS11	0.69
ALS7	0.92	ALS15	0.72	ALS14	0.79
		ALS16	0.85		
		ALS17	0.62		
		ALS18	0.61		
Ordinal alpha	0.91		0.89		0.89

AD, Anxiety/Depression; DE, Depression/Elation; Ang, Anger.

TABLE 4 | Correlations between measures and descriptive statistics ($n = 494$).

	AD	DE	Ang	AL	TDI	Non-acceptance	Goals	IMPULSE	Awareness	Strategies	Clarity
$M \pm SD$	0.68 ± 0.62	0.63 ± 0.50	0.54 ± 0.58	0.62 ± 0.51	26.44 ± 14.02	13.02 ± 5.87	13.54 ± 3.93	12.89 ± 5.23	14.67 ± 4.73	16.69 ± 6.67	11.14 ± 4.33
AD	–				0.59**	0.54**	0.43**	0.51**	0.15**	0.58**	0.48**
DE	0.70**	–			0.52**	0.44**	0.38**	0.43**	0.10*	0.46**	0.40**
Ang	0.78**	0.74**	–		0.47**	0.44**	0.34**	0.49**	0.16**	0.47**	0.36**
AL	0.90**	0.92**	0.91**	–	0.56**	0.52**	0.42**	0.52**	0.15**	0.55**	0.46**

**Significant for $p < 0.01$; Significant for $p < 0.05$. AD, Anxiety/Depression; DE, Depression/Elation; Ang, Anger; AL, Affect Lability; TDI, Teate Depression Inventory.

the design of our study was not conclusive in demonstrating the superiority of one model over the other (MacCallum et al., 1993; Leone, 2009). Furthermore, the bi-factor model was empirical underidentified denoting that our study did not permit a test of the hypothesized bi-factor model (Green and Yang, 2017). However, as far as we know, this was the first tentative study which assessed directly the fit of a hierarchical or bifactor model for the ALS-18.

Inter-correlations among the three dimensions of the ALS-18 were high ($r \geq 0.83$), and despite also other studies reported strong intercorrelations among latent factors (Amstadter, 2008; Look et al., 2010), our figures are higher than those reported in those studies. These data could indicate non-satisfactory discriminant validity of subdimensions scores (Look et al., 2010). The three dimensions and the general factor all had adequate internal consistency.

ALS-18 subdimensions and total score were not associated with sex or age. In the past only Harvey et al. (1989) have investigated this topic for the 54-item ALS and reported sex differences for the depression scale only, suggesting a possible tendency for men to experience depression as a more transient and changeable phenomenon than do women. Conversely, the ALS-18 subdimensions and the general factor were all significantly associated with concurrent measures of depression and difficulties in emotion regulation. Our results are partially discordant from findings of previous studies. For example, Oliver and Simons (2004), in a non-clinical sample of university students, reported significant but negative correlations between the ALS-18 and the Center for Epidemiologic Studies-Depression Scale (r between -0.33 for Ang and -0.47 for AD). Conversely in our sample, the correlations were all significant and positive. Unfortunately, Oliver and Simons (2004) did not comment this result in their article. Nevertheless, Weibel et al. (2017), who administered a short version of the Beck Depression Inventory (BDI), reported an r of 0.34 ($p < 0.05$) between the ALS-18 total score and the BDI. The positive association between depression and affect lability possibly indicates that people with higher lability could experience phases of depression, despite the two constructs are only moderately correlated and both depression and affect lability should be assessed independently (Weibel et al., 2017). The results could also be seen as supportive of the concept of “soft bipolars spectrum” (Akiskal and Mallya, 1987). In fact, several patients who receive a diagnosis of major depression (MDD) have subthreshold symptoms of bipolarity, which includes biphasic mood swings and cyclothymic traits. These patients differ from pure MDD patients and patients with bipolar disorder for their temperamental profile and for clinical variables (Innamorati et al., 2015). Innamorati et al. (2015), investigating the role of cyclothymic temperament in characterizing mood disorder patients, evidenced that around 39% of inpatients with unipolar depression could be included in the soft bipolar spectrum according to their affective temperament. These patients seem to differ from patients with pure major mood disorders for levels of hopelessness and suicide risk.

Correlations with the DERS were also generally moderate ($r \geq 0.4$) with the exclusion of the dimension Awareness of the DERS whose correlations with the ALS-18 were weak (r between 0.10 and 0.16). This means that also the relationship between affect lability and difficulties in emotion regulation could be complex with a partial independence of these two constructs.

Our results must be considered in light of some issues referred to the design of the study. First, in our sample there was a disproportion of female participants compared to males probably associated with the recruitment of university students from the authors' university communities. This bias also prevented us from assessing structural invariance of the questionnaire between sex. Second, our results are based on a general community sample of adults, composed mostly of young adults, which limits the generalizability of these findings to clinical conditions or older adults. Third, we administered only self-report measures potentially affected by social desirability. In conclusion, this may considered only a first necessary step in the process of the cross-cultural validation of the ALS-18.

CONCLUSION

Our results indicate that the Italian version of the ALS-18 can produce valid and reliable assessments of affect lability. Additional studies are needed from clinical samples or samples of older adults with further psychometric assessments of reliability and measurement errors to draw clear recommendations for clinical practice use and research.

ETHICS STATEMENT

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

AUTHOR CONTRIBUTIONS

AC was involved in the study conception and design, acquisition, analysis and interpretation of data, and drafting of the manuscript. CI was involved in the study conception and design, interpretation of data, and the critical revision of the manuscript, provided final approval of the version to be published. IA was involved in the acquisition of data, interpretation of data, and the critical revision of the manuscript. MB was involved in the study conception and design and the analysis and interpretation of data. MI was involved in the study conception and design, acquisition of data, analysis and interpretation of data, and drafting of the manuscript, and provided final approval of the version to be published.

REFERENCES

- Aas, M., Henry, C., Bellivier, F., Lajnef, M., Gard, S., Kahn, J., et al. (2017). Affective lability mediates the association between childhood trauma and suicide attempts, mixed episodes and co-morbid anxiety disorders in bipolar disorders. *Psychol. Med.* 47, 902–912. doi: 10.1017/S0033291716003081
- Aas, M., Pedersen, G., Henry, C., Bjella, T., Bellivier, F., Leboyer, M., et al. (2015). Psychometric properties of the Affective Lability Scale (54 and 18-item version) in patients with bipolar disorder, first-degree relatives, and healthy controls. *J. Affect. Disord.* 172, 375–380. doi: 10.1016/j.jad.2014.10.028
- Akiskal, H. S. (2001). Dysthymia and cyclothymia in psychiatric practice a century after Kraepelin. *J. Affect. Disord.* 62, 17–31. doi: 10.1016/S0165-0327(00)00347-5
- Akiskal, H. S., and Mallya, G. (1987). Criteria for the “soft” bipolar spectrum: treatment implications. *Psychopharmacol. Bull.* 23, 68–73.
- Aldao, A., Nolen-Hoeksema, S., and Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: a meta-analytic review. *Clin. Psychol. Rev.* 30, 217–237. doi: 10.1016/j.cpr.2009.11.004
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edn. Washington, DC: American Psychiatric Association. doi: 10.1176/appi.books.9780890425596
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. Arlington, VA: American Psychiatric Association.
- Aminoff, S. R., Jensen, J., Lagerberg, T. V., Hellvin, T., Sundet, K., Andreassen, O. A., et al. (2012). An association between affective lability and executive functioning in bipolar disorder. *Psychiatry Res.* 198, 58–61. doi: 10.1016/j.psychres.2011.12.044
- Amstadter, A. (2008). Emotion regulation and anxiety disorders. *J. Anxiety Disord.* 22, 211–221. doi: 10.1016/j.janxdis.2007.02.004
- Balsamo, M., Carlucci, L., Sergi, M. R., Klein Murdock, K., and Saggino, A. (2015a). The mediating role of early maladaptive schemas in the relation between co-rumination and depression in young adults. *PLoS One* 10:e0140177. doi: 10.1371/journal.pone.0140177
- Balsamo, M., Innamorati, M., Van Dam, N. T., Carlucci, L., and Saggino, A. (2015b). Measuring anxiety in the elderly: psychometric properties of the state trait inventory of cognitive and somatic anxiety (STICSA) in an elderly Italian sample. *Int. Psychogeriatr.* 27, 999–1008. doi: 10.1017/S1041610214002634
- Balsamo, M., Macchia, A., Carlucci, L., Picconi, L., Tommasi, M., Gilbert, P., et al. (2015c). Measurement of external shame: an inside view. *J. Pers. Assess.* 97, 81–89. doi: 10.1080/00223891.2014.947650
- Balsamo, M., Giampaglia, G., and Saggino, A. (2014). Building a new Rasch-based self-report inventory of depression. *Neuropsychiatr. Dis. Treat.* 13, 153–165. doi: 10.2147/NDT.S53425
- Balsamo, M., Romanelli, R., Innamorati, M., Ciccarese, G., Carlucci, L., and Saggino, A. (2013). The state-trait anxiety inventory: shadows and lights on its construct validity. *J. Psychopathol. Behav. Assess.* 35, 475–486. doi: 10.1007/s10862-013-9354-5
- Balsamo, M., and Saggino, A. (2007). Test per l’assessment della depressione nel contesto italiano: un’analisi critica. [Tests for the assessment of depression in Italian Context: a critical review]. *Psicoterapia Cogn. Comportamentale* 13, 167–199.
- Balsamo, M., and Saggino, A. (2014). Determining a diagnostic cut-off on the Teate Depression Inventory. *Neuropsychiatr. Dis. Treat.* 10, 987–995. doi: 10.2147/NDT.S55706
- Browne, M. W., and Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. *Br. J. Math. Stat. Psychol.* 37, 62–83. doi: 10.1207/s15327906mbr2404_4
- Browne, M. W., and Cudeck, R. (1993). “Alternative ways of assessing model fit” in *Testing Structural Equation Models*, ed. J. S. Long (Newbury Park, CA: Sage), 136–162.
- Contardi, A., Farina, B., Fabbriatore, M., Tamburello, S., Scapellato, P., Penzo, I., et al. (2013). [Difficulties in emotion regulation and personal distress in young adults with social anxiety]. *Riv. Psichiatr.* 48, 155–161. doi: 10.1708/1272.14040
- Gratz, K. L., and Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the Difficulties in Emotion Regulation Scale. *J. Psychopathol. Behav. Assess.* 26, 41–54. doi: 10.1023/B:JOBA.0000007455.08539.94
- Green, S., and Yang, Y. (2017). Empirical underidentification with the bifactor model: a case study. *Educ. Psychol. Meas.* 1–20. doi: 10.1177/0013164417719947
- Harvey, P. D., Greenberg, B. R., and Serper, M. R. (1989). The affective lability scales: development, reliability, and validity. *J. Clin. Psychol.* 45, 786–793. doi: 10.1002/1097-4679(198909)45:5<786::AID-JCLP2270450515>3.0.CO;2-P
- Henry, C., Van den Bulke, D., Bellivier, F., Roy, I., Swendsen, J., M’Bailara, K., et al. (2008). Affective lability and affect intensity as core dimensions of bipolar disorders during euthymic period. *Psychiatry Res.* 159, 1–6. doi: 10.1016/j.psychres.2005.11.016
- 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
- Innamorati, M., Rihmer, Z., Akiskal, H., Gonda, X., Erbutto, D., Belvederi Murri, M., et al. (2015). Cyclothymic temperament rather than polarity is associated with hopelessness and suicidality in hospitalized patients with mood disorders. *J. Affect. Disord.* 170, 161–165. doi: 10.1016/j.jad.2014.08.042
- Innamorati, M., Tamburello, S., Contardi, A., Imperatori, C., Tamburello, A., Saggino, A., et al. (2013). Psychometric properties of the attitudes toward self-revised in Italian young adults. *Depress. Res. Treat.* 2013:209216. doi: 10.1155/2013/209216
- Jöreskog, K. G., and Sörbom, D. (2006). *LISREL for Windows [Computer Software]*. Lincolnwood, IL: Scientific Software International, Inc.
- Leone, L. (2009). Testing conceptual distinctions among Carver and White’s (1994) BAS scales: a comment and extension on Smillie, Jackson, and Dalglish (2006). *Pers. Individ. Dif.* 46, 54–59. doi: 10.1016/j.paid.2008.09.004
- Look, A. E., Flory, J. D., Harvey, P. D., and Siever, L. J. (2010). Psychometric properties of a short form of the Affective Lability Scale (ALS-18). *Pers. Individ. Dif.* 49, 187–191. doi: 10.1016/j.paid.2010.03.030
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., and Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychol. Bull.* 114, 185–199. doi: 10.1037/0033-2909.114.1.185
- Maples, J. L., Guan, L., Carter, N. T., and Miller, J. D. (2014). A test of the International Personality Item Pool representation of the Revised NEO Personality Inventory and development of a 120-item IPIP-based measure of the five-factor model. *Psychol. Assess.* 26, 1070–1084. doi: 10.1037/pas0000004
- McCrae, R. R., and Costa, P. T. Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *J. Pers. Soc. Psychol.* 52, 81–90. doi: 10.1037/0022-3514.52.1.81
- McCrae, R. R., Costa, P. T. Jr., and Martin, T. A. (2005). The NEO-PI-3: a more readable revised NEO personality inventory. *J. Pers. Assess.* 84, 261–270. doi: 10.1207/s15327752jpa8403_05
- Oliver, M. N. L., and Simons, J. S. (2004). The Affective Lability Scales: development of a short-form measure. *Pers. Individ. Dif.* 37, 1279–1288. doi: 10.1016/j.paid.2003.12.013
- Reich, D. B., Zanarini, M. C., and Fitzmaurice, G. (2012). Affective lability in bipolar disorder and borderline personality disorder. *Compr. Psychiatry* 53, 230–237. doi: 10.1016/j.comppsy.2011.04.003
- Saggino, A., Carlucci, L., Sergi, M. R., D’Ambrosio, I., Fairfield, B., Cera, N., et al. (2017). A validation study of the psychometric properties of the Other as Shamer Scale–2. *SAGE Open* 7, 1–10. doi: 10.1177/2158244017704241
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., and King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: a review. *J. Educ. Res.* 99, 323–337. doi: 10.3200/JOER.99.6.323-338
- Svaldi, J., Griepentstroh, J., Tuschen-Caffier, B., and Ehring, T. (2012). Emotion regulation deficits in eating disorders: a marker of eating pathology or general psychopathology? *Psychiatry Res.* 197, 103–111. doi: 10.1016/j.psychres.2011.11.009
- Thompson, R. J., Berenbaum, H., and Bredemeier, K. (2011). Cross-sectional and longitudinal relations between affective instability and depression. *J. Affect. Disord.* 130, 53–59. doi: 10.1016/j.jad.2010.09.021

- Weibel, S., Micoulaud-Franchi, J. A., Brandejsky, L., Lopez, R., Prada, P., Nicastro, R., et al. (2017). Psychometric properties and factor structure of the short form of the Affective Lability Scale in adult patients with ADHD. *J. Atten. Disord.* doi: 10.1177/1087054717690808 [Epub ahead of print].
- Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., and Kraft, M. (1993). A comparison of three structural models for personality: the big three, the big five, and the alternative five. *J. Pers. Soc. Psychol.* 65, 757–768. doi: 10.1037/0022-3514.65.4.757
- Zumbo, B. D., Gadermann, A. M., and Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for likert rating scales. *J. Mod. Appl. Stat. Methods* 6, 21–29. doi: 10.22237/jmasm/1177992180

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

The reviewer EP and handling Editor declared their shared affiliation.

Copyright © 2018 Contardi, Imperatori, Amati, Balsamo and Innamorati. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Psychometric Properties of the Cognitive Emotion Regulation Questionnaire (CERQ) in Patients with Fibromyalgia Syndrome

Albert Feliu-Soler^{1,2,3}, Elvira Reche-Camba⁴, Xavier Borràs⁴, Adrián Pérez-Aranda^{1,2,3}, Laura Andrés-Rodríguez^{1,2,3}, María T. Peñarrubia-María^{5,6}, Mayte Navarro-Gil⁷, Javier García-Campayo^{3,8}, Juan A. Bellón^{3,9,10} and Juan V. Luciano^{1,2,3*}

OPEN ACCESS

Edited by:

Marco Innamorati,
Università Europea di Roma, Italy

Reviewed by:

Dejan Stevanovic,
General Hospital Dr. Radivoj
Simonovic Sombor, Serbia
Federica Andrei,
Università di Bologna, Italy
Marco Lauriola,
Sapienza Università di Roma, Italy

*Correspondence:

Juan V. Luciano
jvluciano@pssjd.org

Specialty section:

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

Received: 29 September 2017

Accepted: 14 November 2017

Published: 13 December 2017

Citation:

Feliu-Soler A, Reche-Camba E, Borràs X, Pérez-Aranda A, Andrés-Rodríguez L, Peñarrubia-María MT, Navarro-Gil M, García-Campayo J, Bellón JA and Luciano JV (2017) Psychometric Properties of the Cognitive Emotion Regulation Questionnaire (CERQ) in Patients with Fibromyalgia Syndrome. *Front. Psychol.* 8:2075. doi: 10.3389/fpsyg.2017.02075

¹ Institut de Recerca Sant Joan de Déu, Barcelona, Spain, ² Teaching, Research & Innovation Unit, Parc Sanitari Sant Joan de Déu, St. Boi de Llobregat, Spain, ³ Primary Care Prevention and Health Promotion Research Network, RedIAPP, Madrid, Spain, ⁴ Facultat de Psicologia, Universitat Autònoma de Barcelona, Barcelona, Spain, ⁵ Primary Health Centre Bartomeu Fabrés Anglada, DAP Costa de Ponent, Institut Català de la Salut- IDIAP Jordi Gol, Gavà, Spain, ⁶ Centre for Biomedical Research in Epidemiology and Public Health CIBERESP, Madrid, Spain, ⁷ Faculty of Psychology, University of Zaragoza, Zaragoza, Spain, ⁸ Instituto de Investigaciones Sanitarias, Psychiatry Service, Hospital Universitario Miguel Servet, Zaragoza, Spain, ⁹ Primary Care Center El Palo, Málaga, Spain, ¹⁰ Department of Preventive Medicine, Public Health and Psychiatry, University of Málaga, Málaga, Spain

Given that Fibromyalgia Syndrome (FMS) is associated with problems in emotion regulation, the importance of assessing this construct is widely acknowledged by clinical psychologists and pain specialists. Although the Cognitive Emotion Regulation Questionnaire (CERQ) is a self-report measure used worldwide, there are no data on its psychometric properties in patients with FMS. This study analyzed the dimensionality, reliability, and validity of the CERQ in a sample of 231 patients with FMS. Given that “fibrofog” is one of the most disabling FMS symptoms, in the present study, items in the CERQ were grouped by dimension. This change in item presentation was conceived as an efficient way of facilitating responses as a result of a clear understanding of what the items related to each dimension are attempting to measure. The following battery of measures was administered: the CERQ, the Revised Fibromyalgia Impact Questionnaire, the Pain Catastrophizing Scale, the Center for Epidemiologic Studies Depression Scale, and the State-Trait Anxiety Inventory. Four models of the CERQ structure were examined and confirmatory factor analyses supported the original factor model, consisting of nine factors—Self-blame, Acceptance, Rumination, Positive refocusing, Refocus on planning, Positive reappraisal, Putting into perspective, Catastrophizing, and Other-blame. There was minimal overlap between CERQ subscales and their internal consistency was adequate. Correlational and regression analyses supported the construct validity of the CERQ. Our findings indicate that the CERQ (items-grouped version) is a sound instrument for assessing cognitive emotion regulation in patients with FMS.

Keywords: cognitive emotion regulation questionnaire (cerq), fibromyalgia, pain, depression, confirmatory factor analysis

INTRODUCTION

Chronic pain conditions, such as fibromyalgia syndrome (FMS), are amongst the most common health problems managed by general practitioners, rheumatologists, and clinical psychologists (Häuser et al., 2015). FMS is characterized by multifocal pain, fatigue, non-restorative sleep, cognitive complaints (also known as *fibrofog*: lack of attention-concentration, decrease in memory, and loss of vocabulary, which are exacerbated in stressful situations), high levels of distress, and is associated with greater affect intensity, which in turn correlates with more pain and fatigue in those patients with deficient emotion processing skills (van Middendorp et al., 2008; Geenen et al., 2012). Emotion regulation refers to “the extrinsic and intrinsic processes responsible for monitoring, evaluating and modifying emotional reactions, especially their intensive and temporal features, to accomplish one’s goals” (Thompson, 1994, p. 27). According to van Middendorp et al. (2008), the strategies to regulate unpleasant emotions such as sadness or anger play an important role in the maintenance or exacerbation of FMS symptoms. Moreover, impaired emotion regulation is a transdiagnostic risk factor that has been implicated in many disorders, including those related to mood, anxiety, substance use, personality, and eating (Naragon-Gainey et al., 2017). Emotion regulation strategies have been incorporated into some models of psychopathology and various therapeutic approaches (Aldao et al., 2010). For instance, Catastrophizing is a critically important risk factor for adverse pain-related outcomes and is directly associated with amplification of pain processing in the brain, whereas Reappraisal has a beneficial impact on an individual’s emotional state. In the long-term, it reduces chronic arousal of the hypothalamic-pituitary-adrenal axis (Edwards et al., 2009; Malfliet et al., 2017).

Hence, the availability of conceptually and psychometrically sound measures of *emotion reactivity* (how readily one experiences an emotion, how intensely, and for how long) and *emotion regulation* is an important component in the comprehensive assessment of patients in clinical research and practice (Zelkowitz and Cole, 2016). Focusing on the self-regulatory, conscious, cognitive components of emotion regulation, Garnefski et al. (2001) developed the Cognitive Emotion Regulation Questionnaire (CERQ). The authors revised existing measures to take out or reformulate items capturing cognitive dimensions, to transform non-cognitive coping strategies into cognitive dimensions, and to add new strategies taking into account rational grounds. The CERQ is a 36-item self-report measure that captures stable-dispositional cognitive emotion regulation strategies when people experience stressful or threatening life experiences. Specifically, the following strategies are measured: Self-blame, Blaming others, Acceptance, Refocusing on planning, Positive refocusing, Rumination, Positive reappraisal, Putting into perspective, and Catastrophizing. *Self-blame and Blaming others* are the cognitive strategies which refer to causal attribution of the negative event to oneself or the others; *Rumination* consists in overthinking about the consequences of the negative event; *Catastrophizing* is described as anticipating thoughts about

exaggerated consequences of the negative event; on the other hand, *Putting into perspective* refers to relativizing the unpleasant event by comparing it to others or considering its impact over time; *Positive refocusing* consists of trying to keep the attention on pleasant thoughts after the occurrence of a negative situation; *Positive reappraisal*, is the strategy by which the individual tries to find the silver lining in the negative event; *Acceptance* refers to the cognitive process by which the individual stops trying to change the negative situation or the emotions that it caused and just experiences them; finally, *Planning* is described as the strategy by which the attention is focused on what the individual can do to solve the unpleasant situation or make it easier to deal with. A detailed explanation of the cognitive strategies is provided in the pioneer study by Garnefski et al. (2001).

When adults from a clinical sample with clinically relevant depression and anxiety, and subjects from a matched non-clinical sample both completed the CERQ, Garnefski et al. (2002) found Cronbach’s α values that ranged from 0.72 (Acceptance) to 0.85 (Self-blame). For cognitive research to remain linked to clinical practice, it is crucial for instruments to perform well in both clinical and non-clinical samples. Garnefski et al. (2002) found significant differences between the clinical and the non-clinical groups in Catastrophizing, Self-blame, Rumination, Other-blame, Positive reappraisal, and Acceptance. Of these strategies, only Positive reappraisal appeared to be reported significantly more often by the non-clinical group than by the clinical group. Garnefski and Kraaij (2006) compared early adolescent, late adolescent, adult, elderly and psychiatric samples on their reported use of cognitive emotion regulation strategies. As expected, data analyses revealed significantly higher scores for Self-blame, Rumination, Catastrophizing and Other-blame in the adult psychiatric sample, supporting the construct validity of the CERQ. In another study, Garnefski and Kraaij (2007) reported adequate goodness-of fit values for the nine-factor model (CFI = 0.92 and 0.97 in two different time points), which confirmed the robustness of the CERQ factor structure.

The CERQ has been translated and validated into many languages and cultures, such as French (Jermann et al., 2006), Chinese (Zhu et al., 2008), Turkish (Tuna and Bozo, 2012), Persian (Abdi et al., 2012), Spanish (Domínguez-Sánchez et al., 2013; Medrano et al., 2013; Domínguez-Lara and Medrano, 2016) and Arabic (Megreya et al., 2016), showing adequate reliability and validity. A recent cross-cultural study (Potthoff et al., 2016) compared CERQ scores across six European countries (Netherlands, Hungary, Spain, Italy, Portugal, and Germany) using general population samples, all comparable in terms of age and educational backgrounds. Although some between-country differences were observed in subscale scores, there was a consistent link between cognitive emotion regulation strategies and psychopathology. More recently, Ireland et al. (2017) examined the dimensionality, and construct validity of the CERQ, both short (18 items) and long (36 items) form, in 795 community residents evaluated online. Although model fit was better for the 18-item CERQ, the correlational analyses with difficulties in emotion regulation and positive/negative affect values indicated a statistically significant small to medium drop in variance explained by the CERQ-short when

compared with the full CERQ, which suggests better convergent validity for the full version of the instrument. To sum up, the CERQ seems to be an optimal candidate for the assessment of emotion regulation in clinical and non-clinical samples.

To date, none of the published studies on the CERQ has examined the psychometric properties of the instrument in patients with FMS. Verification of the original nine-factor model, as well as of adequate reliability and validity in these patients, is lacking. Taking this as its foundation, the present study examines the internal consistency and convergent-discriminant validity of the Spanish CERQ and evaluates its dimensionality using confirmatory factor analysis (CFA) in a pooled sample of patients with FMS. In line with previous studies, a nine-factor solution in addition to unidimensional and hierarchical factor solutions were tested. We expected that the original nine-factor model would provide the best fit. Second, the internal consistency (Cronbach's α) of the best fitting factor structure of the CERQ was determined. Third, construct validity (convergent validity) of the best fitting factor structure of the CERQ was assessed by investigating the relationships with self-report measures of psychological symptoms (anxiety and depression) and pain-related constructs such as pain catastrophizing and functional status in FMS. Given that depression is a disorder characterized by impaired emotion regulation (Joormann and Stanton, 2016), we compared the CERQ scores of subgroups of FMS patients with distinct levels of depressive symptoms to establish the discriminant validity of the CERQ.

MATERIALS AND METHODS

In the present study, we utilized the dataset from the Fibromyalgia Subtypes study (Luciano et al., 2016) and early-stage data from the EUDAIMON study (Feliu-Soler et al., 2016). Study data are available from the corresponding author. Written informed consent was obtained from patients of both studies. **Table 1** displays participant characteristics for the two samples.

Sample 1 (Fibromyalgia Subtypes study) consisted of a convenience sample of 160 adult patients with FMS recruited from 14 physician practices within the Barcelona metropolitan area (Spain). The family physicians at these centers referred suspected FMS cases to Viladecans Hospital or Sant Joan de Déu Hospital (the two reference hospitals in the area). Rheumatologists from these hospitals confirmed or ruled out the diagnosis of FMS following American College of Rheumatology (ACR) 1990 criteria (Wolfe et al., 1990), and added the patients to a database if they received a FMS diagnosis. Adult patients (≥ 18 years-old) in these databases were candidates for inclusion in the study. A detailed description of the study protocol and inclusion/exclusion criteria can be found elsewhere (Luciano et al., 2016). The study protocol was approved by the Ethics Committee at the Sant Joan de Déu Foundation (CEIC PIC-33-11; Esplugues de Llobregat, Spain) and by the Jordi Gol i Gurina Foundation research ethics committee (P12/94; Barcelona, Spain).

TABLE 1 | Participant Characteristics for the Two Samples and the Entire Sample.

Socio-demographic variables	Sample 1 (n = 160)	Sample 2 (n = 71)	Total sample (n = 231)
Gender (n females, %)	156 (97.5)	71 (100)	227 (98.3)
Age, M (SD)	57.28 (8.8)	52.63 (7.2)	55.89 (8.6)
Marital status, n (%)			
Single	5 (3.1)	3 (4.2)	8 (3.5)
Married/Living with a partner	118 (73.8)	57 (80.3)	175 (76.1)
Separated/divorced	20 (12.5)	9 (12.7)	29 (12.6)
Widowed	17 (10.6)	1 (1.4)	18 (7.8)
Living arrangements, n (%)			
Living alone	18 (11.3)	1 (1.4)	19 (8.2)
Living with someone (spouse/partner/relatives)	142 (88.8)	69 (97.2)	211 (91.3)
Educational level, n (%)			
No formal education	33 (20.6)	1 (1.4)	33 (14.3)
Did not graduate from primary school	30 (18.8)	2 (2.8)	32 (13.9)
Primary school	56 (35)	36 (50.7)	92 (39.8)
Secondary school	35 (21.9)	30 (42.3)	65 (28.1)
University	6 (3.8)	2 (2.8)	8 (3.5)
Work status, n (%)			
Homemaker	40 (25)	9 (12.9)	49 (21.3)
Paid employment	25 (15.6)	23 (32.9)	48 (20.8)
Paid employment but on sick leave	7 (4.4)	6 (8.6)	13 (5.7)
Unemployed with allowance	25 (15.6)	6 (8.6)	31 (13.4)
Unemployed without allowance	14 (8.8)	11 (15.7)	25 (10.8)
Retired/pensioner	25 (15.6)	7 (10)	32 (13.9)
Temporarily disabled	–	1 (1.4)	1 (0.4)
Others (e.g., student)	24 (15)	7 (10)	31 (13.8)
Clinical variables, M (SD)			
FIQ-R (0-100)	68.90 (18.87)	59.41 (21.23)	65.99 (20.07)
Function (0-30)	20.54 (6.67)	18.20 (6.50)	19.81 (6.70)
Overall impact (0-20)	12.79 (7.33)	9.29 (7.31)	11.71 (7.49)
Severity of symptoms (0-50)	35.58 (8.07)	31.91 (9.83)	34.45 (8.80)
PCS (0-52)	31.47 (14.06)	21.63 (13.25)	28.48 (14.52)
Rumination (0-16)	10.36 (4.82)	7.70 (4.93)	9.55 (4.98)
Magnification (0-12)	6.49 (3.43)	4.23 (2.92)	5.80 (3.44)
Helplessness (0-24)	14.63 (7.08)	9.70 (6.70)	13.13 (7.32)
CES-D (0-60)	34.34 (11.79)	–	–
STAI-T (0-60)	37.50 (10.56)	–	–

CES-D, Center for Epidemiologic Studies Depression Scale; FIQ-R, Fibromyalgia Impact Questionnaire Revised; PCS, Pain Catastrophizing Scale; STAI-T, Trait Anxiety Inventory.

Sample 2 consisted of 71 patients with FMS recruited for the EUDAIMON study. This ongoing study is a 12-month, randomized controlled trial, the main aim of which is to assess the effectiveness and cost-utility of a mindfulness-based intervention for FMS patients compared with a psycho-educational intervention (FibroQoL) and treatment as usual. For the present work, we used only the EUDAIMON baseline dataset. Patients were selected following a multi-stage recruitment process. All recruited patients are adults diagnosed with FMS according to the ACR 1990 by rheumatologists from the

Sant Joan de Déu Hospital. A detailed description of the study protocol and inclusion/exclusion criteria can be found elsewhere (Feliu-Soler et al., 2016). The RCT is being performed in accordance with ethical standards laid down in the 1964 Declaration of Helsinki and its subsequent updates. The Ethics Committee at the Sant Joan de Déu Foundation evaluated and approved the study protocol in May 2015 (PIC-102-15).

Procedure

In both studies (Feliu-Soler et al., 2016; Luciano et al., 2016), a randomized list of potential participants was delivered to a research assistant (health psychologist) who screened patients through a phone interview until the targeted sample size was achieved. The research assistant then made an appointment for those patients that agreed to participate in the study. In the Fibromyalgia Subtypes study (Luciano et al., 2016), the research assistant performed all the face-to-face interviews in the general practices or in the reference hospitals once written consent had been obtained, whereas in the EUDAIMON study (Feliu-Soler et al., 2016), the CERQ was completed at home and collected by the research assistant (blind to group allocation) on the participants' following visit to the hospital (1–2 weeks later).

Study Measures

Participants from both studies completed the following paper-and-pencil measures:

The Socio-Demographic questionnaire collected information on the following variables: gender, date of birth, marital status, living arrangements, educational level, employment status, type of contract (question for employees), and years since FMS diagnosis.

The *Cognitive Emotion Regulation Questionnaire* (CERQ; Garnefski et al., 2001) is a 36-item self-report measure designed to assess individual differences in cognitive regulation of emotions in response to stressful, threatening or traumatic life events. The instrument assesses nine 4-item dimensions: Self-blame, Blaming others, Acceptance, Refocusing on planning, Positive refocusing, Rumination, Positive reappraisal, Putting into perspective, and Catastrophizing. Responses are given on a 5-point Likert scale ranging from 1 “(almost) never” to “(almost) always.” Therefore, subscale scores can range from 4 to 20 with higher subscale scores indicating greater frequency of use of the specific cognitive strategy. Regarding the Spanish version, it was tested in a large non-clinical sample ($n = 615$ students) by Domínguez-Sánchez et al. (2013), who obtained a hierarchical structure composed of nine dimensions distributed into two second-order factors (adaptive strategies and less adaptive strategies). The internal consistency, test-retest reliability and criterion validity were adequate or acceptable. A characteristic of the CERQ, in common with most multidimensional instruments, is that items are not grouped by dimension, but are dispersed throughout the instrument. Specifically, the questionnaire developers chose a rotating selection strategy, so that every ninth item is presupposed to belong to the same dimension. For instance, items 1, 10, 19, and 28 are considered to belong to Self-blame. Given that *fibrofog* is one of the most prominent FMS symptoms, in this study, items in the CERQ were grouped (but

not labeled) by dimension. This change in item presentation was conceived as an efficient way of facilitating responses as a result of a clear understanding of what the items related to each dimension are attempting to measure (Schell and Oswald, 2013). Thus, we expected to have an instrument perfectly aligned with our target sample that could provide more trustworthy information about emotion regulation with the confidence that there is available empirical evidence that item order, within honest conditions (when faking is not presupposed), does not alter the underlying measurement properties of psychological instruments (Schell and Oswald, 2013).

The *Revised Fibromyalgia Impact Questionnaire* (FIQR; Bennett et al., 2009; Luciano et al., 2013) is the recommended instrument for measuring functional status in FMS patients. It includes 21 items that are all answered on an 11-point numeric rating scale of 0-to-10, with 10 reflecting greater impairment. The time frame is the previous 7 days, with the items distributed across three associated domains: “function” (9 items); “overall impact” (2 items); and “severity of symptoms” (10 items). The scoring system is as follows: the physical function domain (0-to-90) is divided by 3, the overall impact domain (0-to-20) is not transformed, and the severity of symptoms domain (0-to-100) is divided by 2. FIQR reliability in our pooled sample was good (Cronbach's $\alpha = 0.89$).

The *Pain Catastrophizing Scale* (PCS; Sullivan et al., 1995; García-Campayo et al., 2008) is a 13-item instrument that consists of 3 dimensions: Rumination (tendency to focus excessively on pain sensations), Magnification (tendency to magnify the threat value of pain sensations), and Helplessness (tendency to perceive oneself as unable to control the intensity of pain). The PCS total score and subscale scores are computed as the algebraic sum of ratings for each item. PCS items are rated in relation to the frequency of occurrence on 5-point scales (0 = never, 4 = almost always), and total scores can vary from 0 to 52. Higher scores indicate greater pain catastrophizing. Internal consistency was excellent in the pooled sample (Cronbach's $\alpha = 0.94$).

In addition, the participants from the Fibromyalgia Subtypes study completed the following paper-and-pencil measures:

The *Center for Epidemiologic Studies Depression Scale* (CES-D; Radloff, 1977; Vázquez et al., 2007) is a 20-item scale frequently used to assess depressive symptom severity. The time frame is the previous week. Item responses range from 0 to 3 [0 = rarely or none of the time (<1 day in the past week), 1 = some or a little of the time (1–2 days), 2 = occasionally or a moderate amount of the time (3–4 days), and 3 = most or all of the time (5–7 days)]. Therefore, total scores can vary from 0 to 60, with higher scores reflecting increased depression severity. The CES-D has been widely used to detect mood disturbances in many populations, including FMS patients, demonstrating adequate psychometric properties (Smarr and Keefer, 2011). A recent meta-analysis (Vilagut et al., 2016) focused on CES-D screening accuracy for depression observed that a cut-off score ≥ 20 yielded the best trade-off between sensitivity (0.83) and specificity (0.78). The CES-D had high internal consistency (Cronbach's $\alpha = 0.86$).

The *Spanish State-Trait Anxiety Inventory* (STAI—form X; Spielberger et al., 1986) is a 40-item, self-report measure of general anxiety. The first 20 items (STAI-S) measure state anxiety,

or how the subject feels right now. The second 20 items (STAI-T) assess trait anxiety, or how the subject generally feels. We only used the STAI-T. Individuals have to rate each item using a Likert-type scale from 0 (not at all) to 3 (very much so). Total scores on the STAI-T vary from 0 to 60, with higher scores indicating more trait anxiety. Cronbach's α for the STAI-T was 0.84.

Statistical Analyses

SPSS v22.0 and MPlus v7.4 were used to compute the data analyses.

First, we conducted a CFA to test the fit of the following factor structures: the one-factor model with all CERQ items loading on one latent factor, the original nine-factor model by Garnefski et al. (2001) with Self-blame, Other-blame, Catastrophizing, Rumination, Acceptance, Positive refocusing, Refocus on planning, Positive reappraisal, and Putting into perspective. Finally, we tested the higher order factor model reported by Domínguez-Sánchez et al. (2013) with the nine dimensions grouped into two general latent dimensions of *adaptive strategies* (Acceptance, Positive refocusing, Refocus on planning, Positive reappraisal, Putting into perspective) and *less adaptive strategies* (Self-blame, Rumination, Catastrophizing, and Other-blame). In ordinal items with a non-normal distribution, such as those in the CERQ, it may be expected that the covariance matrix will underestimate the true extent of relationships among items. Therefore, we proceeded to estimate the models from the polychoric correlation matrix. Mean and Variance corrected Weighted Least Squares (WLSMV) was applied to test the fit of the three factor models. The following indices were examined to evaluate model fit: χ^2 (a non-significant estimate reflects good fit), the Tucker-Lewis Index (TLI ≥ 0.90), the comparative fit index (CFI ≥ 0.90), and the root means square error of approximation (RMSEA ≤ 0.08).

Second, we calculated the internal consistency for each CERQ domain by computing Cronbach's α in the pooled sample. A common rule of thumb criterion is a Cronbach's α of 0.6 for exploratory research and of 0.7 for confirmatory research (Hair et al., 1998). In addition, we assessed homogeneity of the CERQ subscales by inspecting the corrected item total correlation (correlation of the designated item with the total score for all other subscale items). A cut-off score of 0.3 is recommended for the corrected item-total correlations (Nunnally and Bernstein, 1994).

Third, we examined the correlations among the CERQ subscales as well as their construct validity by computing Pearson product moment correlations between each of the CERQ subscales with the measures of functional status (FIQR), pain catastrophizing (PCS), depressive symptoms (CES-D), and trait anxiety (STAI-T). We took Cohen (1988) into account to evaluate the substantive significance of correlations (large correlations are those >0.5 , medium correlations are from 0.3 to 0.49, and small correlations are from 0.1 to 0.29).

Finally, the known-groups' validity approach is founded on the hypothesis that specific subgroups of patients might be expected to score differently from others. In this study, a set of Student's *t*-tests for independent samples was computed to assess

the validity of the CERQ subscales to discriminate between the FMS patients with clinically relevant depressive symptoms and those without (according to the CES-D cut-off value ≥ 20 ; Vilagut et al., 2016). We calculated between-groups effect sizes using Cohen's *d* with a 95% confidence interval. The rule of thumb for Cohen's *d* is that 0.2 is small, 0.5 is medium, and 0.8 is large. Additionally, bearing in mind that the separate cognitive emotion regulation strategies have overlapping processes and due to the likely significant subscale intercorrelations, multivariate analyses accounting for the intercorrelations are needed to identify unique relationships between cognitive emotion regulation strategies and clinical subgroup membership (FMS with vs. without depression). Therefore, we computed a logistic regression analysis to examine the unique "influence" of each strategy on subgroup membership, while controlling for the influence of the other strategies (Garnefski et al., 2002). In this analysis, the binary dependent variable was subgroup membership (FMS with vs. without depression), whereas the independent variable set consisted of the nine cognitive emotion regulation strategies.

RESULTS

Testing Competing Confirmatory Factor Analytic CERQ Models

In the CFA involving the one-factor model, we found that it provided a very poor fit to the sample data: $\chi^2_{(594, N=229)} = 5,564.958$, $p < 0.001$, CFI = 0.527, TLI = 0.498, and RMSEA = 0.191 (90% CI, 0.187–0.196). Consistent with Garnefski et al. (2001), a nine-factor model adequately fit the data, $\chi^2_{(558, N=229)} = 1,302.203$, $p < 0.001$, CFI = 0.929, TLI = 0.920, and RMSEA = 0.076 (90% CI, 0.071–0.082). Standardized factor loadings for the nine-factor model were all statistically significant and ranged from 0.542 (item 29) to 0.957 (item 34). See **Table 2** for standardized factor loading estimates. For the sake of comparability, **Table 2** also shows factor loadings reported by Garnefski and Kraaij (2007) in a sample of 611 Dutch adults from the general population and by Domínguez-Sánchez et al. (2013) in 615 Spanish students.

The hierarchical factor model revealed that the inclusion of two second-order factors (adaptive and less adaptive strategies) produced a worse fit to the data compared to the nine-factor model, $\chi^2_{(584, N=229)} = 1,519.054$, $p < 0.001$, CFI = 0.911, TLI = 0.904, and RMSEA = 0.084 (90% CI, 0.078–0.089). One of the reasons for the worse fit was the low factor loading ($\lambda = 0.135$, $p = 0.044$) of Acceptance with the second-order factor labeled as adaptive strategies. Therefore, we tested a respecification of the second-order factor model that incorporated Acceptance on the latent factor labeled as less adaptive strategies. This hierarchical model showed a slightly better fit across all indices, compared with the previously estimated hierarchical model $\chi^2_{(584, N=229)} = 1,462.583$, $p < 0.001$, CFI = 0.916, TLI = 0.910, and RMSEA = 0.081 (90% CI, 0.076–0.086). The Acceptance dimension was more strongly related to the less adaptive strategies latent factor ($\lambda = 0.287$, $p < 0.001$) than with the adaptive strategies factor. For illustrative purposes, the second hierarchical model is displayed in **Figure 1**. Therefore, we decided to retain the

TABLE 2 | Item Content, Mean (M), Standard Deviation (SD), and Factor Loadings (λ , 9-factor solution) of the CERQ Items.

Scale names (Cronbach α values) and items	Sample 1 M (SD)	Sample 2 M (SD)	Total sample M (SD)	λ	Domínguez-Sánchez et al. (2013) λ n= 615	Garnefski and Kraaij (2007) λ T1 / T2 n= 611
Self-blame ($\alpha = 0.86$)						
1. I feel that I am the one to blame for it (1)	2.33 (1.27)	2.17 (1.19)	2.33 (1.26)	0.89	0.79	0.70/0.70
2. I feel that I am the one who is responsible for what has happened (10)	2.23 (1.21)	2.14 (1.20)	2.23 (1.20)	0.92	0.68	0.71/0.70
3. I think about the mistakes I have made in this matter (19)	2.98 (1.34)	3.06 (1.28)	2.98 (1.33)	0.73	-0.11	0.55/0.57
4. I think that basically the cause must lie within myself (28)	2.26 (1.24)	2.27 (1.14)	2.26 (1.24)	0.82	0.69	0.80/0.77
Acceptance ($\alpha = 0.77$)						
5. I think that I have to accept that this has happened (2)	3.36 (1.35)	3.21 (1.29)	3.36 (1.35)	0.93	0.72	0.73/0.77
6. I think that I have to accept the situation (11)	3.47 (1.35)	3.46 (1.27)	3.47 (1.35)	0.96	0.87	0.70/0.71
7. I think that I cannot change anything about it (20)	3.13 (1.43)	2.91 (1.34)	3.13 (1.43)	0.87	-0.10	0.66/0.65
8. I think that I must learn to live with it (29)	3.80 (1.30)	3.51 (1.33)	3.80 (1.29)	0.91	0.61	0.69/0.61
Rumination ($\alpha = 0.84$)						
9. I often think about how I feel about what I have experienced (3)	3.23 (1.35)	2.97 (1.32)	3.23 (1.34)	0.80	0.71	0.75/0.66
10. I am preoccupied with what I think and feel about what I have experienced (12)	3.19 (1.41)	3.21 (1.35)	3.19 (1.40)	0.82	0.74	0.77/0.74
11. I want to understand why I feel the way I do about what I have experienced (21)	3.23 (1.40)	3.24 (1.43)	3.23 (1.39)	0.70	0.52	0.66/0.69
12. I dwell upon the feelings the situation has evoked in me (30)	3.36 (1.40)	3.25 (1.42)	3.36 (1.40)	0.88	0.75	0.68/0.77
Positive refocusing ($\alpha = 0.93$)						
13. I think of nicer things than what I have experienced (4)	2.61 (1.40)	2.83 (1.23)	2.61 (1.40)	0.86	0.83	0.76/0.79
14. I think of pleasant things that have nothing to do with it (13)	2.60 (1.48)	2.94 (1.42)	2.60 (1.47)	0.95	0.86	0.85/0.87
15. I think of something nice instead of what has happened (22)	2.27 (1.28)	2.46 (1.23)	2.27 (1.27)	0.93	0.85	0.83/0.80
16. I think about pleasant experiences (31)	2.40 (1.34)	2.82 (1.36)	2.40 (1.33)	0.95	0.91	0.67/0.74
Refocus on planning ($\alpha = 0.83$)						
17. I think of what I can do best (5)	3.37 (1.26)	3.46 (1.24)	3.37 (1.25)	0.76	0.72	0.69/0.81
18. I think about how I can best cope with the situation (14)	3.30 (1.24)	3.45 (1.08)	3.30 (1.23)	0.84	0.84	0.75/0.80
19. I think about how to change the situation (23)	3.23 (1.31)	3.15 (1.28)	3.23 (1.30)	0.75	0.71	0.74/0.71
20. I think about a plan of what I can do best (32)	3.12 (1.37)	3.08 (1.28)	3.12 (1.37)	0.82	0.81	0.78/0.77
Positive reappraisal ($\alpha = 0.80$)						
21. I think I can learn something from the situation (6)	3.10 (1.41)	2.89 (1.30)	3.10 (1.40)	0.74	0.83	0.67/0.72
22. I think that I can become a stronger person as a result of what has happened (15)	2.74 (1.44)	2.65 (1.36)	2.74 (1.44)	0.71	0.81	0.59/0.59
23. I think that the situation also has its positive sides (24)	2.65 (1.44)	2.46 (1.41)	2.65 (1.44)	0.78	0.79	0.64/0.52
24. I look for the positive sides to the matter (33)	2.85 (1.47)	2.63 (1.40)	2.85 (1.46)	0.86	0.94	0.73/0.70
Putting into perspective ($\alpha = 0.79$)						
25. I think that it all could have been much worse (7)	3.10 (1.36)	2.89 (1.35)	3.10 (1.35)	0.59	0.68	0.62/0.60
26. I think that other people go through much worse experiences (16)	3.64 (1.37)	3.14 (1.42)	3.64 (1.36)	0.82	0.80	0.77/0.79
27. I think that it hasn't been too bad compared to other things (25)	2.96 (1.31)	2.91 (1.26)	2.96 (1.31)	0.82	0.87	0.68/0.79
28. I tell myself that there are worse things in life (34)	3.48 (1.37)	3.24 (1.34)	3.48 (1.37)	0.80	0.81	0.70/0.80
Catastrophizing ($\alpha = 0.82$)						
29. I often think that what I have experienced is much worse than what others have experienced (8)	2.16 (1.27)	2.03 (1.08)	2.16 (1.26)	0.55	0.46	0.75/0.34
30. I keep thinking about how terrible it is what I have experienced (17)	2.39 (1.30)	2.30 (1.19)	2.39 (1.30)	0.85	0.87	0.64/0.75
31. I often think that what I have experienced is the worst that can happen to a person (26)	2.07 (1.26)	1.86 (1.14)	2.07 (1.26)	0.79	0.63	0.70/0.80
32. I continually think how horrible the situation has been (35)	2.42 (1.32)	2.13 (1.13)	2.42 (1.31)	0.91	0.81	0.59/0.78
Other-blame ($\alpha = 0.92$)						
33. I feel that others are to blame for it (9)	1.77 (1.24)	1.51 (0.95)	1.77 (1.24)	0.93	0.81	0.75/0.71
34. I feel that others are responsible for what has happened (18)	1.80 (1.20)	1.52 (0.89)	1.80 (1.19)	0.96	0.79	0.82/0.79
35. I think about the mistakes others have made in this matter (27)	2.22 (1.31)	1.81 (1.08)	2.22 (1.30)	0.87	0.50	0.72/0.72
36. I feel that basically the cause lies with others (36)	1.82 (1.27)	1.54 (1.00)	1.82 (1.27)	0.91	0.87	0.83/0.81

Original item numbering is presented between brackets. T1 = Time 1; T2 = Time 2.

nine CERQ domains for further analyses (reliability and validity) given that, among the tested models, the first-order nine-factor model showed the best fit to the data and because of parsimony considerations¹.

Reliability and Homogeneity of the CERQ Subscales

As can be seen in **Table 2**, Cronbach's α reliability scores for the CERQ subscales in FMS patients ranged from 0.77 (Acceptance) to 0.93 (Positive refocusing) and the values of the corrected item-total correlations ranged from 0.44 (item 25) to 0.87 (items 14, 16, and 34). The average corrected item-total correlation was $r = 0.7$ (Self-blame), 0.58 (Acceptance), 0.67 (Focus on thoughts), 0.85 (Positive refocusing), 0.67 (Refocus on planning), 0.61 (Positive reappraisal), 0.6 (Putting into perspective), 0.64 (Catastrophizing), and 0.81 (Other-blame). Squaring that value shows that 49, 34, 45, 72, 45, 37, 36, and 41% of the variance of the average item overlaps with the remaining subscale items, respectively.

Intercorrelations among the CERQ Subscales

As displayed in **Table 3**, correlations among the CERQ subscales fell between non-significant (n.s) and one large value (0.54 for Self-blame and Rumination). Notably, half of the computed correlations (18/36) were not statistically significant. The majority of the significant relationships were small or medium in magnitude, suggesting that the subscales are relatively independent. Following Cohen's (1988) criteria to evaluate the substantive significance of correlations, the average size of the significant intercorrelations found among the adaptive and less adaptive subscales was medium in both cases ($r = 0.38$ and 0.33 , respectively).

Convergent Validity: Association of the CERQ Subscales with Study Measures

The results are shown in **Table 4**. On the one hand, it is interesting to note that Acceptance presented significant, positive, small correlations with the CES-D and STAI-T and the FMS-related measures (FIQR and PCS) as well, which supported the second-order factor model reported above. In a similar vein, the other less adaptive strategies (Self-blame, Rumination, Catastrophizing, and Other-blame) showed a significant pattern of positive correlations with the study measures. On the other hand, two of the adaptive CERQ strategies (Refocus on planning and Putting into perspective) presented null correlations with the

study measures. Only Positive refocusing and Positive reappraisal presented the expected significant negative relationships with trait anxiety, depression symptoms, functional impairment and pain catastrophizing. All these correlations were of small magnitude with the exception of those obtained by Positive refocusing with depressive symptoms and trait anxiety, which were medium-to-large.

Discriminant Validity: Differences in Cognitive Emotion Regulation between FMS Patients with vs. without Clinically Significant Depression

More than three-quarters of our participants (84.4%) presented clinically relevant depressive symptoms. Student's t and χ^2 tests revealed that the two subgroups (FMS + depression vs. FMS) were fully comparable in their demographic characteristics (including duration of illness). As shown in **Table 5**, the FMS patients with clinically relevant depression scored significantly higher on the Self-blame, Rumination, Catastrophizing, and Other-Blame subscales than the FMS participants without depression. The differences in Positive refocusing and Positive reappraisal were also significant, but in the opposite direction. The significant differences oscillated from medium to large in magnitude according to Cohen's criteria. Some null differences were obtained. Specifically, those patients that were depressed did not differ from the non-depressed subgroup on the Acceptance, Refocus on planning, and Putting into perspective subscales. Overall, our data on the criterion-related validity of the CERQ subscales support the FMS-relevance of some of the measured cognitive emotion regulation strategies for discriminating among patients with/without affective comorbidity. Means and standard deviations of the CERQ scales are shown in **Table 5**. For the sake of comparability, **Table 5** also shows the descriptive CERQ data obtained in a sample of 615 Spanish students (Domínguez-Sánchez et al., 2013) and 99 Dutch patients with clinically relevant depression and anxiety (Garnefski et al., 2002). With the exception of Catastrophizing, it seems that FMS patients do not use the *a priori* less adaptive cognitive emotion regulation strategies (including Acceptance) more frequently when compared with non-clinical Spanish subjects. In contrast, with the exception of Putting into perspective, patients report having used the more adaptive strategies less often. Patients with FMS in our study that had clinically relevant depressive symptoms had similar CERQ subscale scores compared with patients referred for treatment at an outpatient psychiatric clinic in the Netherlands who had significant depressive and anxiety symptoms. These comparisons should be interpreted with caution due to the absence of statistical analyses and matching in relevant variables such as gender or age.

Finally, given that the two subgroups were almost identical in their sociodemographic characteristics, it was unnecessary to control for these variables in the subsequent logistic regression analysis. The regression model explained 24.9% of the total variance [$\chi^2_{(9)} = 45.88, p < 0.001$]. The Wald statistic was used to determine the significance of the contribution of the

¹As suggested by one anonymous reviewer, a bifactor structure was also fitted, examining whether the CERQ could be modeled using two general factors of 'adaptive' and 'less adaptive' strategies, as measured by a priori adaptive and less adaptive items, respectively and nine specific factors, as measured by item subsets. A bifactor approach (Rodríguez et al., 2016) helps to determine whether the CERQ items are multidimensional, allowing the computation of sub-scale scores, or whether the items are mainly unidimensional, for which only two total scores should be computed and reported (one total 'adaptive' score + one total 'less adaptive' score). Unfortunately, this model had estimation problems (empirically unidentified) that preclude its reporting in the manuscript as potential factor solution for the CERQ.

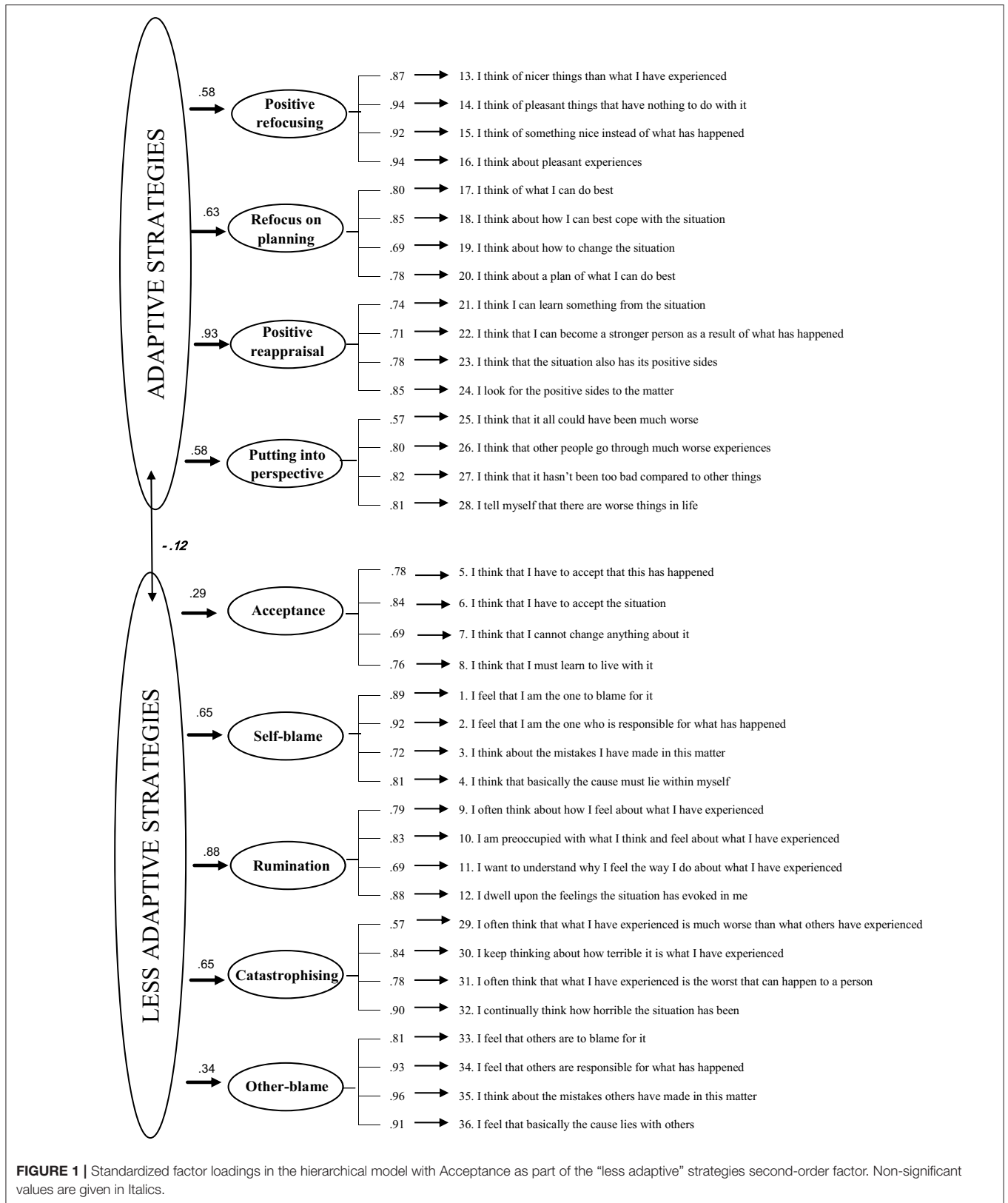


FIGURE 1 | Standardized factor loadings in the hierarchical model with Acceptance as part of the “less adaptive” strategies second-order factor. Non-significant values are given in *italics*.

TABLE 3 | Intercorrelations among the CERQ Subscales.

CERQ Subscales	SB	A	RUM	PR	RP	POSR	PP	CAT	OB
Self-Blame (SB)	–	0.25**	0.54**	–0.29**	0.17**	n.s.	n.s.	0.29**	n.s.
Acceptance (A)		–	0.28**	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Rumination (RUM)			–	–0.21**	0.26**	n.s.	n.s.	0.42**	0.18**
Positive refocusing (PR)				–	0.31**	0.41**	0.24**	–0.21**	n.s.
Refocus on planning (RP)					–	0.48**	0.35**	n.s.	n.s.
Positive reappraisal (POSR)						–	0.47**	n.s.	n.s.
Putting into perspective (PP)							–	n.s.	n.s.
Catastrophizing (CAT)								–	0.38**
Other-Blame (OB)									–

n.s., non-significant; * $p < 0.05$; ** $p < 0.01$.

TABLE 4 | Intercorrelations between the CERQ Subscales and Study Measures.

CERQ Subscales	FIQR	PCS	CES-D	STAI-T
Self-Blame	0.31**	0.35**	0.42**	0.45**
Acceptance	0.29**	0.21**	0.18*	0.21**
Rumination	0.34**	0.32**	0.39**	0.46**
Positive refocusing	–0.14*	–0.24**	–0.48**	–0.55**
Refocus on planning	n.s.	n.s.	n.s.	n.s.
Positive reappraisal	–0.14*	–0.13*	–0.26**	–0.32**
Putting into perspective	n.s.	n.s.	n.s.	n.s.
Catastrophizing	0.39**	0.41**	0.47**	0.43**
Other-Blame	0.14*	0.17**	0.19*	0.22**

CES-D, Center for Epidemiologic Studies Depression Scale; FIQR, Fibromyalgia Impact Questionnaire Revised; PCS, Pain Catastrophizing Scale; STAI-T, Trait Anxiety Inventory. n.s., non-significant * $p < 0.05$; ** $p < 0.01$.

independent variables and the standardized β to ascertain the relative influence of each independent variable. As can be seen in **Table 6**, only two cognitive emotion regulation strategies were independent predictors of subgroup membership: Positive refocusing (standardized $\beta = 0.13$) and Catastrophizing (standardized $\beta = 0.22$). Therefore, subgroup membership was related to higher reported use of Catastrophizing and lower reported use of Positive refocusing. A new logistic regression model was computed including the two significant predictors only. This model yielded a slightly lower percentage of total explained variance (16.9%), but both predictors remained significant.

DISCUSSION

The CFAs computed on the CERQ supported the original nine-factor model in a Spanish sample of adult patients with FMS. This factor solution best fit the data, which is consistent with previous published psychometric studies carried out in other countries. For instance, a sample of French-speaking, young community volunteers completed the CERQ in the study by Jermann et al. (2006). The principal component analysis (PCA) suggested extracting nine factors that explained 56.7% of the

variance and the CFA with the maximum likelihood (ML) method supported the nine-factor model (CFI = 0.94; RMSEA = 0.06). As with our study, the authors also tested a second-order factor model with adaptive and less adaptive strategies which provided good fit to the data. Similarly, Zhu et al. (2008) examined the dimensionality of the CERQ in Chinese university students performing a CFA with ML as the estimation method. The first-order nine-factor model fit the data well (CFI = 0.91, NNFI = 0.9, RMSEA = 0.05). More recently, Megreya et al. (2016) analyzed the psychometric properties of the Arabic version of the CERQ in four Arabic-speaking Middle Eastern countries (Egypt, Kingdom of Saudi Arabia, Kuwait, and Qatar). In line with our study, due to the ordinal nature of the items, the WLSMV estimator was used in the CFA. Overall, the goodness-of-fit indices indicated a good fit of the nine-factor model in the cases of Egypt, Kingdom of Saudi Arabia, and Qatar. The subsequent second-order CFA for each country, yielded poorer fit for the four countries in all indices compared with the first-order factor models. Therefore, the accumulated empirical evidence suggests that the first-order nine-factor structure is retained beyond the cultural context.

Inspection of the specific item-loadings is also in line with previous factor analytic studies performed on the CERQ. However, studies of different cultural versions of the CERQ have reported low or null standardized factor loadings for some items. For example, Domínguez-Sánchez et al. (2013) reported factor loadings of –0.11 for item 19 and –0.10 for item 20. Similarly, the PCA conducted by Jermann et al. (2006) indicated that the maximum loading of each CERQ item was found on the assigned factor, except for items 19 and 20. The saturation of item 8 on its factor was below 0.3. In contrast, we found that all 36 items could be retained taking common cut-off criteria for item retention into account. The lowest factor loading was 0.55 in the present work (item 8 in the original form). In our opinion, the main reason of this increase in factor loadings in our study is that items were grouped by factor. Our change in item presentation, taking the possible impact of *fibrofog* (Katz et al., 2004) into account, may have facilitated patients' responses to items as a result of a clearer understanding of what the four items per dimension are attempting to measure (Schell and Oswald, 2013). Further

TABLE 5 | Discriminant Validity: Subgroup Comparisons (FMS vs. FMS + Depression) on the CERQ Subscales in Subsample 1 ($n = 160$).

CERQ Subscales [‡] (4–20)	Domínguez-Sánchez et al. (2013) $n = 615$ Spanish students	Garnefski and Kraaij (2007) $n = 99$ with Anx/Dep	FMS total $n = 160$	FMS $n = 24$	FMS + depression $n = 136$	Student's t FMS vs. FMS+DEP	Cohen's d (90%CI)
Self-Blame	10.59 (2.65)	10.97 (4.21)	9.86 (4.34)	7.17 (2.58)	10.34 (4.42)	4.89**	0.75 (0.57–1.01)
Acceptance	13.24 (3.14)	11.68 (3.74)	14.06 (4.22)	12.79 (3.90)	14.28 (4.25)	1.60	–
Rumination	13.34 (3.49)	12.64 (4.04)	13.16 (4.48)	9.42 (3.50)	13.82 (4.32)	4.72**	1.05 (0.87–1.40)
Positive refocusing	10.87 (4.00)	9.21 (3.65)	9.36 (5.08)	13.17 (4.72)	8.69 (4.86)	4.18**	0.93 (0.46–1.14)
Refocus on planning	15.58 (3.25)	12.62 (3.86)	12.96 (4.32)	13.29 (3.93)	12.90 (4.39)	0.41	–
Positive reappraisal	15.21 (3.89)	10.19 (4.09)	11.66 (4.49)	13.67 (4.55)	11.30 (4.40)	2.42*	0.54 (0.08–0.72)
Putting into perspective	13.72 (3.89)	10.54 (3.86)	13.63 (4.02)	13.96 (4.48)	13.57 (3.95)	0.43	–
Catastrophizing	7.96 (2.98)	9.11 (4.19)	9.35 (4.26)	6.25 (2.36)	9.90 (4.29)	6.01**	0.90 (0.71–1.13)
Other-Blame	7.80 (2.53)	7.76 (3.55)	8.16 (4.76)	5.54 (1.93)	8.62 (4.96)	5.30**	0.66 (0.45–0.85)

[‡]Data expressed as means (standard deviation). *n.s.* = non-significant * $p < 0.05$; ** $p < 0.01$.

TABLE 6 | Identification of Cognitive Emotion Regulation Strategies Discriminating Subgroup Membership (FMS with vs. without Depression): Initial Logistic Regression Model and Final Logistic Regression Model (between brackets).

Predictors	Standardized β	SE β	Wald	p
Self-Blame	0.05	0.11	0.18	0.67
Acceptance	0.11	0.07	2.41	0.12
Rumination	0.12	0.08	2.14	0.14
Positive refocusing	–0.13 (–0.15)	0.06 (0.05)	4.20 (10.26)	0.04 (0.01)
Refocus on planning	0.01	0.08	0.01	0.92
Positive reappraisal	–0.06	0.08	0.55	0.46
Putting into perspective	–0.04	0.08	0.23	0.63
Catastrophizing	0.22 (0.28)	0.10 (0.09)	4.68 (10.35)	0.03 (0.01)
Other-Blame	0.20	0.11	3.12	0.08

Total explained variance (Cox & Snell R^2): 24.9%.

Total explained variance (Cox & Snell R^2) of the final model (with the two significant predictors only): 16.9%.

Significance model: $\chi^2_{(9)} = 45.88$, $p < 0.001$.

Significance of the final model (with the two significant predictors only): $\chi^2_{(2)} = 29.61$, $p < 0.001$.

studies are needed to discern in which evaluation circumstances and for whom item grouping or item randomization is most recommended.

All CERQ subscales showed high internal consistency, ranging from 0.77 (Acceptance) to 0.93 (Positive refocusing) and, with minimal exceptions, were null or modestly correlated with each other, indicating that some subscales share common variance but also represent unique dimensions. Only Rumination and Self-blame presented a large correlation (>0.5). In general, the Cronbach's alphas and subscale correlations found here do not differ from those reported by other authors (e.g., Garnefski et al., 2002; Jermann et al., 2006; Zhu et al., 2008; Tuna and Bozo, 2012; Ireland et al., 2017). When 396 Turkish university students

completed the Turkish version of the CERQ, Tuna and Bozo (2012) observed that the subscales were relatively independent with a mean correlation coefficient of 0.2. Internal consistency of the subscales ranged between 0.72 (Self-blame) and 0.83 (Catastrophizing). In a clinical adult population with symptoms of depression and anxiety (Garnefski et al., 2002), Cronbach's alpha values for the CERQ ranged from 0.72 (Acceptance) to 0.85 (Self-blame). We consider it particularly important in our case to establish comparisons because psychometric evidence of the CERQ has mainly been obtained in non-clinical samples composed of healthy community adults or university students.

Although we could not establish causal relationships due to the cross-sectional nature of our data, it is reasonable to infer that some specific cognitive emotion-regulation strategies might be considered risk factors for or protective factors against depressive and anxiety symptoms and functional status in patients with FMS. The following findings are noteworthy. The strategies Refocus on planning and Putting into perspective had non-significant correlations with functional status, pain catastrophizing, depressive symptoms and trait anxiety. The strategies of Catastrophizing, Rumination, and Self-blame emerged as counterproductive strategies. Positive refocusing negatively correlated with the aforementioned pain-related and psychological variables and, finally, Acceptance and Positive reappraisal had relatively small relationships with these variables. In fact, the apparently counterintuitive positive significant correlation between Acceptance and the pain-related and psychological variables is not surprising. Jermann et al. (2006) pointed out that items related to thoughts of acceptance and resignation are mixed up within this strategy. From a clinical perspective, Acceptance is considered to be an adaptive strategy whereas resignation is similar to helplessness. Higher Acceptance measured with the CERQ has been found to be positively associated with higher depressive symptoms in both Chinese and North-American samples (Martin and Dahlen, 2005; Zhu et al., 2008). Acceptance exhibited significant positive correlations with general symptoms of psychopathology in a Turkish sample (Tuna and Bozo, 2012). Even the designers of the instrument found that Acceptance had significant positive

relationships with depressive symptoms in a general adult sample and in the elderly (Garnefski and Kraaij, 2006). Thus, taking the body of literature and our higher-order factor models into account, we can conclude that Acceptance (as measured in the full version of the CERQ) cannot be considered as part of the repertoire of adaptive cognitive emotion regulation strategies. We agree with Martin and Dahlen (2005, p. 1256) when they stated that “*the presumably adaptive role of acceptance needs to be reconceptualised.*”

In addition, we were interested in analyzing whether frequency of use of *a priori* adaptive and less adaptive emotion regulation strategies was influenced by the presence of comorbid depression. FMS patients with clinically relevant depression were expected to use less adaptive strategies more frequently than those patients without comorbid depressive symptoms. We used a CES-D cut-off to dichotomize the FMS sample (depressed vs. non-depressed). Although it is well-known that splitting a variable into categories results in loss of information and might increase the probability of type II errors (Altman and Royston, 2006), we observed additive effects of depression, that is, the relationship between FMS and cognitive emotion regulation was influenced by the presence of clinically relevant depressive symptoms. Specifically, those participants suffering clinically relevant depression reported more frequent use of Self-blame, Other-blame, Rumination, and Catastrophizing and less use of Positive refocusing and Positive reappraisal, which is clinically coherent. The subsequent regression analyses revealed that Catastrophizing and Positive refocusing were the strategies that significantly discriminated between patients with/without depression. Bearing in mind the high prevalence of clinically relevant depressive symptoms detected in our sample and that depression is characterized by impaired emotion regulation (Joormann and Stanton, 2016), the innovative *Emotion Regulation Therapy* (ERT; Mennin and Fresco, 2014; Renna et al., 2017) might be a potential add-on treatment for patients with FMS plus co-occurring depression. Originally developed for generalized anxiety disorder comorbid with major depression, ERT is a transdiagnostic mechanism-targeted treatment for distress disorders, which makes it an interesting therapeutic option for FMS, a distress-related disorder according to some specialists in this syndrome (Schweinhardt et al., 2012).

Our study is limited by the use of self-report measures and by its cross-sectional nature, which prevents causal inferences and the assessment of important psychometric aspects such as test-retest reliability, sensitivity to change, or longitudinal prediction of clinically relevant and pain-related constructs. Moreover, assessment of the habitual use of cognitive emotion regulation strategies relies on recall, which may be particularly problematic for strategies whose use is highly contextually dependent, such as Acceptance or Positive reappraisal. In addition, due to the predominance of women among participants, we were not able to examine gender differences in the use of CERQ strategies, as has been done in many previous studies carried out in Western, Middle Eastern, and Eastern countries (Martin and Dahlen, 2005; Megreya et al., 2016). We did not implement statistical techniques to mitigate potential “method biases” (Podsakoff et al.,

2012) in our data because we judged that our participants were able to provide accurate answers. In fact, the CERQ items were grouped by dimension in the present work, a change in item presentation that facilitates responses as a result of a clear understanding of what the items related to each dimension are attempting to measure. Moreover, method biases are less likely in respondents that are motivated to provide optimal responses to the items. Patients with FMS have a strong desire for self-expression, CERQ items imply intellectual challenge and in part some emotional catharsis; and patients have the desire to help clinicians improve available treatments for their condition. In summary, stylistically or non-differentiated responding was not expected *a priori*.

To sum up, our findings indicate that the CERQ is a sound instrument for assessing cognitive emotion regulation in patients with FMS and the reported results add to several previous studies that have found a consistent association between cognitive emotion regulation strategies and depressive-anxious symptoms across countries and across clinical and non-clinical samples.

AUTHOR CONTRIBUTIONS

AF-S and JL made substantial contribution to the analysis and to the interpretation of the data, drafted the manuscript, provided final approval of the version to be published, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. ER-C, XB, and MP-M made substantial contributions to the conception and the design of the study, drafted the manuscript, provided final approval of the version to be published, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. AP-A and LA-R helped out in the interpretation of data for the work, revised the manuscript critically for important intellectual content, provided final approval of the version to be published, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. MN-G, JG-C, and JB helped out in the interpretation of data for the work, revised the manuscript critically for important intellectual content, provided final approval of the version to be published, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

The project has been funded in part by the Instituto de Salud Carlos III (ISCIII) of the Ministry of Economy and Competitiveness (Spain) through the Network for Prevention and Health Promotion in Primary Care (RD12/0005/0001; RD16/0007/0005; RD16/0007/0010; RD16/0007/0012), by a grant for research projects on health from ISCIII (PI15/00383) and

cofinanced with European Union ERDF funds. The first listed author (AS) has a “Sara Borrell” research contract from the ISCIII (CD16/00147). The fourth listed author has a FI predoctoral contract awarded by the Agency for Management of University

and Research Grants (AGAUR; 2017; FI_B 00754). The last listed author (JL) has a “Miguel Servet” research contract from the ISCIII (CP14/00087). Thanks to Stephen Kelly for English editing.

REFERENCES

- Abdi, S., Taban, S., and Ghaemian, A. (2012). Cognitive emotion regulation questionnaire: validity and reliability of the persian translation of the CERQ (36-item). *Procedia-Soc. Behav. Sci.* 32, 2–7. doi: 10.1016/j.sbspro.2012.01.001
- Aldao, A., Nolen-Hoeksema, S., and Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: a meta-analytic review. *Clin. Psychol. Rev.* 30, 217–237. doi: 10.1016/j.cpr.2009.11.004
- Altman, D. G., and Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ* 332:1080. doi: 10.1136/bmj.332.7549.1080
- Bennett, R. M., Friend, R., Jones, K. D., Ward, R., Han, B. K., and Ross, R. L. (2009). The revised fibromyalgia impact questionnaire (FIQR): validation and psychometric properties. *Arthritis Res. Ther.* 11:R120. doi: 10.1186/ar2783
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd Edn.* Hillsdale, NJ: Erlbaum.
- Domínguez-Lara, S. A., and Medrano, L. A. (2016). Propiedades psicométricas del cuestionario de regulación cognitiva de las emociones (CERQ) en estudiantes universitarios de Lima [Psychometric properties of the cognitive emotional regulation questionnaire (CERQ) in college students from Lima]. *Psychologia. Avances de la Disciplina* 10, 53–67. doi: 10.21500/19002386.2466
- Domínguez-Sánchez, F. J., Lasa-Aristu, A., Amor, P. J., and Holgado-Tello, F. P. (2013). Psychometric properties of the Spanish version of the cognitive emotion regulation questionnaire. *Assessment* 20, 253–261. doi: 10.1177/1073191110397274
- Edwards, R. R., Campbell, C., Jamison, R. N., and Wiech, K. (2009). The neurobiological underpinnings of coping with pain. *Curr. Dir. Psychol. Sci.* 18, 237–241. doi: 10.1111/j.1467-8721.2009.01643.x
- Feliu-Soler, A., Borràs, X., Peñarrubia-María, M. T., Rozadilla-Sacanell, A., D'Amico, F., Moss-Morris, R., et al. (2016). Cost-utility and biological underpinnings of mindfulness-based stress reduction (MBSR) versus a psychoeducational programme (FibroQoL) for fibromyalgia: a 12-month randomised controlled trial (EUDAIMON study). *BMC Complement. Altern. Med.* 16, 81. doi: 10.1186/s12906-016-1068-2
- García-Campayo, J., Rodero, B., Alda, M., Sobradie, N., Montero, J., and Moreno, S. (2008). Validación de la versión española de la escala de la catastrofización ante el dolor (pain catastrophizing scale) en la fibromialgia [validation of the spanish version of the pain catastrophizing scale in fibromyalgia]. *Med. Clín.* 131, 487–492. doi: 10.1157/13127277
- Garnefski, N., and Kraaij, V. (2006). Relationships between cognitive emotion regulation strategies and depressive symptoms: a comparative study of five specific samples. *Pers. Individ. Dif.* 40, 1659–1669. doi: 10.1016/j.paid.2005.12.009
- Garnefski, N., and Kraaij, V. (2007). The cognitive emotion regulation questionnaire. *Eur. J. Psychol. Assess.* 23, 141–149. doi: 10.1027/1015-5759.23.3.141
- Garnefski, N., Kraaij, V., and Spinoven, P. (2001). Negative life events, cognitive emotion regulation and emotional problems. *Pers. Individ. Dif.* 30, 1311–1327. doi: 10.1016/S0191-8869(00)00113-6
- Garnefski, N., Van Den Kommer, T., Kraaij, V., Teerds, J., Legerstee, J., and Onstein, E. (2002). The relationship between cognitive emotion regulation strategies and emotional problems: comparison between a clinical and a non-clinical sample. *Eur. J. Pers.* 16, 403–420. doi: 10.1002/per.458
- Geenen, R., van Ooijen-van der Linden, L., Lumley, M. A., Bijlsma, J. W., and van Middendorp, H. (2012). The match-mismatch model of emotion processing styles and emotion regulation strategies in fibromyalgia. *J. Psychosom. Res.* 72, 45–50. doi: 10.1016/j.jpsychores.2011.09.004
- Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C. (1998). *Multivariate Data Analysis, 5th Edn.* Upper Saddle River, NJ: Prentice Hall.
- Häuser, W., Ablin, J., Fitzcharles, M. A., Littlejohn, G., Luciano, J. V., Usui, C., et al. (2015). Fibromyalgia. *Nat. Rev. Dis. Primers* 1:15022. doi: 10.1038/nrdp.2015.22
- Ireland, M. J., Clough, B. A., and Day, J. J. (2017). The cognitive emotion regulation questionnaire: factorial, convergent, and criterion validity analyses of the full and short versions. *Pers. Individ. Dif.* 110, 90–95. doi: 10.1016/j.paid.2017.01.035
- Jermann, F., Van der Linden, M., d'Acremont, M., and Zermatten, A. (2006). Cognitive emotion regulation questionnaire (CERQ): confirmatory factor analysis and psychometric properties of the french translation. *Eur. J. Psychol. Assess.* 22, 126–131. doi: 10.1027/1015-5759.22.2.126
- Joormann, J., and Stanton, C. H. (2016). Examining emotion regulation in depression: a review and future directions. *Behav. Res. Ther.* 86, 35–49. doi: 10.1016/j.brat.2016.07.007
- Katz, R. S., Heard, A. R., Mills, M., and Leavitt, F. (2004). The prevalence and clinical impact of reported cognitive difficulties (fibrofog) in patients with rheumatic disease with and without fibromyalgia. *J. Clin. Rheumatol.* 10, 53–58. doi: 10.1097/01.rhu.0000120895.20623.9f
- Luciano, J. V., Aguado, J., Serrano-Blanco, A., Calandre, E. P., and Rodríguez-Lopez, C. M. (2013). Dimensionality, reliability, and validity of the revised fibromyalgia impact questionnaire in two Spanish samples. *Arthritis Care Res.* 65, 1682–1689. doi: 10.1002/acr.22034
- Luciano, J. V., Forero, C. G., Cerdà-Lafont, M., Peñarrubia-María, M. T., Fernández-Vergel, R., Cuesta-Vargas, A. I., et al. (2016). Functional status, quality of life, and costs associated with fibromyalgia subgroups. *Clin. J. Pain* 32, 829–840. doi: 10.1097/AJP.0000000000000336
- Malfliet, A., Coppieters, I., Van Wilgen, P., Kregel, J., De Pauw, R., Dolphens, M., et al. (2017). Brain changes associated with cognitive and emotional factors in chronic pain: a systematic review. *Eur. J. Pain* 21, 769–786. doi: 10.1002/ejp.1003
- Martin, R. C., and Dahlen, E. R. (2005). Cognitive emotion regulation in the prediction of depression, anxiety, stress, and anger. *Pers. Individ. Dif.* 39, 1249–1260. doi: 10.1016/j.paid.2005.06.004
- Medrano, L., Moretti, L., Ortiz, A., and Pereno, G. (2013). Validación del cuestionario de regulación emocional cognitiva en universitarios de Córdoba, Argentina [validation of the cognitive emotion regulation questionnaire in university students of Córdoba, Argentina]. *Psykhe* 22, 83–96. doi: 10.7764/psykhe.22.1.473
- Megreya, A. M., Latzman, R. D., Al-Attayah, A. A., and Alrashidi, M. (2016). The robustness of the nine-factor structure of the cognitive emotion regulation questionnaire across four arabic-speaking middle eastern countries. *J. Cross Cult. Psychol.* 47, 875–890. doi: 10.1177/0022022116644785
- Mennin, D. S., and Fresco, D. M. (2014). “Emotion regulation therapy,” in *Handbook of Emotion Regulation, 2nd Edn.*, ed J. J. Gross New York, NY: Guilford Press.
- Naragon-Gainey, K., McMahon, T. P., and Chacko, T. P. (2017). The structure of common emotion regulation strategies: a meta-analytic examination. *Psychol. Bull.* 143, 384–427. doi: 10.1037/bul0000093
- Nunnally, J., and Bernstein, I. (1994). *Psychometric Theory, 3rd Edn.* New York, NY: McGraw Hill.
- Podsakoff, P. M., MacKenzie, S. B., and Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annu. Rev. Psychol.* 63, 539–569. doi: 10.1146/annurev-psych-120710-100452
- Potthoff, S., Garnefski, N., Miklósi, M., Ubbiali, A., Domínguez-Sánchez, F. J., Martins, E. C., et al. (2016). Cognitive emotion regulation and psychopathology across cultures: a comparison between six European countries. *Pers. Individ. Dif.* 98, 218–224. doi: 10.1016/j.paid.2016.04.022
- Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Appl. Psychol. Meas.* 1, 385–401. doi: 10.1177/014662167700100306

- Renna, M. E., Quintero, J. M., Fresco, D. M., and Mennin, D. S. (2017). Emotion regulation therapy: a mechanism-targeted treatment for disorders of distress. *Front. Psychol.* 8:98. doi: 10.3389/fpsyg.2017.00098
- Rodriguez, A., Reise, S. P., and Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *J. Pers. Assess.* 98, 223–237. doi: 10.1080/00223891.2015.1089249
- Schell, K. L., and Oswald, F. L. (2013). Item grouping and item randomization in personality measurement. *Pers. Individ. Dif.* 55, 317–321. doi: 10.1016/j.paid.2013.03.008
- Schweinhardt, P., Fitzcharles, M. A., Boomershine, C., Vierck, C., and Yunus, M. B. (2012). Fibromyalgia as a disorder related to distress and its therapeutic implications. *Pain Res. Treat.* 2012:950602. doi: 10.1155/2012/950602
- Smarr, K. L., and Keefe, A. L. (2011). Measures of depression and depressive symptoms: beck depression inventory-II (BDI-II), center for epidemiologic studies depression scale (CES-D), geriatric depression scale (GDS), hospital anxiety and depression scale (HADS), and patient health questionnaire-9 (PHQ-9). *Arthritis Care Res.* 638, S454–S466. doi: 10.1002/acr.20556
- Spielberger, C. D., Gorsuch, R. L., and Lushene, R. E. (1986). *Cuestionario de Ansiedad Estado-Rasgo, 2nd Edn.* Madrid, TEA.
- Sullivan, M. J. L., Bishop, S., and Pivik, J. (1995). The pain catastrophizing scale: development and validation. *Psychol. Assess.* 7, 524–532. doi: 10.1037/1040-3590.7.4.524
- Thompson, R. A. (1994). Emotion regulation: a theme in search of definition. *Monogr. Soc. Res. Child Dev.* 59, 25–52. doi: 10.2307/1166137
- Tuna, E., and Bozo, Ö. (2012). The cognitive emotion regulation questionnaire: factor structure and psychometric properties of the Turkish version. *J. Psychopathol. Behav. Assess.* 34, 564–570. doi: 10.1007/s10862-012-9303-8
- van Middendorp, H., Lumley, M. A., Jacobs, J. W., van Doornen, L. J., Bijlsma, J. W., and Geenen, R. (2008). Emotions and emotional approach and avoidance strategies in fibromyalgia. *J. Psychosom. Res.* 64, 159–167. doi: 10.1016/j.jpsychores.2007.08.009
- Vázquez, F. L., Blanco, V., and López, M. (2007). An adaptation of the Center for Epidemiologic Studies Depression Scale for use in non-psychiatric Spanish populations. *Psychiatry Res.* 149, 247–252. doi: 10.1016/j.psychres.2006.03.004
- Vilagut, G., Forero, C. G., Barbaglia, G., and Alonso, J. (2016). Screening for depression in the general population with the center for epidemiologic studies depression (CES-D): a systematic review with meta-analysis. *PLoS ONE* 11:e0155431. doi: 10.1371/journal.pone.0155431
- Wolfe, F., Smythe, H. A., Yunus, M. B., Bennett, R. M., Bombardier, C., Goldenberg, D. L., et al. (1990). The American college of rheumatology 1990 criteria for the classification of fibromyalgia. report of the multicenter criteria committee. *Arthritis Rheum.* 33, 160–172. doi: 10.1002/art.1780330203
- Zelkowitz, R. L., and Cole, D. A. (2016). Measures of emotion reactivity and emotion regulation: convergent and discriminant validity. *Pers. Individ. Dif.* 102, 123–132. doi: 10.1016/j.paid.2016.06.045
- Zhu, X., Auerbach, R. P., Yao, S., Abela, J. R., Xiao, J., and Tong, X. (2008). Psychometric properties of the cognitive emotion regulation questionnaire: chinese version. *Cogn. Emot.* 22, 288–307. doi: 10.1080/02699930701369035

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

Copyright © 2017 Feliu-Soler, Reche-Camba, Borràs, Pérez-Aranda, Andrés-Rodríguez, Peñarrubia-María, Navarro-Gil, García-Campayo, Bellón and Luciano. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Confirmatory Factor Analysis of the French Version of the Savoring Beliefs Inventory

Philippe Golay¹, Bénédicte Thonon², Alexandra Nguyen³, Caroline Fankhauser³ and Jérôme Favrod^{3*}

¹ Community Psychiatry Service, Department of Psychiatry, University Hospital Centre, Lausanne, Switzerland, ² Psychology and Neuroscience of Cognition Research Unit, Department of Psychology, University of Liège, Liège, Belgium, ³ La Source, School of Nursing Sciences, HES-SO University of Applied Sciences and Arts of Western Switzerland, Lausanne, Switzerland

OPEN ACCESS

Edited by:

Michela Balsamo,
Università Degli Studi "G. D'Annunzio"
Chieti - Pescara, Italy

Reviewed by:

Roberta Romanelli,
Università Degli Studi "G. D'Annunzio"
Chieti - Pescara, Italy
Paul Easton Jose,
Victoria University of Wellington,
New Zealand

*Correspondence:

Jérôme Favrod
j.favrod@ecolelasource.ch

Specialty section:

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

Received: 20 October 2017

Accepted: 02 February 2018

Published: 19 February 2018

Citation:

Golay P, Thonon B, Nguyen A,
Fankhauser C and Favrod J (2018)
Confirmatory Factor Analysis of the
French Version of the Savoring Beliefs
Inventory. *Front. Psychol.* 9:181.
doi: 10.3389/fpsyg.2018.00181

The Savoring Beliefs Inventory (SBI) is a measure designed to assess attitudes toward savoring positive experience within three temporal orientations: the past (reminiscence), the present moment (present enjoyment), and the future (anticipation). The aim of this study was to validate the structure of the SBI—French version. The scale was tested with 335 French-speaking participants. Two models were estimated: a one-factor model representing a general construct of savoring and a three-factor model differentiating between anticipation, present enjoyment, and reminiscence. Several indicators of model fit were used: the root mean square error of approximation (RMSEA), the comparison fit index (CFI), the Tucker–Lewis fit index (TLI), and the standardized root mean residual (SRMR). A chi-square difference test was used to compare the two models. The model fit of the three-factor model assessed by the SRMR showed to be excellent, while it could be considered as satisfactory according to the CFI and TLI coefficients. RMSEA, however, was slightly less adequate. The model fit for the one-factor model seemed less adequate than the three-factor solution. Further, the chi-square difference test revealed that the three-factor model had significantly better fit than the one-factor model. Finally, the reliability of the four scores (anticipating pleasure, present moment pleasure, reminiscing pleasure, and total score) was very good. These results show that the French version of the SBI is a valid and valuable scale to measure attitudes regarding the ability to savor positive experience, whether it be in anticipation, reminiscence, or the present moment.

Keywords: savoring, positive affect, emotion regulation, wellbeing, happiness

INTRODUCTION

Subjective wellbeing does not rely solely on the absence of distress, dysfunctional psychological processes, and mental disorders, nor on the ability to cope with negative experiences (Bryant and Veroff, 1984; Trompeter et al., 2017). The experience of positive emotions and, above all, the savoring of these pleasant emotions, have an independent and singular input for subjective wellbeing (Bryant, 1989; Carl et al., 2013; Hurley and Kwon, 2013). Savoring characterizes the ability to generate, increase, and prolong enjoyment, with a deliberate attentiveness to and awareness of the pleasure (Bryant, 2003; Jose et al., 2012). Facing the same positive event, two individuals will anticipate, enjoy, and reminisce to different extents and, therefore, experience different levels of positive emotions and wellbeing. Thus, it is not only the frequency of pleasant

experiences or the ability to feel pleasure that matters to wellbeing but also the capacity to upregulate positive emotions. The ability to savor positive emotions has as much importance as dampening negative emotions has for subjective wellbeing (Nelis et al., 2011).

A large number of scales measure dysfunctional attitudes and emotional regulations (e.g., Garnefski and Kraaij, 2006; Innamorati et al., 2013). However, the exclusive use of these scales might not paint a reliable picture of one's emotional functioning, nor of one's subjective wellbeing (Nelis et al., 2011). A few measurements exist that capture one's ability to savor. Such scales would shed light on the strengths and limitations of an individual. Such evaluation would, consequently, guide therapy into relying on some emotional competencies and reinforcing or developing the weaker savoring abilities. Scales evaluating the ability to savor would also enable the evaluation of the effects of psychotherapy and any approach that intends to foster wellbeing, for example, interventions targeting emotional regulation and anhedonia (Meyer et al., 2012; Favrod et al., 2015).

To date, scales measuring positive emotion regulation in depth are limited to a few, including the Responses to Positive Affect scale (RPA) (Feldman et al., 2008), the Emotion Regulation Profile-Revised (ERP-R) scale (Nelis et al., 2011), and the Savoring Beliefs Inventory (SBI) (Bryant, 2003). The RPA focuses on the tendency to dampen positive emotions and to ruminate positively. Positive rumination consists of recurrently thinking of positive emotions or events (e.g., successes). The ERP-R measures several strategies related to emotion down-regulation and upregulation and includes both maladaptive and adaptive strategies. The adaptive positive emotion upregulation strategies include displaying positive emotions, mindfully savoring the present moment, capitalizing (i.e., celebrating and communicating about positive events), and positive mental time traveling (i.e., reminiscing or anticipating positive events). Other scales include only a few items that focus on positive emotion upregulation, such as the Emotion Regulation Questionnaire (Gross and John, 2003). The SBI was created to evaluate individuals' attitudes regarding savoring positive experiences. Its strength is its focus on positive upregulation of emotions and its inclusion of the three temporal orientations: the past (reminiscence), the present moment (present enjoyment), and the future (anticipation) (Bryant, 2003).

The SBI is composed of 24 items, each temporal orientation being represented by 8 items. Half of the items are positively formulated (e.g., "I find it easy to enjoy myself when I want to"), while the other half is negatively framed (e.g., "I don't like to look forward too much"). Thus, the scale measures, on the one hand, the propensity to savor pleasure and the beliefs in the capacity of savoring, and on the other hand, the negative attitudes concerning savoring and the difficulties one might have regarding the ability to savor.

The SBI has been validated in English-speaking populations (college students and elderly people) and shows good psychometric properties, as seen in the six studies conducted by Bryant (2003). Indeed, the total score of the SBI showed very good internal consistency (Cronbach's alpha between 0.88 and 0.94), and the subscales demonstrated moderate to high internal consistency (Cronbach's alpha between 0.68 and 0.89).

Three-week test-retest correlations indicated strong temporal reliability (SBI total score, $r = 0.84$; Anticipating subscale, $r = 0.80$; Present moment subscale, $r = 0.88$; and Reminiscing subscale, $r = 0.85$, all $p < 0.0001$). The SBI total score correlated positively with various variables indicating good convergent validity, i.e., affect intensity (study 3, $r = 0.48$), optimism (study 4, $r = 0.50$), extraversion (study 4, $r = 0.42$), happiness intensity (study 3, $r = 0.45$; study 6, $r = 0.56$), percent of time happy (study 3, $r = 0.55$; study 6, $r = 0.61$), gratification (study 1, $r = 0.39$; study 2, $r = 0.37$), and self-esteem (study 1, $r = 0.39$; all $p < 0.05$, Bonferroni-adjusted). Good discriminant validity was evidenced by negative correlations between the SBI total score and hopelessness (study 4, $r = -0.41$), neuroticism (study 2, $r = -0.38$), physical anhedonia (study 4, $r = -0.56$), social anhedonia (study 4, $r = -0.57$), strain (study 2, $r = -0.33$), and percent of time unhappy (study 3, $r = -0.35$; study 6, $r = -0.57$; all $p < 0.05$, Bonferroni-adjusted).

Regarding gender differences, numerous studies have provided evidence that women experience joy and naturally savor pleasure to a greater extent than men do (e.g., Diener et al., 1999; Gentzler et al., 2016; but for a more complex review of the question, please refer to Zuckerman et al., 2017). Bryant found that women scored higher than did men on the SBI total scale [$F_{(1,445)} = 11.21$, $p < 0.001$], the Anticipating subscale [$F_{(1,445)} = 9.18$, $p < 0.003$], the Present moment subscale [$F_{(1,445)} = 4.97$, $p < 0.03$], and the Reminiscing subscale [$F_{(1,445)} = 10.96$, $p < 0.001$] (Bryant, 2003).

To date, there has not been any translation of the SBI into other languages. The goal of this study was to validate the French translation of the SBI and to determine which factor structure is more appropriate for the scale.

MATERIALS AND METHODS

Participants

Participants were 335 volunteers who were enrolled in the La Source School of Nursing Sciences in Lausanne as pre-graduate students or as professionals in continuous education courses (19.09% male and 80.91% female). The mean age was 28.09 years ($SD = 9.72$). Participants responded voluntarily and anonymously, there was no way they could be identified, and no personal data concerning their health were collected. They did not receive credit to participate. This study is outside the scope of the Swiss Human Research Act because no personal data concerning human diseases and concerning the structure and function of the human body (HRA art. 2) were collected. Therefore, this study did not need to be authorized by an ethics committee.

Instrument

The SBI is a self-assessment questionnaire composed of 24 items, divided into three temporal orientations, past, present, and future, each represented by 8 items. Half of the items are positively formulated, while the other half are negatively framed. Each item is rated on a 7-point Likert scale ranging from "strongly disagree" to "strongly agree." The total score of the SBI is calculated by subtracting the sum score of the negatively

framed items from the sum score of positively phrased items. The three subscales—Anticipating pleasure, Present moment pleasure, and Reminiscing pleasure—are calculated in the same fashion. The Anticipating pleasure subscale measures savoring a future positive event beforehand, the Present moment pleasure subscale measures enjoying positive events when they occur and the Reminiscing pleasure subscale measures recalling past positive events after they have occurred.

The original English version of the SBI was independently translated by three native French-speaking members of our workgroup, JF, CF and AN, and compared until full agreement was found. The translation was authorized by the author of the original version.

Statistical Analyses

All reverse-scored items were re-coded before data-analysis. For the confirmatory factor analysis (CFA), item data were treated as categorical ordinal, and the models were estimated using a robust weighted least squares estimator with adjustments for the mean and variance (WLSMV). The hypothesized three-factor scoring structure was first tested (Bryant, 2003). It included an Anticipating pleasure factor (items 1, 4, 7, 10, 13, 16, 19, and 22), a Present moment pleasure factor (items 2, 5, 8, 11, 14, 17, 20, and 23), and a Reminiscing pleasure factor (items 3, 6, 9, 12, 15, 18, 21, and 24). Because a total score was also considered in the original scale, this model was compared to a more parsimonious structure including one general savoring factor.

Several indicators of model fit were used, such as the root mean square error of approximation (RMSEA), the comparison fit index (CFI), the Tucker–Lewis fit index (TLI), and the standardized root mean residual (SRMR). RMSEA < 0.06, SRMR < 0.08, and CFI/TLI > 0.95 are interpreted as having good fit, while values of RMSEA ≤ 0.08, SRMR < 0.10, and CFI/TLI ≥ 0.90 are considered as indicating acceptable fit (Hu and Bentler, 1999; Kline, 2005). One should note, however, that interpretation of global fit indexes in models with ordered categorical indicators is not as well established as it is with continuous indicators (Hu and Bentler, 1999). While simulation studies suggest that these cut-off values work reasonably well with categorical outcomes (Yu, 2002; Muthén, 2004), exact cut-off scores may not perfectly apply in the context of this study. Accordingly, alternative models were compared using a robust chi-square test using the DIFFTEST procedure. The reliability of the three subscales was estimated with McDonald's model-based Omega (ω) coefficient (Canivez, 2016). Age and gender differences were assessed by regressing each of the latent scores on the age and gender variables. All statistical analyses were performed with the Mplus statistical package version 7.4.

RESULTS

CFA

As shown in **Table 1**, the model fit of the three-factor model assessed by the SRMR was shown to be excellent, while it could be considered satisfactory according to the CFI and TLI coefficients. RMSEA, however, was slightly less adequate. Overall model fit could be considered as satisfactory and, as

TABLE 1 | Comparisons of model fit for the SBI ($N = 335$).

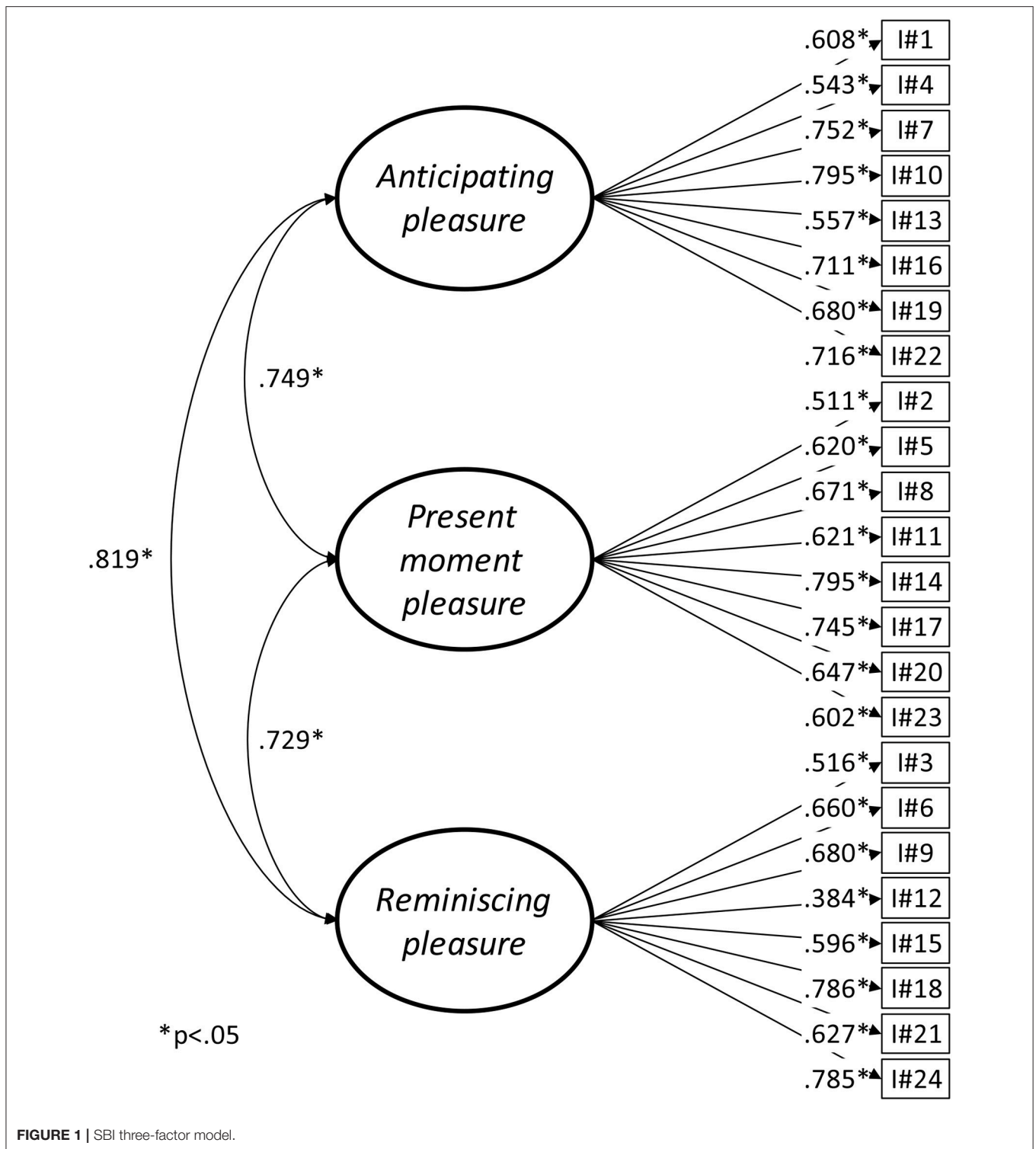
Model	χ^2	df	p-value	RMSEA	CFI	TLI	SRMR
Three-factor model	793.752	249	<0.001	0.081	0.915	0.905	0.069
One-factor model	1067.243	252	<0.001	0.098	0.872	0.860	0.082

df, degree of freedom; RMSEA, Root Mean Square Error of Approximation; CFI, Comparative Fit Index; TLI, Tucker-Lewis Index; SRMR, Standardized Root Mean Square Residual.

indicated on **Figure 1**, all factor loadings were supported. Factor correlations were high, suggesting that items could potentially be explained by a single dimension. A simpler one-factor model was estimated and compared to the three-factor structure. Model fit seemed less adequate than the three-factor solution although all factor loadings were supported (**Figure 2**). Because these two models were statistically nested, they could be compared using a robust chi-square difference test. Result confirmed that the three-factor model had significantly better fit than the one-factor model and should therefore be preferred ($\Delta\chi^2 = 130.598$, $\Delta df = 3$, $p < 0.001$). Although statistically equivalent to the three-factor model, a higher-order model with three first-order factors loading onto a single overarching latent construct of savoring was estimated. The goal was to allow the determination of which of the three factors had the highest or lowest loading on the overarching construct. The loadings were high and quite similar (Anticipating pleasure = 0.917, Present moment pleasure = 0.817, Reminiscing pleasure = 0.893). The reliability of the four scores (ω Anticipating pleasure = 0.879, ω Present moment pleasure = 0.860, ω Reminiscing pleasure factor = 0.851, ω Total score = 0.941) was very good (Canivez, 2016). Additionally, when regressed on age and gender, the four latent scores (Total score, Anticipating pleasure, Present moment pleasure, and Reminiscing pleasure) were not significantly related with these socio-demographic variables (**Table 2**).

DISCUSSION

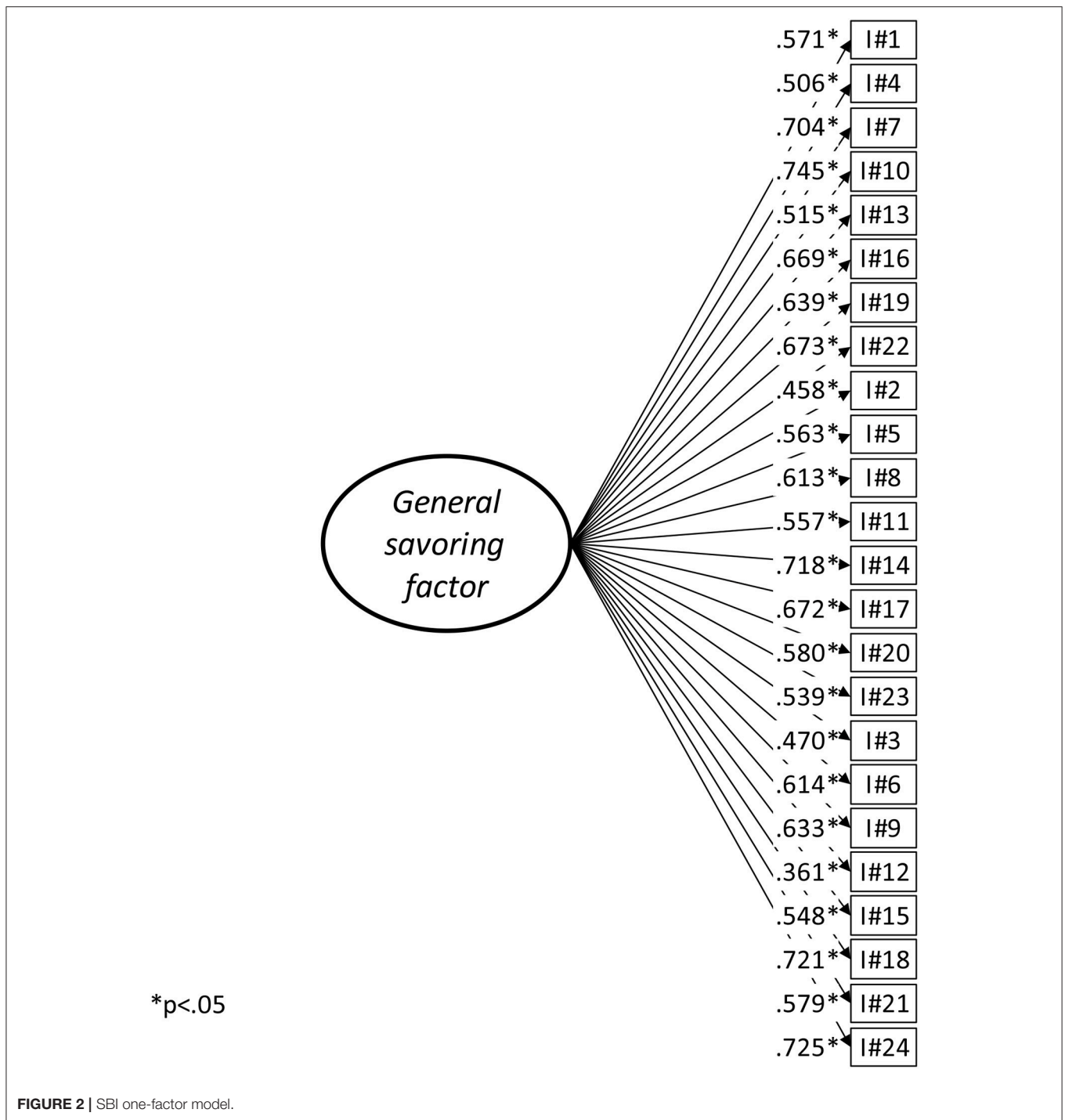
This study investigated the factor structure of the French version of the SBI. The results of the CFA indicated that the hypothesized three-factor structure of the French SBI was adequate, and all items contributed significantly to their corresponding factor: Anticipating pleasure, Present moment pleasure, and Reminiscing pleasure. The model-based reliability of all scores was very good. The three types of pleasure savoring were substantially correlated and shared between 53 and 67% of their variance. These results suggest that individuals able to experience pleasure in one of these three subdomains were more likely to be able to do so in the two other dimensions. However, based on the comparison between the one- and three-factor models, these three types of savoring may not be considered as undifferentiated and may represent theoretically meaningful and distinct dimensions. Despite the large amount of shared variance, there are theoretical benefits to conceptualizing savoring beliefs with three subscales rather than one (Bryant, 2003). The high



correlation between the three subscales suggests that respondents are likely to have similar scores, on average. However, this will not always be the case, and, in our opinion, these differences may allow the identification of important clinical conditions.

Taken together, these results show that the SBI is a valid instrument to investigate savoring capacities in the three examined time frames.

In the original scale, compared to men, women showed higher mean scores on SBI total score and the different subscales



(Bryant, 2003), which was not the case here. Further, in our sample, age was not related to any of the scores of the SBI.

Our study has several limitations that could be the focus of future studies. First, our sample comprised only students in the mental health field, which might not be representative of the general population in terms of savoring abilities. Further, our sample was relatively young and included a large majority of females. A study involving a more representative

sample of the French-speaking population, including more males and elderly people, would further help understanding the different savoring abilities. Second, to further validate the French version of the SBI, concurrent and divergent validity must be examined. Further research on the psychometric characteristics of this scale may also include different clinical groups (e.g., people diagnosed with depression or schizophrenia). Finally, experimental designs may be used to examine the

TABLE 2 | Relationship between age and gender and the four factors.

	Standardized Estimate	Standard Error	p-value
ANTICIPATING PLEASURE			
Gender	0.036	0.062	0.557
Age	-0.001	0.065	0.989
PRESENT MOMENT PLEASURE			
Gender	-0.021	0.064	0.738
Age	-0.019	0.067	0.779
REMINISCING PLEASURE			
Gender	0.122	0.064	0.055
Age	0.085	0.068	0.212
GENERAL SAVORING FACTOR			
Gender	0.047	0.060	0.435
Age	0.022	0.066	0.738

scale's sensitivity to change before and after psychosocial interventions.

The current study showed that the French version of the SBI is an internally valid instrument with very good model-based reliability. The results showed that the French version of the SBI was successfully adapted from the American version. This scale may, therefore, be a valuable tool for French-speaking clinicians and researchers who need to explore savoring attitudes, for instance, in relation to the maintenance or the development of wellbeing, as well as for the development of new interventions focusing on pleasure with clinical populations

REFERENCES

- Bryant, F. (2003). Savoring Beliefs Inventory (SBI): a scale for measuring beliefs about savouring. *J. Ment. Health* 12, 175–196. doi: 10.1080/0963823031000103489
- Bryant, F. B. (1989). A four-factor model of perceived control: avoiding, coping, obtaining, and savoring. *J. Pers.* 57, 773–797. doi: 10.1111/j.1467-6494.1989.tb00494.x
- Bryant, F. B., and Veroff, J. (1984). Dimensions of subjective mental health in American men and women. *J. Health Soc. Behav.* 25, 116–135. doi: 10.2307/2136664
- Canivez, G. L. (2016). “Bifactor modeling in construct validation of multifaceted tests: implications for understanding multidimensional constructs and test interpretation,” in *Principles and Methods of Test Construction: Standards and Recent Advancements*, eds. K. Schweizer and C. DiStefano (Göttingen: Hogrefe), 247–271.
- Carl, J. R., Soskin, D. P., Kerns, C., and Barlow, D. H. (2013). Positive emotion regulation in emotional disorders: a theoretical review. *Clin. Psychol. Rev.* 33, 343–360. doi: 10.1016/j.cpr.2013.01.003
- Chaix, J., Golay, P., Fankhauser, C., Nguyen, A., Gooding, D. C., and Favrod, J. (2017). Confirmatory factor analysis of the french version of the anticipatory and consummatory interpersonal pleasure scale. *Front. Psychol.* 8:1296. doi: 10.3389/fpsyg.2017.01296
- Diener, E., Suh, E. M., Lucas, R. E., and Smith, H. L. (1999). Subjective well-being: three decades of progress. *Psychol. Bull.* 125, 276–302. doi: 10.1037/0033-2909.125.2.276
- Favrod, J., Ernst, F., Giuliani, F., and Bonsack, C. (2009). Validation of the Temporal Experience of Pleasure Scale (TEPS) in a French-speaking environment. *Encephale* 35, 241–248. doi: 10.1016/j.encep.2008.02.013

(Nguyen et al., 2016). The French version of the SBI completes the available scales for assessing pleasure in this language (Favrod et al., 2009; Chaix et al., 2017).

ETHICS STATEMENT

The project was conducted in accordance with ethical code regarding research with human participants and was exempted from the institutional review board.

AUTHOR CONTRIBUTIONS

JF, PG, and AN designed this research. JF, AN, and CF independently translated the scale into French and found translation agreement. JF and AN acquired the data. PG and JF analyzed and interpreted the data. BT, PG, and JF drafted the first version of the manuscript. All the authors approved the final version for publication. All the authors agree to be accountable for all aspects of the work by ensuring that any questions related to its accuracy or integrity can be appropriately investigated and resolved.

FUNDING

This work was supported by the Swiss National Science Foundation grant number 105319_163355. BT is funded by a doctoral research fellow grant (FRESH) from the “Belgian National Fund for Scientific Research” (FNRS).

- Favrod, J., Nguyen, A., Fankhauser, C., Ismailaj, A., Hasler, J.-D., Ringuet, A., et al. (2015). Positive Emotions Program for Schizophrenia (PEPS): a pilot intervention to reduce anhedonia and apathy. *BMC Psychiatry* 15:231. doi: 10.1186/s12888-015-0610-y
- Feldman, G. C., Joormann, J., and Johnson, S. L. (2008). Responses to positive affect: a self-report measure of rumination and dampening. *Cognit. Ther. Res.* 32, 507–525. doi: 10.1007/s10608-006-9083-0
- Garnefski, N., and Kraaij, V. (2006). Cognitive emotion regulation questionnaire - development of a short 18-item version (CERQ-short). *Pers. Individ. Diff.* 41, 1045–1053. doi: 10.1016/j.paid.2006.04.010
- Gentzler, A. L., Palmer, C. A., and Ramsey, M. A. (2016). Savoring with intent: investigating types of and motives for responses to positive events. *J. Happiness Stud.* 17, 937–958. doi: 10.1007/s10902-015-9625-9
- Gross, J. J., and John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *J. Pers. Soc. Psychol.* 85, 348–362. doi: 10.1037/0022-3514.85.2.348
- Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equat. Model. A Multidiscip. J.* 6, 1–55. doi: 10.1080/10705519909540118
- Hurley, D. B., and Kwon, P. (2013). Savoring helps most when you have little: interaction between savoring the moment and uplifts on positive affect and satisfaction with life. *J. Happiness Stud.* 14, 1261–1271. doi: 10.1007/s10902-012-9377-8
- Innamorati, M., Tamburello, S., Contardi, A., Imperatori, C., Tamburello, A., Saggino, A., et al. (2013). Psychometric properties of the attitudes toward self-revised in Italian young adults. *Depress Res. Treat.* 54, 46–52. doi: 10.1155/2013/209216
- Jose, P. E., Lim, B. T., and Bryant, F. B. (2012). Does savoring increase happiness? A daily diary study. *J. Posit. Psychol.* 7, 176–187. doi: 10.1080/17439760.2012.671345

- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Press.
- Meyer, P. S., Johnson, D. P., Parks, A., Iwanski, C., and Penn, D. L. (2012). Positive living: a pilot study of group positive psychotherapy for people with schizophrenia. *J. Posit. Psychol.* 7, 239–248. doi: 10.1080/17439760.2012.677467
- Muthén, B. O. (2004). *Mplus Technical Appendices*. Los Angeles, CA: Muthén & Muthén.
- Nelis, D., Quoidbach, J., Hansenne, M., and Mikolajczak, M. (2011). Measuring individual differences in emotion regulation: the Emotion Regulation Profile-Revised (Erp-R). *Psychol. Belg.* 51, 49–91. doi: 10.5334/pb-51-1-49
- Nguyen, A., Frobert, L., McCluskey, I., Golay, P., Bonsack, C., and Favrod, J. (2016). Development of the positive emotions program for schizophrenia: an intervention to improve pleasure and motivation in schizophrenia. *Front. Psychiatry* 7:13. doi: 10.3389/fpsy.2016.00013
- Trompetter, H. R., Lamers, S. M. A., Westerhof, G. J., Fledderus, M., and Bohlmeijer, E. T. (2017). Both positive mental health and psychopathology should be monitored in psychotherapy: confirmation for the dual-factor model in acceptance and commitment therapy. *Behav. Res. Ther.* 91, 58–63. doi: 10.1016/j.brat.2017.01.008
- Yu, C.-Y. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes*. Los Angeles, CA: University of California, Los Angeles.
- Zuckerman, M., Li, C., and Diener, E. F. (2017). Societal conditions and the gender difference in well-being. *Pers. Soc. Psychol. Bull.* 43, 329–336. doi: 10.1177/0146167216684133

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

The reviewer, RR, and handling Editor declared their shared affiliation.

Copyright © 2018 Golay, Thonon, Nguyen, Fankhauser and Favrod. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Measuring the Capacity to Love: Development of the CTL-Inventory

Nestor D. Kapusta^{1*}, Konrad S. Jankowski², Viktoria Wolf¹,
Magalie Chéron-Le Guludec¹, Madlen Lopatka¹, Christopher Hammerer¹,
Alina Schnieder¹, David Kealy³, John S. Ogrodniczuk³ and Victor Blüml¹

¹ Department for Psychoanalysis and Psychotherapy, Medical University of Vienna, Vienna, Austria, ² Faculty of Psychology, University of Warsaw, Warsaw, Poland, ³ Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

OPEN ACCESS

Edited by:

Marco Innamorati,
Università Europea di Roma, Italy

Reviewed by:

Marco Tommasi,
Università degli Studi "G. d'Annunzio"
Chieti-Pescara, Italy

Elisa Pedroli,
Istituto Auxologico Italiano (IRCCS),
Italy

*Correspondence:

Nestor D. Kapusta
nestor.kapusta@meduniwien.ac.at

Specialty section:

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

Received: 16 March 2018

Accepted: 11 June 2018

Published: 24 July 2018

Citation:

Kapusta ND, Jankowski KS, Wolf V,
Chéron-Le Guludec M, Lopatka M,
Hammerer C, Schnieder A, Kealy D,
Ogrodniczuk JS and Blüml V (2018)
Measuring the Capacity to Love:
Development of the CTL-Inventory.
Front. Psychol. 9:1115.
doi: 10.3389/fpsyg.2018.01115

Objective: The individual capacity to love (CTL) has been linked to various mental health parameters and is considered to be an important outcome parameter of psychotherapeutic treatment. However, empirical examinations of the concept have not been conducted up to now. The aim of this study was to develop a valid and reliable instrument for the assessment of CTL [Capacity to Love Inventory (CTL-I)] as a trait of personality, which is shown to be related to clinically relevant symptoms and conditions.

Method: Four independent healthy samples in Austria ($n = 547$, $n = 174$, and $n = 85$) and Poland ($n = 240$) were assessed by a prototype of the CTL-I and its final shorter version in a confirmatory factor analysis (CFA). Internal consistency of the total questionnaire and each subscale was assessed by Cronbach alpha. External validity was measured against Beck Depression Inventory, Quality of Relationship Inventory, Sociosexual Orientation Inventory, Pathological Narcissism Inventory, and Narcissistic Personality Inventory according to the theoretical framework of the CTL concept. Further test–retest reliability was assessed.

Results: The CFA confirmed 41 items in six dimensions: Interest in the life project of the other, Basic trust, Humility and gratitude, Common ego ideal, Permanence of sexual passion, and Acceptance of loss/jealousy/mourning. The Cronbach alphas of the total CTL-I and its subscales ranged between 0.67 and 0.90 in all samples, suggesting a valid construct. The CTL-I was moderately positively associated with quality of relationship (Support $r = 0.63$, Conflict $r = -0.66$, and Depth $r = 0.66$) and inversely associated with symptoms of depression ($r = -0.37$), pathological narcissism ($r = -0.29$) and promiscuity ($r = -0.42$). The test–retest reliability of the total CTL-I was high with $r = 0.81$, suggesting the stability of answers over time.

Conclusion: The proposed 41-item version of the CTL-I is a psychometrically sound and validated instrument measuring six dimensions of the concept of the CTL. The reported negative associations with clinically relevant parameters such as depression, pathological narcissism and promiscuity as well as associations with relationship qualities such as conflicts, support, and depth warrant its future use in burdened populations including couples in clinical settings.

Keywords: Capacity to Love Inventory (CTL-I), psychometrics properties, validity and reliability, psychotherapy, psychoanalytic theory

“...we must begin to love in order not to fall ill. . .”

Freud, 1914, p. 85

INTRODUCTION

Love is one of the most fundamental human phenomena and it has been the subject of poets, philosophers and religious considerations for millennia. As an evolved commitment love is linked to better health and survival, and plays a critical role in the evolution of humans (Fletcher et al., 2015). Despite its ubiquity, research has not agreed upon a theoretical basis for the phenomenon of love, resulting in many open questions for research (Levin, 2000). Some of the first empirical psychological and sociological approaches focused on declared aspects of love such as its romantic features (Rubin, 1970) or attempted to define love-styles such as eros, agape, pragma, mania, storge, ludus (Lee, 1973, 1977; Hendrick and Hendrick, 1986). Following the scientific differentiation of love aspects and styles, the first attempt of a unified theory of love was established empirically by Sternberg (1986) with the introduction of love-components such as *intimacy*, *passion* and *commitment*, which facilitate description of different phenotypes of love relations. However, although these descriptive typologies helped to establish phenotypic love styles, there is a lack of love concepts that take into account etiological dimensions and links to psychopathological vicissitudes of love. This would allow understanding love relations from a more dynamic functional perspective.

Given the frequency with which difficulties in love relationships are linked with clinical complaints, it would be helpful to obtain a better empirical understanding not of love itself, but of the individual capacity to love (CTL), and to develop means of assessing the components of this construct. This would allow for an enhanced study of personality characteristics associated with difficulties in love relationships, as well as those characteristics which may be strengthened in order to enhance relational functioning. Although close relationships are indisputably associated with well-being, the mechanisms involved remain less well-understood (Feeney and Collins, 2015). Measurement of the CTL may also be useful for clinicians who, in the course of addressing their patients' psychological concerns, encounter various manifestations of impairments in the area of committed love relations.

Research suggests that love-related aspects such as social contact, libido and sexual activity are reduced during episodes of psychiatric disorders (Mathew and Weinman, 1982; Davidson and Turnbull, 1986). For example, a poor quality of social relationships or the relationship with partner and family is independently related to an increased risk of depression (Mamun et al., 2009; Teo et al., 2013) and highly anxious persons experience more conflicts in relationships than non-anxious individuals (Campbell et al., 2005). While stable intimate bonds are associated with psychological health (Burman and Margolin, 1992), the incapacity to maintain close relationships is associated with emotional distress (Bloom et al., 1978; Simon and Marcussen, 1999). Also, a propensity to engage in casual sex or sexual activity in uncommitted relationships (referred to as sociosexual orientation), is a predictor of instability of romantic relationships (Simpson et al., 2004; Penke and Asendorpf, 2008). Similarly, love themes are frequently present in suicide notes among both genders (Canetto and Lester, 2002) and suicidal

behavior is often associated with disappointed love relationships (Séguin et al., 2014; Andreoli et al., 2016).

The here proposed concept of CTL is based on an integrated psychoanalytic theory formulated in terms of object relations theory (Modell, 1963; Bergmann, 1971; Kernberg, 1974a,b, 1977, 2011; Garza-Guerrero, 2000; Gottlieb, 2002) and empirically-based relationship science. The concept involves multiple components and refers to the ability to engage in, invest in, and sustain a committed romantic love relationship (Kernberg, 2011a). These components reflect critical aspects of psychological development theorized to contribute to successful partner relationship involvement. Indeed, from a developmental perspective, the CTL may be regarded as a culmination of complex processes that begin in early caregiving relationships (Zayas et al., 2011; Fraley and Roisman, 2015) and continue to be shaped throughout childhood, adolescent, and early adult developmental experience (Collins and Sroufe, 1999).

In terms of Erikson's model of psychosocial development, the basic trust established in early caregiving relationships may provide a foundation for similar feelings of security in adult romantic relationships (Erikson, 1963; Marcia, 2014). This is consistent with attachment theory and research regarding the role of early experience in later partnerships (Bowlby, 1969; Hazan and Shaver, 1987; Pittman et al., 2011; Zayas et al., 2011). Later childhood and adolescent developmental achievements – characterized as autonomy, initiative, and identity – contribute to the individual's ability to invest in relationships that involve intermittent disappointments, compromises, and potential separations (Erikson, 1963; Beyers and Seiffge-Krenke, 2010; Marcia, 2014). In this way, the ability to tolerate loss and mourning allows for the management of potential affronts to the individual's sense of self in the face of inevitable relationship ruptures. Those who are averse to psychological pain may struggle to fully invest in a gratifying and meaningful – though imperfect – love relationship.

As a function of personality development, the CTL is thought to undergo considerable intensification during Erikson's stage of Intimacy versus Isolation. In the successful negotiation of this phase of development, typically occurring in the emerging adulthood period, the individual develops the ability to share life interests and goals with another person (Erikson, 1963). Intimacy is thus both an interpersonal process involving the interactions of two people (Reis and Shaver, 1988), and an individual intrapsychic developmental achievement that portends for the health of long-term relationships (Weinberger et al., 2008). In the context of committed romantic relationships, intimacy involves the prioritization of a partner's needs, as well as the acceptance of one's vulnerability towards and dependence upon the partner (Kernberg, 2011a). Such vulnerability and dependence is likely to be intermittently tested during conflicts, calling for an ability to forgive and to repair relationship ruptures in the interest of the greater good of the couple as a unit. The blending of two identities into a common relational identity is further symbolized in passionate sexual relations in which transient experiences of merger may occur (Kernberg, 2011b). Individuals who lack the ability to develop a sustained absorption in the interests and goals of another,

whilst simultaneously preserving a robust sense of personal identity, are likely to encounter difficulties in committed love relationships.

Numerous psychoanalytic theorists have drawn attention to personality structures that are organized at least in part around the management of intimate relatedness and its implications for the individual's emotional equilibrium (Sullivan, 1953; Balint, 1979; Guntrip, 1992; Kernberg, 1995; Akhtar, 2000). Some individuals, for example, though desiring a love relationship, may dread an imagined engulfment – a perceived loss of autonomy – should they invest deeply in an intimate partnership. Some individuals may experience the investment in another as a depletion of the self, preferring instead an unrestricted sexuality with relatively limited emotional commitment. For others, the interdependence of a close relationship may evoke anxieties about the individual's vulnerability and acceptability to a partner, potentially stimulating controlling behaviors aimed at both influencing the partner and regulating the individual's feelings of insecurity. In this way, the failure to acquire a mature capacity for intimate, mutually gratifying, and deeply committed relatedness may be associated with self-regulatory psychopathology.

Pathological narcissism, a personality syndrome involving deficits in and maladaptive mechanisms regarding the maintenance of self-image, is perhaps exemplary of psychological circumstances involving an impaired CTL (Kernberg, 1995, 2011a; Kealy and Ogrodniczuk, 2014). Indeed, individuals with high levels of pathological narcissism tend to report anxious attachment patterns and histories of unsatisfactory relationships (Kealy et al., 2015), as well as domineering, vindictive, and intrusive interpersonal behaviors (Ogrodniczuk et al., 2009).

The present study had three objectives. The first objective was to operationalize the theory driven construct of CTL by developing a psychometric tool, the Capacity to Love Inventory (CTL-I), for assessing CTL and confirming the results using samples from two different countries. The second objective was to test the scale's convergent and divergent validity relative to other presumably related psychological concepts (narcissism, depression, relationship quality, and sociosexual orientation). The third objective was to closer examine associations between dimensions of the CTL and these related psychological concepts as a way to advance the construct validity of CTL-I.

MATERIALS AND METHODS

In order to avoid problems in operationalization of the CTL construct, the study concept was oriented on the unified theory of construct validity (Messick, 1995). In synthesizing psychoanalytic theories, clinical observations, and empirical science regarding the CTL, Kernberg (2011a) has furnished a conceptualization of several areas that represent critical potential impediments in the area of romantic relational functioning. The components identified by Kernberg (2011a) served as our guide in developing a psychometric scale capable of measuring the CTL and comprised: Falling in love (FIL), Interest in the other (INT), Basic trust (BTR), Forgiveness (FRG), Gratitude (GRT), Common ego

ideal (CEI), Mature dependency, Permanence of sexual passion (PSP), and Loss and mourning (LOM). To reflect each of the domains, psychoanalytic literature referenced by Kernberg (2011a) and additional theories on CTL were incorporated (Modell, 1963; Bergmann, 1971; Kernberg, 1974a,b, 1977; Garza-Guerrero, 2000; Gottlieb, 2002; Kealy and Ogrodniczuk, 2014). Based on these theories, content validity of items was assured by a joint discussion between research group members at the Capacity to Love Research Lab, and 70 items were generated in English and German simultaneously (NK, VW, MCL, and VB) then translated into Polish language and back translated by native speakers (KJ and NK).

Participants

The present study utilized four different samples for the development and testing of the CTL-I. *Sample 1* (Austrian sample) was recruited to permit determination of the factorial structure of the initial 70-item questionnaire. The sample consisted of 547 (82.1% females) full datasets (out of 942 subjects who started but dropped off during some point of assessment) aged 16 to 66 ($M = 28.92$, $SD = 10.22$). They were recruited by snowball sampling procedure within the social network of medical students at the Medical University of Vienna, their families and friends, and were invited to fill an online questionnaire (German language). *Sample 2* (in Poland) consisted of 240 participants (82.9% females) aged 18 to 50 ($M = 23.24$, $SD = 4.21$), mainly psychology students at University of Warsaw, who were contacted by email addressed to students of the department. No dropouts and missing values were reported. The Polish sample was used to confirm the structure of the CTL-I that was derived from Sample 1. *Sample 3* (in Austria) consisted of 174 full datasets (out of 233) subjects (58.6% females) aged 18 to 70 ($M = 29.53$, $SD = 12.10$) recruited with the intention of assessing construct validity with reference to pathological narcissism. The same recruiting procedure was applied to another independent *Sample 4* ($N = 85$, out of 125 approaching subjects in Austria), which was recruited to enable investigation of test-retest reliability of the confirmed scale structure based on Samples 1 and 2. The participants were asked to fill the questionnaire at baseline and 4 weeks later. In all studies a forced-item procedure was adopted which does not allow participants to proceed if items were left blank. In all studies participants were asked to refer to an ongoing relationship or in absence of such to the last significant relationship they had. The studies were conducted under the code of the Declaration of Helsinki and received a positive decision (1515/2013, 1179/2015, and 1184/2015) from the ethics committee at the Medical University of Vienna.

Questionnaires

Capacity to Love Inventory (CTL-I)

The initial 70-items of the prospective questionnaire was applied in sample 1 (German translation) and the final reduced version with 41 items was used in sample 2 (polish translation), samples 3 (German) and 4 (German). The full item list of the final version is presented in **Table 1**. The items were rated on a four-point Likert scale ranging between 1 and 4.

TABLE 1 | Factor loadings (residual variances) from confirmatory factor analysis and item statistics in Austrian (AT) and Polish (PL) samples.

	<i>Interest in the other (INT)</i>	AT				PL			
		Loading (variance)	M(SD)	Skewness	Kurtosis	Loading (variance)	M(SD)	Skewness	Kurtosis
1	It is important to me to know the life plan of my partner.	0.49(0.26)	3,66(0.58)	-1.65	2.52	0.34(0.27)	3.75(0.55)	-2.43	6.26
2	I share my life plans with my partner.	0.62(0.33)	3,42(0.73)	-1.07	0.52	0.50(0.60)	3.18(0.90)	-0.86	-0.13
3	I am joyful to share my partner's success.	0.68(0.14)	3,70(0.51)	-1.48	1.27	0.69(0.16)	3.75(0.56)	-2.59	7.46
4	I feel enriched to see the personal growth and life experience of my partner.	0.64(0.18)	3,67(0.55)	-1.54	1.78	0.63(0.25)	3.62(0.64)	-1.63	2.24
5	When my partner is unhappy, I also feel sad.	0.48(0.25)	3,60(0.57)	-1.28	1.59	0.53(0.51)	3.24(0.85)	-0.89	0.03
6	I can be empathic with my partner and try to understand her/him.	0.50(0.26)	3,53(0.58)	-0.97	0.83	0.58(0.23)	3.59(0.59)	-1.35	2.11
7	I am often bored with my partner.*	0.45(0.43)	3,37(0.73)	-1.03	0.77	0.50(0.42)	3.36(0.75)	-1.00	0.54
	Basic trust (BTR)								
8	I trust that my partner is empathic with me when necessary.	0.57(0.40)	3,21(0.77)	-0.89	0.64	0.75(0.33)	3.02(0.87)	-0.69	-0.08
9	My weaknesses, inner conflicts and problems are open to the other.	0.71(0.32)	3,14(0.8)	-0.69	0.03	0.46(0.56)	3.20(0.85)	-0.78	-0.21
10	I can express my feelings and needs to my partner openly.	0.78(0.25)	3,31(0.79)	-0.87	-0.07	0.76(0.22)	3.43(0.73)	-1.26	1.42
11	I can trust my partner that in situations of uncertainty and ambivalence she/he can be emotionally supportive.	0.69(0.33)	3,33(0.79)	-1.01	0.40	0.78(0.22)	3.46(0.75)	-1.46	1.91
12	I feel being honest to my partner.	0.72(0.20)	3,60(0.64)	-1.60	2.32	0.82(0.15)	3.56(0.68)	-1.59	2.29
13	I keep secrets from my partner.*	0.57(0.38)	3,37(0.75)	-1.09	0.86	0.36(0.81)	2.73(0.97)	-0.47	-0.70
14	I can confess my weaknesses to my partner.	0.62(0.33)	3,33(0.74)	-0.91	0.43	0.56(0.34)	3.48(0.70)	-1.29	1.35
15	I sometimes feel that my relationships are limited.*	0.49(0.70)	2,62(0.96)	-0.16	-0.91	0.38(0.89)	2.72(1.02)	-0.37	-0.95
16	I am comfortable with my partner and I usually feel safe in his/her company.	0.67(0.30)	3,45(0.73)	-1.29	1.32	0.70(0.18)	3.60(0.60)	-1.46	2.23
	Gratitude (GRT)								
17	I feel gratitude for the existence of my partner.	0.67(0.22)	3,63(0.63)	-1.77	3.13	0.72(0.26)	3.57(0.73)	-1.73	2.48
18	I feel gratitude for the love received.	0.71(0.19)	3,65(0.61)	-1.90	3.80	0.75(0.25)	3.51(0.77)	-1.51	1.55
19	When separated I still feel connected with the partner.	0.64(0.34)	3,44(0.76)	-1.38	1.62	0.66(0.25)	3.57(0.67)	-1.60	2.50
20	I accept that I need my partner.	0.54(0.54)	3,19(0.87)	-0.75	-0.38	0.76(0.23)	3.56(0.74)	-1.78	2.71
21	I like to convey comfort to my partner.	0.66(0.22)	3,62(0.62)	-1.61	2.46	0.53(0.35)	3.59(0.70)	-1.79	2.92
22	I like to take care of the other, when he/she needs my help.	0.66(0.24)	3,56(0.65)	-1.52	2.37	0.58(0.23)	3.65(0.59)	-1.72	3.12
23	I like to share responsibilities in our daily life in order to take pressure off my partner.	0.54(0.36)	3,32(0.71)	-0.86	0.56	0.37(0.69)	2.96(0.90)	-0.45	-0.65
	Common ego ideal (CEI)								
24	I am dedicated to my relationships.	0.60(0.32)	3,28(0.70)	-0.73	0.40	0.69(0.20)	3.61(0.61)	-1.55	2.35
25	We always try to work on our relationship.	0.61(0.35)	3,13(0.74)	-0.57	0.02	0.57(0.32)	3.28(0.69)	-0.67	.25
26	I respect the personality and essential values of my partner.	0.53(0.27)	3,51(0.61)	-0.98	0.68	0.59(0.24)	3.59(0.61)	-1.53	2.84

(Continued)

TABLE 1 | Continued

	<i>Interest in the other (INT)</i>	AT				PL			
		Loading (variance)	M(SD)	Skewness	Kurtosis	Loading (variance)	M(SD)	Skewness	Kurtosis
27	I love to watch my partner's gestures and reactions.	0.55(0.29)	3,54(0.64)	-1.19	0.81	0.51(0.29)	3.65(0.63)	-1.88	3.49
28	I feel committed to our joint life.	0.71(0.24)	3,46(0.70)	-1.28	1.49	0.71(0.19)	3.68(0.62)	-2.05	4.16
29	I search for compromise solutions when conflicts and competing agendas arise.	0.41(0.31)	3,42(0.61)	-0.71	0.57	0.45(0.37)	3.39(0.68)	-1.16	1.89
30	I often tell my partner that I love him.	0.60(0.59)	3,13(0.96)	-0.83	-0.37	0.60(0.74)	3.12(1.07)	-0.88	-0.58
31	I feel deeply connected with my partner.	0.74(0.25)	3,37(0.75)	-0.99	0.37	0.78(0.21)	3.51(0.73)	-1.47	1.63
	Permanence of sexual passion (PSP)								
32	Sexual boredom arises in long-term relationships.*	0.79(0.31)	2,69(0.92)	-0.21	-0.79	0.95(0.08)	3.24(0.93)	-1.06	0.14
33	The sexual desire diminishes throughout time.*	0.89(0.18)	2,76(0.92)	-0.13	-0.94	0.81(0.27)	3.25(0.88)	-1.06	0.36
	Loss and mourning (LOM)								
34	It is hard for me to accept when a loved person is not able to respond to my love.*	0.46(0.62)	3,10(0.89)	-0.67	-0.42	0.41(0.77)	1.76(0.97)	-1.12	0.16
35	When a relationship is over, I often blame my ex-partner.*	0.57(0.67)	3,05(1,00)	-0.72	-0.62	0.40(0.84)	2.78(1.00)	0.32	-0.99
36	I sometimes have wishes for revenge when my partner dismisses me.*	0.57(0.68)	3,18(1,00)	-0.94	-0.34	0.45(0.84)	3.03(1.03)	0.68	-0.77
37	I am often unwilling to accept the end of my relationships.*	0.66(0.63)	2,74(1,07)	-0.24	-1.21	0.59(0.82)	2.59(1.12)	0.09	-1.36
38	I am often jealous.*	0.49(0.68)	2,73(0.94)	-0.31	-0.78	0.55(0.80)	2.47(1.07)	0.09	-1.26
39	I have feelings of guilt after a separation.*	0.45(0.79)	2,78(0.99)	-0.31	-0.97	0.50(0.70)	2.53(0.97)	0.06	-0.96
40	I sometimes devalue myself if my partner abandoned me.*	0.62(0.74)	2,61(1,10)	-0.12	-1.31	0.72(0.59)	2.47(1.11)	-0.11	-1.34
41	It is hard for me to move on after a relationship.*	0.40(0.93)	2,43(1,05)	0.06	-1.20	0.33(0.99)	2.26(1.05)	-0.26	-1.15

*Reversed item; $n = 547$ for Austrian (AT) and $n = 240$ for Polish (PL) sample.

Quality of Relationship Inventory (QRI)

The Quality of Relationship Inventory (QRI) (Pierce et al., 1997) in the German translation was used (Reiner et al., 2012). It is a self-report questionnaire consisting of 25 items that are evaluated on a four-point Likert scale ranging from 1 (not true) to 4 (almost always true). QRI has 25 items forming three dimensions: Support (seven items, e.g., 'To what extent could you count on this person to help with a problem?'), Conflict (12 items, e.g., 'How critical of you is this person?'), and Depth (6 items, e.g., 'How much do you depend on this person?'). The internal consistency for the subscales was 0.84, 0.89, and 0.82 for the respective subscales in a representative German sample (Reiner et al., 2012). Higher scores in Support and Depth dimensions mean better quality of relationship, whereas higher scores in the Conflict are interpreted in terms of lower quality of relationship. The questionnaire was administered in sample 1.

Beck Depression Inventory (BDI-II-R)

Beck Depression Inventory (BDI-II) is a well-established measure of depressive traits (Beck et al., 1988). It allows to be used as a screening measure as well as a measure of severity of depression based on 21 items rated between 0 and 3. Its German translation by Hautzinger et al. (1994) yields satisfying internal consistencies as measured by Cronbach's α ranging between 0.76 and 0.95 in clinical samples and between 0.73 and 0.92 in non-clinical samples. The questionnaire was administered in sample 1.

Revised Version of the Sociosexual Orientation Inventory (SOI-R)

Sociosexual Orientation Inventory (SOI-R) by Penke and Asendorpf (2008) in the Polish translation (Jankowski, 2016) was used to measure sociosexual orientation. The questionnaire was administered in sample 2. Higher scores in SOI-R indicated

unrestricted sociosexuality, whereas lower scores indicated more restricted orientation. The scale used in the study has nine items with a five-point Likert scale response format. It allows for quantification of three facets of sociosexual orientation, i.e., behavior, attitude, desire, and a total score. Each of the three dimensions consists of three items. Sample questions are: behavior ‘With how many different partners have you had sex within the past 12 months?’; attitude ‘Sex without love is OK’; desire ‘In everyday life, how often do you have spontaneous fantasies about having sex with someone you have just met?’ Typically, scores of each scale are expressed as the average of scores obtained from adherent items, and the total score is an average of the scores for the three facets. This allows for comparisons between subscales and between subscales and total score, and produces values between 1 and 9 for each subscale and for the total score. Cronbach’s α in the present study were high for behavior (0.79), attitude (0.82), and desire (0.88), and the total score (0.87).

Pathological Narcissism Inventory (PNI)

The original Pathological Narcissism Inventory (PNI) is a 52-item self-report measure assessing seven dimensions of pathological narcissism including measures of narcissistic grandiosity (Entitlement Rage, Exploitativeness, Grandiose Fantasy, and Self-sacrificing Self-enhancement) and narcissistic vulnerability (Contingent Self-esteem, Hiding the Self, and Devaluing) (Pincus et al., 2009). The applied German version includes a translation of the original and two additional items constructed to extend the exploitative subscale based on DSM diagnostic criteria (Morf et al., 2017). In the German validation study, Cronbach alphas for the subscales ranged between 0.82 (SSSE) and 0.92 (CSE) with an alpha coefficient for the total scale of 0.94. The retest reliability for the subscales ranged from $r = 0.75$ (DEV and SSSE) to 0.87 (CSE and GF) and the reliability for the total score was 0.86 (Morf et al., 2017). The questionnaire was used in sample 3.

Narcissistic Personality Inventory (NPI)

The original Narcissistic Personality Inventory (NPI) is a 40-item self-report measure developed by Raskin and Terry (1988) also available in a 15-items version (Schütz et al., 2004) assessing two dimensions of narcissism (Leadership and Grandiosity) from a social-personality psychology perspective. Leadership represents the ability to lead groups and others, while Grandiosity describes features of personality such as feelings to be a special and unique person. Some research indicates that the NPI assesses adaptive characteristics of narcissism such as achievement motivation, self-esteem, emotional resilience, and extraversion rather than pathological features (Pincus et al., 2009). The applied German NPI-15 translation (Schütz et al., 2004) showed good Cronbach alphas for the subscales with 0.73 and 0.82 and a good test-retest reliability $r = 0.86$ (Schütz et al., 2004). Recently, the two-factor structure was re-examined in a representative German population, resulting in Cronbach alpha of 0.82 for the Leadership and an acceptable 0.69 for the Grandiosity subscale (Spangenberg et al., 2013). The questionnaire was used in sample 3.

Statistical Analysis

The factor structure of the CTL-I was examined by maximum likelihood confirmatory factor analysis (CFA) and goodness of fit was established basing on Hu and Bentler (1998) two-index presentation strategy. Specifically, for complex models, as in the presented work, it is suggested to interfere on fit, based on standardized root mean square residual (SRMR) together with root-mean-square error of approximation (RMSEA) with cutoff values indicative for acceptable fit of around 0.80 or less for SRMR and around 0.60 or less for RMSEA (Hu and Bentler, 1998). We supplemented the above fit indices with Akaike information criterion (AIC) and Bayesian information criterion (BIC) allowing for comparison between competing models (lower the values represent a better fit). Associations with other scales were tested with Pearson correlation and differences between groups were checked using t -test. Internal consistency reliability was assessed by Cronbach alpha. The calculations were performed by IBM SPSS and AMOS (version 22.0). Significance levels were set at 0.05.

RESULTS

Factor Analysis

At first, a 70-item, eight-factor model consisting of FIL, INT, BTR, FRG, GRT, CEI, PSP, LOM was tested in the Austrian sample (sample 1) using CFA, with factors allowed to correlate with each other. Fit indices were: $\chi^2(2317) = 7026.8$, $\chi^2/\text{degrees of freedom} = 3.03$, $SRMR = 0.086$, $RMSEA = 0.061$ (95% CI : 0.059; 0.063), $AIC = 7362.8$, $BIC = 8085.9$, thus they showed mediocre fit to the data due to SRMR exceeding the commonly acknowledged threshold of 0.080 for good fit.

To improve the model fit, we retained items with factor loadings greater than 0.40 and, next, examined internal consistency of each of the eight scales by means of Cronbach alpha. Only scales with internal consistency of at least 0.70 were retained. As a result, the scales FIL (all nine items) and FRG (all seven items) could not be retained, and 13 further items from other subscales dropped out due to too low factor loadings. The resulting six-factor model (with 41 items) was re-tested with CFA, with scales being allowed to correlate with each other. Fit indices were: $\chi^2(764) = 2391.9$, $\chi^2/\text{degrees of freedom} = 3.13$, $SRMR = 0.060$, $RMSEA = 0.062$ (95% CI : 0.060; 0.065), $AIC = 2585.9$, $BIC = 3003.4$. Thus, similarly to the previous model $\chi^2/\text{degrees of freedom}$ and $RMSEA$ were acceptable, and SRMR lowered below threshold value of 0.080. What is more, both AIC and BIC values were lower for the six-factor model. Consequently, in comparison to the initial eight-factor model, the six-factor model (model 2) consisting of INT, BTR, GRT, CEI, PSP, LOM was improved and accepted as the final one.

The next step was to retest the six-factor model (model 2) by CFA in an independent (Polish) sample 2. The results confirmed item loadings on the six-factors. The fit indices were: $\chi^2(764) = 1482.3$, $\chi^2/\text{degrees of freedom} = 1.94$, $SRMR = 0.070$, $RMSEA = 0.063$ (95% CI : 0.058; 0.067) indicating acceptable fit, which was comparable to that observed in sample 1 (Austria).

Factor loadings and Cronbach alphas for the scales for the final model in both samples are shown in **Tables 1, 2**, respectively.

Internal Reliability and Validity of CTL-I Subscales

As shown in **Table 2**, the internal consistency after item reduction in each of the six subscales was good and comparable in both samples 1 and 2. The corresponding Cronbach's alpha values for the total scale were 0.90 and 0.88 respectively. The total alpha was further confirmed in sample 3 with 0.89 (subscale alpha range = 0.68 to 0.82). Similarly, in sample 4, alphas for the CTL total score at time 1 was 0.90 and 0.92 ($N = 85$) at time 2 and alpha values for the subscales at both time points ranged between 0.67 and 0.86.

All six subscales correlated consistently with each other, with moderate associations of the subscale 'Loss and Mourning' with the subscales 'Basic Trust' in Sample 1 and 'Permanence of Sexual Passion' as shown in **Table 2**. When Bonferroni correction is adopted to the correlational analyses and twenty one coefficients are considered, p level should equal 0.002 or less to be considered statistically significant. Using this conservative method associations of PSP with LOM in both samples and with INT and BTR in the Polish sample could be regarded as less firm.

Interestingly, age was only associated with the subscales 'Gratitude' and 'Loss and Mourning' (association with CEI in the Austrian sample would not survive the Bonferroni correction).

Table 3 shows small but significant differences between the Austrian (1) and Polish (2) sample in most subscales. The gender comparison showed that males scored slightly higher than females on the subscale 'Loss and Mourning'. Additionally, Austrian males showed a lower mean than females in the 'Interest for the other' subscale (see **Table 4**), but the p -value of this association exceeds the value of 0.008 imposed by the Bonferroni correction considering six t -tests, making the association less firm.

Convergent and Divergent Validity

Validity was examined by correlations between the CTL-I subscales and other variables. We hypothesized a positive association between CTL and relationship quality in sample 1. Each of the six scales of the CTL-I was moderately correlated

with each of the three dimensions of the QRI. According to expectations, the total CTL-I score and all scales of the CTL-I correlated positively with the dimensions Depth and Support and negatively with the Conflict dimension of QRI. The only exception was the CTL-I subscale 'Loss and Mourning,' which was not significantly correlated with the dimension Depth of QRI (see **Table 5**).

In line with the hypothesis that depressive symptoms are associated with limitations to the CTL, all scales of CTL-I were inversely correlated with depression scores (BDI) (see **Table 5**).

We hypothesized that unrestricted SOI-R, which is a propensity to engage in casual sex or sexual activity in uncommitted relationships, would be negatively correlated with CTL. In fact, we found that the total SOI-R score was negatively correlated with five scales of CTL-I with the exception of 'Loss and Mourning.' This result seems to be based mainly on the correlation of the two dimensions of SOI-R *Attitude* and *Desire*. The third SOI-R dimension *Behavior* was not related to CTL-I subscales with the exception of the 'Permanence of Sexual Passion' scale (see **Table 6**).

The association between CTL-I subscales and pathological narcissism (see **Table 7**) was examined within sample 3. In line with expectations, the total CTL-I and total PNI score were moderately and inversely correlated. The CTL-I subscales 'Loss and Mourning' as well as 'Basic Trust' contributed most to the association. On the other hand, the narcissistic aspects 'Hiding the self,' 'Vulnerability,' and 'Devaluing' contributed most to restrictions in CTL. As further expected, none of the NPI dimensions nor the total score was substantially associated with CTL-I total and subscales.

Test-Retest Reliability

Within sample 4, the test-retest reliability for the total CTL-I score was $r_{tt} = 0.81$. The reliabilities for the subscales ranged from $r_{tt} = 0.64$ (GRT) to $r_{tt} = 0.85$ (LOM) (see **Table 8**).

DISCUSSION

The aim of the study was the psychometric operationalization of the construct of CTL. The underlying theoretical basis of the construct was derived from an integrated psychoanalytic

TABLE 2 | Correlations among factor scores in Samples 1 and 2 (Austria/Poland).

	INT	BTR	GRT	CEI	PSP	LOM	Age
INT	–	0.70***/0.69***	0.80***/0.84***	0.80***/0.94***	0.40***/0.25**	0.11/–0.03	–0.09/0.01
BTR			0.75***/0.69***	0.75***/0.77***	0.34***/0.22**	0.25***/–0.07	–0.08/0.00
GRT				0.91***/0.90***	0.39***/0.30***	0.04/–0.16	–0.15***/0.01
CEI					0.49***/0.32***	0.06/–0.12	–0.12**/0.01
PSP						0.15**/0.17*	–0.06/0.03
LOM						–	0.22***/0.28***
CTL-I total	0.54***/0.76***	0.63***/0.69***	0.75***/0.74***	0.79***/0.80***	0.69***/0.66***	0.40***/0.28***	–0.05/0.10
Cronbach alpha	0.73/0.72	0.86/0.83	0.81/0.80	0.81/0.82	0.83/0.87	0.75/0.72	–

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; $n = 547$ for Austrian sample 1 and $n = 240$ for Polish sample 2. CTL-I scales: INT, interest in the other; BTR, basic trust; GRT, gratitude; CEI, common ego ideal; PSP, permanence of sexual passion; LOM, loss and mourning.

TABLE 3 | Descriptive statistics for the studied measures and comparison of means between countries.

	Austria			Poland		
	Mean (SD)	Skewness	Kurtosis	Mean (SD)	Skewness	Kurtosis
Interest in the other (INT)***	3.32 (0.32)	-1.25	2.30	3.50 (0.43)	-1.22	2.16
Basic trust (BTR)***	3.04 (0.36)	-0.97	0.73	3.24 (0.52)	-0.96	0.68
Gratitude (GRT)	3.49 (0.48)	-1.39	2.73	3.49 (0.49)	-1.85	4.76
Common ego ideal (CEI)***	3.35 (0.47)	-1.11	1.87	3.48 (0.47)	-1.66	4.45
Perm. of sexual passion (PSP)***	2.72 (0.85)	-0.17	-0.73	3.25 (0.85)	-1.10	0.44
Loss and mourning (LOM)***	2.68 (0.61)	-0.29	-0.47	2.49 (0.60)	0.07	-0.50
CTL-I total***	18.60 (1.97)	-0.72	0.88	19.47 (2.16)	-0.91	1.01

*** $p < 0.001$; $n = 547$ for Austrian sample 1 and $n = 240$ for Polish sample 2.

TABLE 4 | Gender differences in CTL-I scales in AT and PL.

	Austria		Poland	
	Women mean (SD)	Men mean (SD)	Women mean (SD)	Men mean (SD)
Interest in the other (INT)	3.33 (0.32)	3.25 (0.34)*	3.52 (0.42)	3.40 (0.45)
Basic trust (BTR)	3.04 (0.36)	3.03 (0.39)	3.25 (0.52)	3.20 (0.56)
Gratitude (GRT)	3.49 (0.48)	3.47 (0.44)	3.50 (0.47)	3.41 (0.58)
Common ego ideal (CEI)	3.37 (0.47)	3.30 (0.45)	3.49 (0.46)	3.41 (0.53)
Permanence of sexual passion (PSP)	2.72 (0.85)	2.74 (0.83)	3.26 (0.85)	3.16 (0.88)
Loss and mourning (LOM)	2.64 (0.61)	2.81 (0.50)**	2.41 (0.58)	2.84 (0.61)***
CTL-I total	18.60 (1.92)	18.60 (1.98)	19.45 (2.17)	19.41 (2.10)

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. $n = 547$ for Austrian sample 1 and $n = 240$ for Polish sample 2.

TABLE 5 | Correlations between dimensions of capacity to love and quality of relationship inventory and depression scores ($M = 9.13$, $SD = 8.59$) (Austrian sample 1, $n = 531$).

	QRI support	QRI conflict	QRI depth	BDI
Interest in the other (INT)	0.49***	-0.36**	0.58***	-0.19**
Basic trust (BTR)	0.68***	-0.51***	0.62***	-0.31**
Gratitude (GRT)	0.54***	-0.37***	0.74***	-0.19**
Common ego ideal (CEI)	0.52***	-0.37***	0.67***	-0.19**
Permanence of sexual passion (PSP)	0.28***	-0.26***	0.32***	-0.21**
Loss and mourning (LOM)	0.21***	-0.36***	-0.01	-0.44**
CTL-I total	0.63***	-0.53***	0.66***	-0.37**

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

TABLE 6 | Correlations between the CTL and SOI-R scales ($n = 240$, Polish sample 2).

	SOI-R behavior	SOI-R attitude	SOI-R desire	SOI-R total score
Mean (SD)	1.86 (0.78)	2.60 (1.22)	2.39 (1.05)	2.38 (0.82)
Interest in the other (INT)	-0.10	-0.34***	-0.34***	-0.35***
Basic trust (BTR)	-0.11	-0.27***	-0.27***	-0.28***
Gratitude (GRT)	-0.04	-0.28***	-0.29***	-0.27***
Common ego ideal (CEI)	-0.01	-0.29***	-0.32***	-0.28***
Permanence of sexual passion (PSP)	-0.19**	-0.27***	-0.40***	-0.37***
Loss and mourning (LOM)	-0.02	0.04	-0.12	-0.04
CTL-I total	-0.14*	-0.36***	-0.46***	-0.42***

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

TABLE 7 | Correlations between the CTL dimensions and narcissistic personality and pathological narcissism scores ($n = 180$, Austrian sample 3).

		INT	BTR	GRT	CEI	PSP	LOM	CTL-I total
PNI	CSE	0.03	-0.15*	-0.03	0.00	-0.15	-0.60**	-0.26**
	DEV	-0.09	-0.26**	-0.16*	-0.11	-0.08	-0.51**	-0.31**
	ER	-0.05	-0.17*	-0.11	-0.10	-0.16*	-0.45**	-0.28**
	EXP	-0.03	-0.01	-0.03	-0.04	-0.07	-0.08	-0.08
	GF	-0.03	-0.10	-0.04	-0.02	-0.01	-0.47**	-0.18*
	HS	-0.13	-0.30**	-0.23**	-0.20**	-0.16*	-0.47**	-0.38**
	SSSE	0.05	-0.05	0.04	0.06	0.00	-0.40**	-0.09
	GRAND	-0.02	-0.11	-0.05	-0.03	-0.07	-0.46**	-0.20**
	VULN	-0.05	-0.25**	-0.13	-0.09	-0.15*	-0.60**	-0.34**
PNI total	-0.04	-0.19**	-0.10	-0.07	-0.12	-0.57**	-0.29**	
NPI	LEAD	0.03	-0.10	0.13	0.09	-0.10	0.04	0.01
	GRAN	0.06	0.11	0.15*	0.16*	0.00	0.16*	0.15*
	NPI total	0.04	-0.04	0.16*	0.12	-0.08	0.09	0.05

** $p < 0.01$; * $p < 0.05$. PNI scales: CSE, Contingent Self-Esteem; DEV, Devaluing; ER, Entitlement Rage; EXP, Exploitativeness; GF, Grandiose Fantasy; HS, Hiding the Self; SSSE, Self-Sacrificing Self-Enhancement; GRAND, Narcissistic Grandiosity Subscale; VULN, Narcissistic Vulnerability Scale. NPI scales: LEAD, Leadership; GRAN, Grandiosity. CTL-I scales: INT, interest in the other; BTR, basic trust; GRT, gratitude; CEI, common ego ideal; PSP, permanence of sexual passion; LOM, loss and mourning.

TABLE 8 | Test-retest correlations ($N = 85$, Austrian sample 4).

	INT 1	BTR 1	GRT 1	CEI 1	PSP 1	LOM 1	CTL total 1
INT 2	0.71**	0.49**	0.63**	0.59**	0.42**	0.01	0.62**
BTR 2	0.59**	0.77**	0.57**	0.53**	0.41**	0.15	0.71**
GRT 2	0.53**	0.53**	0.64**	0.56**	0.39**	-0.03	0.58**
CEI 2	0.59**	0.60**	0.60**	0.73**	0.50**	0.05	0.70**
PSP 2	0.33**	0.40**	0.31**	0.43**	0.81**	0.01	0.48**
LOM 2	0.08	0.27*	-0.05	-0.01	-0.08	0.85**	0.40**
CTL total 2	0.62**	0.72**	0.58**	0.61**	0.46**	0.34**	0.81**

** $p < 0.01$; * $p < 0.05$. CTL-I scales: INT, interest in the other; BTR, basic trust; GRT, gratitude; CEI, common ego ideal; PSP, permanence of sexual passion; LOM, loss and mourning.

theory of the CTL, with emphasis on recent object relations theory which understands current interpersonal relations as linked to early childhood development (Modell, 1963; Bergmann, 1971; Kernberg, 1974a,b, 1977, 2011a; Gottlieb, 2002; Kealy and Ogrodniczuk, 2014). Rather than descriptively characterizing styles of loving, the theory of CTL refers to functioning in romantic committed relationships commonly referred to as love relationships. Accordingly, the theory is based on the assumption that various inhibitions of personality functioning result in current limitations to the CTL and thus in interpersonal difficulties.

Factor Analysis

In order to test the factor structure of the instrument, a CFA approach was chosen because the coherent theoretical construct of CTL and its constituting dimensions were already defined within the framework of Kernberg (2011a). Out of an initial pool of 70 items constituting eight dimensions, the factor analysis finally confirmed 41 items in six dimensions:

(1) Interest in the life project of the other, (2) Basic trust, (3) Humility and gratitude, (4) Common ego ideal, (5) Permanence of sexual passion, and (6) Acceptance of loss/jealousy/mourning. Only two initial dimensions 'Falling in love' and 'Capacity for authentic forgiveness' could not be confirmed as items needed to be excluded due to low loadings and unacceptable low internal consistencies of these subscales. In general, indices of CFA suggested an acceptable fit of the six-factor model and support the face validity of the concept. While all remaining six dimensions of the CTL-I showed moderate correlations with each other, the subscale 'Acceptance of loss/Jealousy/Mourning' represented by items like 'It is hard for me to accept if a loved person is not able to respond to my love' was only modestly related to 'Basic trust' and 'Permanence of sexual passion.' This scale represents coalesced aspects of reactions to loss of love objects, which are in line with the observed inverse associations with depression scores and conflicts in relationships. However, 'Acceptance of loss/Jealousy/Mourning' is not related to depth of the relationship or perceived support therein. It can therefore be seen as a rather stabilizing or neutralizing function in critical

times rather than one that adds to depth or intensity of love relations.

Validity

CTL and Symptoms of Depression

A framework of Freud's comprehensive theory of depression has only recently been formulated for empirical testing and suggests close links between depression and loss of loved objects (Desmet, 2013). In line with our expectations, CTL is inversely associated with depressive symptoms in our study. The hypothesis has a long tradition among many theorists beyond Freud (1917) and his initial theory on the relation between loss of love objects and depression as a representation of the inability to mourn. In this sense, Balint (1952) described the difference between mature and primitive love as determined by strong narcissistic tendencies and unbearable depressive fears which may impair the ability to maintain loving relations. Thus, while loss and total unity with a love object are antagonistic extremes, mature CTL as the ability to bear depressive feelings (Klein, 1940, 1946) represents a protective feature against its symptomatic expressions at both ends of the continuum in form of depression and narcissism. More recent empirical results show that perceived parental love inconsistency is associated with later proneness to depression (Trumpeter et al., 2008). Conversely, depression in adolescence may impair subsequent romantic relationship qualities into late adolescence and emerging adulthood (Vujeva and Furman, 2011), corresponding to an impaired CTL. In the process of transition to parenthood, for example, love received from husband is seen by Benedek (1949, 1959) as a remedy for postpartum depression, which by restoring a narcissistic loss, in turn allows mother to be the source of love for the child. In this sense, further research might help to understand the early effects of parenting styles (Busch and Kapusta, unpublished) and the early nurturing co-parenting environment (Kapusta et al., 2017) on the development of the CTL in offspring.

CTL and Sociosexual Orientation

The concept of sociosexuality, which is related to promiscuity, describes individual differences in the readiness to engage in uncommitted sexual relationships (Penke and Asendorpf, 2008), and thus reflects the capacity to restrict one's own sexual propensity to one love object. Restricted sociosexuality is also related to romantic relationship stability and quality (Simpson, 1987; Simpson and Gangestad, 1991, 1992; Ellis, 1998; Jones, 1998). In line with these facts, our results show that sociosexuality is inversely related to the CTL. The relevance of the ability to restrict one's own sexual propensity seems to be reflected in the emotional and motivational aspects captured by the SOI, namely *attitudes* toward and *desire* for unrestricted sexuality. In contrast, the *behavior* dimension of the SOI, which counts sexual contacts and describes the lifetime allocation of effort to short-term versus long-term mating tactics, was only marginally related to CTL in our Polish sample. The *behavior* dimension was inversely correlated with the CTL scale '*Permanence of sexual passion*'. It seems possible that the sociosexual behavior of our rather young sample does not adequately differentiate between immature and mature CTL, given that sociosexuality increases

with age (Penke and Asendorpf, 2008; Jankowski, 2016), and the rather low mean of the SOI *behavior* dimension in our sample reflects low sexual experience relative to other comparable studies (Penke and Asendorpf, 2008; Jankowski, 2016).

The gender comparison showed that males in both samples 1 and 2 scored higher than females on the subscale '*Loss and Mourning*' and Austrian males showed a lower '*Interest for the life plan of the other*' than females (Table 4). The mean scores of the CTL-I subscales were similarly high in both the Polish and Austrian samples, with the exception of higher means of PSP among Polish participants. It remains to be elucidated in the future, whether this difference between countries is based on cultural/religious, sampling or linguistic differences (Table 3). Although the demonstration of cultural differences between Austria and Poland is beyond the scope of this work, for example, religious beliefs differ between Austria and Poland considerably, with Poland exhibiting more religiousness (Coutinho, 2016). Given the fact that '*Permanence of sexual passion*' is inversely related to promiscuity as measured by SOI-R, our results are supported by the argumentation that *Permanence of sexual passion* is disclosed at higher levels in a more religious country. However, we also admit that the PSP scale could be improved in future as it consists of only two items in the final 41-item CTL-I version.

CTL and Quality of Relationships

Since the functions of the CTL are experienced within relationships, we hypothesized a positive association between CTL and relationship quality. The QRI measuring the dimensions of *support*, *conflict*, and *depth* of relationships is based on theoretical models of social *support* which include interpersonal, intrapersonal, and situational efforts of exchange between two participants (Pierce et al., 1991, 1997). The QRI is based on the assumption that general predispositions to engage in and respond to social behavior are grounded in expectations, derived from Bowlby's (1980) theory of working models and relations between the self and important others. Perceived qualities of *depth* and *support* in intimate relationships were associated most strongly with the domains *BTR*, *gratitude* and a *common ego ideal* of the CTL-I. In applying Mikulincer and Shaver's (2005) model of the interplay between the caregiving and the attachment behavioral systems, in which one person responds to signals of need emitted by the other, the reduction of another person's suffering by provision of support or experience of positive emotions fosters the experience of gratitude and may strengthen attachment security and *BTR*. Also, a perception of a shared ego ideal may increase mutual understanding and thus increase feelings of depth in a relationship. The opposite is true for *conflict* in intimate relationships, which was inversely related to all CTL-I subscales, and notably reflects a loss of *BTR*, reduced feelings of *gratitude* and restrictions in *common ego ideal*. These results are in line with considerations of the emergence of relational tensions and conflicts in the presence of a malfunctioning of the attachment and caregiving system, which otherwise tend toward a maintenance of stable and mutually satisfactory affectional bonds (Mikulincer and

Shaver, 2008). Future studies should assess the associations between CTL-I and attachment styles, the latter being likely a function of the mature dependency dimension of CTL-I, and test the hypothesis formulated by Hazan and Shaver (1987) who argued for understanding romantic love as an attachment process.

Pathological Narcissism and Limitations to the Capacity to Love

Theoretical assumptions suggest that pathological narcissism is associated with an overall difficulty in the CTL or as a fundamental impairment of it (Gottlieb, 2002; Kernberg, 2011a; Kealy and Ogrodniczuk, 2014). In contrast to pathological narcissism with its incapability to love others, balanced love for oneself is generally held to be an essential component of healthy psychosocial functioning (Kealy and Ogrodniczuk, 2014). Pincus et al. (2009) convincingly show that the two measures of narcissism, namely PNI and NPI assess different aspects of narcissism, with the latter capturing more adaptive expressions of healthier and extraverted narcissistic features.

According to our expectations, CTL was moderately and negatively associated with pathological narcissism in our study, but not with the adaptive narcissism as measured by NPI. Interestingly, it was the vulnerability of narcissistic persons that was associated with limitations in the CTL, most notably the CTL-I domains of *BTR* and *LOM*. This is in line with theory, which points to problematic mourning processes in the context of separation from loved objects, due to lack of object constancy. According to Kernberg (2010), instead of mourning, persons with pathological narcissism blame others for the loss of the loved object, and in this way inhibit (and are thus protected from) the painful mourning process. This is often accompanied by a denial that the object could have his or her own independent existence, or in a more omnipotent processing, narcissistic personalities deny their own dependence on others (Garza-Guerrero, 2000).

Falling in Love and Authentic Forgiveness

Some limitations of our approach need to be discussed. Due to poor psychometric properties, the two initial dimensions of “*falling in love*” and “*authentic forgiveness*” were dropped from the final version of the CTL-I. It is not possible to evaluate if this rather reflects problems of the conceptualization and operationalization of these dimensions in the assessment instrument or if these dimensions are indeed no central prerequisites of the CTL. Some authors argue that falling in love represents a process of idealization which can lead to both successful and frustrating experiences in relationships depending on the level of maturity of the idealization process itself, which means that normal vs. pathological idealizations need to be differentiated (Kernberg, 1976, p. 191ff; Garza-Guerrero, 2000). We believe that our attempt to conceptualize the ‘falling in love’ dimension with items like ‘*I have experienced falling in love in my life often*’ or ‘*It often happens that I idealize my partner*’ may not have sufficiently captured the theoretically ambiguous concept of falling in love. The role of falling in love in mature loving remains unclear and although falling in love seems related, it not necessarily is a characteristic of the capacity for mature love

(Kernberg, 1976, p. 237; Kernberg, 2011a). Future attempts to conceptualize falling in love should take non-pathological aspects of idealization into account.

Similarly, the factor analysis could not confirm the dimension ‘*Capacity for authentic forgiveness*’. Kernberg’s (2011a) understanding of authentic forgiveness is based on the acknowledgment of one’s own aggressive potential, the experience of trust and the communication of feelings of being hurt without blaming. We attempted to operationalize these aspects initially in the 70-item version of the CTL-I with items like ‘*When feeling misunderstood or hurt, I express my feelings to the other*’ or the inverse statement ‘*When hurt, I often try to induce guilt feelings in my partner.*’ However, the items did not integrate into a consistent forgiveness scale as expected. Future attempts at operationalizing authentic forgiveness should try to broaden the concept by including other salient facets such as the empathy for the offending partner’s motives (McCullough et al., 1998; Akhtar, 2013, p. 130), which has been linked to the mentalizing capacity of an individual Fonagy (2009, p. 447).

CONCLUSION

According to the objectives of the study, (1) we were largely able to empirically confirm the concept of CTL by operationalization of its theoretical assumptions and have demonstrated that the 41-item CTL-I yields good internal consistency with stable and consistent results in two culturally different samples, and very good test–retest reliability. (2) The scale’s convergent and divergent validity has coherently been established in relation to narcissism, depression, relationship quality and sociosexual orientation. (3) A closer examination of the associations between dimensions of the CTL suggests a further refinement need for the dimension of ‘*permanence of sexual passion*’ to improve the construct validity of CTL-I. Also, the dimensions ‘*Falling in love*’ and ‘*Forgiveness*’ which could be not confirmed by means of CFA in this work should be re-approached in future.

The so established CTL-I allows self-assessment and empirical testing of the relation between CTL-I and other concepts in future, thereby contributing to further understanding of the construct of CTL. Such an instrument might be suitable for the measurement of changes in psychoanalytic and other psychotherapeutic interventions and to help psychotherapists to understand their patients limitations and resources with respect to relationship issues. The resulting CTL-I also adds to the strong need for operationalization of psychoanalytic concepts to promote further empirical studies in psychoanalysis.

AUTHOR CONTRIBUTIONS

NK has developed the project to operationalize the concept of CTL, conceptualized all study designs 1, 2, 3, 4 and serves also the group leader of the CTL research lab. NK has written the manuscript draft and coordinated all researchers contributions in this project. KJ calculated statistics for studies 1 and 2 (CFA),

NK for studies 3 and 4. The following authors contributed to the design and performed the following studies, VW, NK, VB, and MG study 1, NK, VB, and KJ study 2, CH, NK, and AS study 3, ML, JO, DK, and NK study 4. All authors reviewed the manuscript draft, contributed to manuscript writing and literature search and confirmed the manuscript submission.

REFERENCES

- Akhtar, S. (2000). *Broken Structures: Severe Personality Disorders and Their Treatment*. Lanham, MD: Jason Aronson Inc.
- Akhtar, S. (2013). *Good Stuff: Courage, Resilience, Gratitude, Generosity, Forgiveness, and Sacrifice*. Lanham, MD: Jason Aronson Inc.
- Andreoli, A., Burnand, Y., Cochenec, M. F., Ohlendorf, P., Frambati, L., Gaudry-Maire, D., et al. (2016). Disappointed love and suicide: a randomized controlled trial of "Abandonment Psychotherapy" among borderline patients. *J. Pers. Disord.* 30, 271–287. doi: 10.1521/pedi_2015_29_196
- Balint, M. (1952). On love and hate. *Int. J. Psychoanal.* 33, 355–362.
- Balint, M. (1979). *The Basic Fault: Therapeutic Aspects of Regression*. Evanston, IL: Northwestern University Press.
- Beck, A. T., Steer, R. A., and Garbin, M. G. (1988). Psychometric properties of the beck depression inventory: twenty-five years of evaluation. *Clin. Psychol. Rev.* 8, 77–100. doi: 10.1016/0272-7358(88)90050-5
- Benedek, T. (1949). The psychosomatic implications of the primary unit: mother-child. *Am. J. Orthopsychiatry* 19, 642–654. doi: 10.1111/j.1939-0025.1949.tb06521.x
- Benedek, T. (1959). Parenthood as a developmental phase: a contribution to the libido theory. *J. Am. Psychoanal. Assoc.* 7, 389–417. doi: 10.1177/000306515900700301
- Bergmann, M. S. (1971). "Psychoanalytic observations on the capacity to love," in *Separation-Individuation. Essays in Honor of Margaret S. Mahler*, eds J. B. McDevitt and C. F. Settlage (New York, NY: International Universities Press), 15–40.
- Beyers, W., and Seiffge-Krenke, I. (2010). Does identity precede intimacy? Testing Erikson's theory on romantic development in emerging adults of the 21st century. *J. Adolesc. Res.* 25, 387–415. doi: 10.1177/0743558410361370
- Bloom, B. L., Asher, S. J., and White, S. W. (1978). Marital disruption as a stressor: a review and analysis. *Psychol. Bull.* 85, 867–894. doi: 10.1037/0033-2909.85.4.867
- Bowlby, J. (1969). *Attachment and Loss Attachment*, Vol. 1. New York, NY: Basic Books.
- Bowlby, J. (1980). *Attachment and Loss: Loss: Sadness and Depression*, Vol. 3. New York, NY: Basic Books.
- Burman, B., and Margolin, G. (1992). Analysis of the association between marital relationships and health problems: an interactional perspective. *Psychol. Bull.* 112, 39–63. doi: 10.1037/0033-2909.112.1.39
- Campbell, L., Simpson, J. A., Boldry, J., and Kashy, D. A. (2005). Perceptions of conflict and support in romantic relationships: the role of attachment anxiety. *J. Pers. Soc. Psychol.* 88, 510–531. doi: 10.1037/0022-3514.88.3.510
- Canetto, S. S., and Lester, D. (2002). Love and achievement motives in women's and men's suicide notes. *J. Psychol.* 136, 573–576. doi: 10.1080/00223980209605552
- Collins, A., and Sroufe, L. A. (1999). "Capacity for intimate relationships: a developmental construction," in *Development of Romantic Relationships in Adolescence*, eds W. Furman, B. B. Brown, and C. Feiring (New York, NY: Cambridge University Press), 125–147. doi: 10.1017/CBO9781316182185.007
- Coutinho, J. P. (2016). Religiosity in Europe: an index, factors, and clusters of religiosity. *Sociol. Prob. Prát.* 81, 163–188. doi: 10.7458/SPP2016816251
- Davidson, J., and Turnbull, C. D. (1986). Diagnostic significance of vegetative symptoms in depression. *Br. J. Psychiatry* 148, 442–446. doi: 10.1192/bjp.148.4.442
- Desmet, M. (2013). Some preliminary notes on an empirical test of Freud's theory on depression. *Front. Psychol.* 4:158. doi: 10.3389/fpsyg.2013.00158
- Ellis, B. J. (1998). The partner-specific investment inventory: an evolutionary approach to individual differences in investment. *J. Pers.* 66, 383–442. doi: 10.1111/1467-6494.00017
- Erikson, E. H. (1963). *Childhood and Society*, 2nd Edn. New York, NY: Norton.
- Feeney, B. C., and Collins, N. L. (2015). Thriving through relationships. *Curr. Opin. Psychol.* 1, 22–28. doi: 10.1016/j.copsyc.2014.11.001
- Fletcher, G. J., Simpson, J. A., Campbell, L., and Overall, N. C. (2015). Pair-bonding, romantic love, and evolution: the curious case of Homo sapiens. *Perspect. Psychol. Sci.* 10, 20–36. doi: 10.1177/1745691614561683
- Fonagy, P. (2009). "Commentary on forgiveness," in *Good Feelings: Psychoanalytic Reflections on Positive Emotions and Attitudes*, ed. S. Akhtar (London: Karnac Books), 441–452.
- Fraley, R. C., and Roisman, G. I. (2015). Do early caregiving experiences leave an enduring or transient mark on developmental adaptation? *Curr. Opin. Psychol.* 1, 101–106. doi: 10.1016/j.copsyc.2014.11.007
- Freud, S. (1914). "On Narcissism," in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. XIV, ed. J. Strachey (London: Hogarth Press), 67–102.
- Freud, S. (1917). "Mourning and melancholia," in *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, Vol. XIV, 237–258.
- Garza-Guerrero, C. (2000). Idealization and mourning in love relationships: normal and pathological spectra. *Psychoanal. Q.* 69, 121–150. doi: 10.1002/j.2167-4086.2000.tb00557.x
- Gottlieb, S. (2002). "The capacity for love," in *Love and Hate: Psychoanalytic Perspectives*, ed. D. Mann (Hove: Brunner-Routledge).
- Guntrip, H. (1992). *Schizoid Phenomena, Object Relations and the Self*. London: Karnac Books.
- Hautzinger, M., Bauer, M., Worall, H., and Keller, F. (1994). *Beck Depressions-Inventar (BDI). Bearbeitung Der Deutschen Ausgabe. Testhandbuch*. Bern: Huber.
- Hazan, C., and Shaver, P. (1987). Romantic love conceptualized as an attachment process. *J. Pers. Soc. Psychol.* 52, 511–524. doi: 10.1037/0022-3514.52.3.511
- Hendrick, C., and Hendrick, S. (1986). A theory and method of love. *J. Pers. Soc. Psychol.* 50, 392–402. doi: 10.1037/0022-3514.50.2.392
- Hu, L.-T., and Bentler, P. M. (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychol. Methods* 3, 424–453. doi: 10.1037/1082-989X.3.4.424
- Jankowski, K. S. (2016). "Charakterystyka psychometryczna polskiej wersji zrewidowanego Inwentarza Orientacji Socjoseksualnej (SOI-R)," in *Wybrane Metody i Paradygmaty Badawcze w Psychologii*, eds A. Rynkiewicz, K. S. Jankowski, and W. Oniszczenko (Warsaw: Scholar), 77–92.
- Jones, M. (1998). Sociosexuality and motivations for romantic involvement. *J. Res. Pers.* 32, 173–182. doi: 10.1006/jrpe.1997.2212
- Kapusta, N. D., Philipp, A. C., and Feinberg, M. E. (2017). Co-parenting during transition to parenthood: creating a healthy nurturing environment for children. *Austin Child. Adolesc. Psychiatry* 2:1008.
- Kealy, D., and Ogrodniczuk, J. S. (2014). Pathological narcissism and the obstruction of love. *Psychodyn. Psychiatry* 42, 101–119. doi: 10.1521/pdps.2014.42.1.101
- Kealy, D., Ogrodniczuk, J. S., Joyce, A. S., Steinberg, P. I., and Piper, W. E. (2015). Narcissism and relational representations among psychiatric outpatients. *J. Personal. Disord.* 29, 393–407. doi: 10.1521/pedi_2013_27_084
- Kernberg, O. F. (1974a). Barriers to falling and remaining in love. *J. Am. Psychoanal. Assoc.* 22, 486–511. doi: 10.1177/000306517402200302
- Kernberg, O. F. (1974b). Mature Love: prerequisites and characteristics. *J. Am. Psychoanal. Assoc.* 22, 743–768. doi: 10.1177/000306517402200402
- Kernberg, O. F. (1976). *Object Relations Theory and Clinical Psychoanalysis*. Lanham, MD: Jason Aronson Inc.
- Kernberg, O. F. (1977). Boundaries and structure in love relations. *J. Am. Psychoanal. Assoc.* 25, 81–114. doi: 10.1177/000306517702500104
- Kernberg, O. F. (1995). *Love Relations: Normality and Pathology*. New Haven, CT: Yale University Press.
- Kernberg, O. F. (2010). Some observations on the process of mourning. *Int. J. Psychoanal.* 91, 601–619. doi: 10.1111/j.1745-8315.2010.00286.x

FUNDING

The studies in this publication have been supported by a grant funding of NK at the Medical University of Vienna. The contribution of KJ was supported from a grant provided by Faculty of Psychology, University of Warsaw.

- Kernberg, O. F. (2011a). Limitations to the capacity to love. *Inter. J. Psychoanal.* 92, 1501–1515. doi: 10.1111/j.1745-8315.2011.00456.x
- Kernberg, O. F. (2011b). The sexual couple: a psychoanalytic exploration. *Psychoanal. Rev.* 98, 217–245. doi: 10.1521/prev.2011.98.2.217
- Klein, M. (1940). Mourning and its relation to manic-depressive states. *Int. J. Psychoanal.* 21, 125–153.
- Klein, M. (1946). Notes on Some Schizoid Mechanisms. *Int. J. Psychoanal.* 27, 99–110.
- Lee, J. A. (1977). A typology of styles of loving. *Pers. Soc. Psychol. Bull.* 3, 173–182.
- Lee, J. A. (1973). *Colours of Love Toronto*. New York, NY: New Press.
- Levin, J. (2000). A prolegomenon to an epidemiology of love: theory, measurement, and health outcomes. *J. Soc. Clin. Psychol.* 19, 117–136. doi: 10.1521/jscp.2000.19.1.117
- Mamun, A. A., Clavarino, A. M., Najman, J. M., Williams, G. M., O'Callaghan, M. J., and Bor, W. (2009). Maternal depression and the quality of marital relationship: a 14-year prospective study. *J. Womens Health* 18, 2023–2031. doi: 10.1089/jwh.2008.1050
- Marcia, J. E. (2014). From industry to integrity. *Identity* 14, 165–176. doi: 10.1080/15283488.2014.924413
- Mathew, R. J., and Weinman, M. L. (1982). Sexual dysfunctions in depression. *Arch Sex Behav.* 11, 323–328. doi: 10.1007/BF01541593
- McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington, E. L. Jr., Brown, S. W., and Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *J. Pers. Soc. Psychol.* 75, 1586–1603. doi: 10.1037/0022-3514.75.6.1586
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educ. Meas. Issues Pract.* 14, 5–8. doi: 10.1111/j.1745-3992.1995.tb00881.x
- Mikulincer, M., and Shaver, P. R. (2008). Commentary on “Is There a Drive to Love?” *Neuropsychanalysis* 10, 154–165. doi: 10.1080/15294145.2008.10773581
- Mikulincer, M., and Shaver, P. R. (2005). Attachment theory and emotions in close relationships: exploring the attachment-related dynamics of emotional reactions to relational events. *Pers. Relat.* 12, 149–168. doi: 10.1111/j.1350-4126.2005.00108.x
- Modell, A. H. (1963). Primitive object relationships and the predisposition to schizophrenia. *Int. J. Psychoanal.* 44, 282–292.
- Morf, C. C., Schürch, E., Künfer, A., Siegrist, P., Vater, A., Back, M., et al. (2017). Expanding the nomological net of the pathological narcissism inventory: german validation and extension in a clinical inpatient sample. *Assessment* 24, 419–443. doi: 10.1177/1073191115627010
- Ogrodniczuk, J. S., Piper, W. E., Joyce, A. S., Steinberg, P. I., and Duggal, S. (2009). Interpersonal problems associated with narcissism among psychiatric outpatients. *J. Psychiatr. Res.* 43, 837–842. doi: 10.1016/j.jpsychires.2008.12.005
- Penke, L., and Asendorpf, J. B. (2008). Beyond global sociosexual orientations: a more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *J. Pers. Soc. Psychol.* 95, 1113–1135. doi: 10.1037/0022-3514.95.5.1113
- Pierce, G. R., Sarason, I. G., Sarason, B. R., Solky-Butzel, J. A., and Nagle, L. C. (1997). Assessing the quality of personal relationships. *J. Soc. Pers. Relat.* 14, 339–356. doi: 10.1177/0265407597143004
- Pierce, G. R., Sarason, I. G., and Sarason, B. R. (1991). General and relationship-based perceptions of social support: are two constructs better than one? *J. Pers. Soc. Psychol.* 61, 1028–1039. doi: 10.1037/0022-3514.61.6.1028
- Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G., and Levy, K. N. (2009). Initial construction and validation of the Pathological Narcissism Inventory. *Psychol. Assess.* 21, 365–379. doi: 10.1037/a0016530
- Pittman, J. F., Keiley, M. K., Kerpelman, J. L., and Vaughn, B. E. (2011). Attachment, identity, and intimacy: parallels between Bowlby's and Erikson's paradigms. *J. Fam. Theory Rev.* 3, 32–46. doi: 10.1111/j.1756-2589.2010.00079.x
- Raskin, R., and Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *J. Pers. Soc. Psychol.* 54, 890–902. doi: 10.1037/0022-3514.54.5.890
- Reiner, I., Beutel, M., Skaletz, C., Brähler, E., and Stöbel-Richter, Y. (2012). Validating the german version of the quality of relationship inventory: confirming the three-factor structure and report of psychometric properties. *PLoS One* 7:e37380. doi: 10.1371/journal.pone.0037380
- Reis, H. T., and Shaver, P. (1988). “Intimacy as an interpersonal process,” in *Handbook of Personal Relationships* ed. S. W. Duck (Oxford: Wiley & Sons), 367–389.
- Rubin, Z. (1970). Measurement of romantic love. *J. Pers. Soc. Psychol.* 16, 265–273. doi: 10.1037/h0029841
- Schütz, A., Marcus, B., and Sellin, I. (2004). Die Messung von narzissmus als persönlichkeitskonstrukt: psychometrische eigenschaften einer lang- und einer kurzform des deutschen NPI (Narcissistic Personality Inventory) [Measuring narcissism as a personality construct: psychometric properties of a long and a short version of the German Narcissistic Personality Inventory]. *Diagnostica* 50, 202–218. doi: 10.1026/0012-1924.50.4.202
- Séguin, M., Beauchamp, G., Robert, M., DiMambro, M., and Turecki, G. (2014). Developmental model of suicide trajectories. *Br. J. Psychiatry* 205, 120–126. doi: 10.1192/bjp.bp.113.139949
- Simon, R. W., and Marcussen, K. (1999). Marital transitions, marital beliefs, and mental health. *J. Health Soc. Behav.* 40, 111–125. doi: 10.2307/2676367
- Simpson, J. A. (1987). The dissolution of romantic relationships: factors involved in relationship stability and emotional distress. *J. Pers. Soc. Psychol.* 53, 683–692. doi: 10.1037/0022-3514.53.4.683
- Simpson, J. A., and Gangestad, S. W. (1991). Individual differences in sociosexuality: evidence for convergent and discriminant validity. *J. Pers. Soc. Psychol.* 60, 870–883. doi: 10.1037/0022-3514.60.6.870
- Simpson, J. A., and Gangestad, S. W. (1992). Sociosexuality and romantic partner choice. *J. Pers.* 60, 31–51. doi: 10.1111/j.1467-6494.1992.tb00264.x
- Simpson, J. A., Wilson, C. L., and Winterheld, H. A. (2004). “Sociosexuality and romantic relationships,” in *Handbook of Sexuality in Close Relationships*, eds J. H. Harvey, A. Wenzel and S. Sprecher (Mahwah: Erlbaum), 87–111.
- Spangenberg, L., Romppel, M., Bormann, B., Hofmeister, D., Brähler, E., and Strauß, B. (2013). A short version of the Narcissistic Personality Inventory (NPI-15): dimensionality and psychometric properties in a representative sample of the German population. *Psychother. Psychosom. Med. Psychol.* 63, 341–347.
- Sternberg, R. J. (1986). A triangular theory of love. *Psychol. Rev.* 93, 119–135. doi: 10.1037/0033-295X.93.2.119
- Sullivan, H. S. (1953). *The Interpersonal Theory of Psychiatry*. New York, NY: W W Norton & Co.
- Teo, A. R., Choi, H., and Valenstein, M. (2013). Social relationships and depression: ten-year follow-up from a nationally representative study. *PLoS One* 8:e62396. doi: 10.1371/journal.pone.0062396
- Trumpeter, N. N., Watson, P. J., O'Leary, B. J., and Weathington, B. L. (2008). Self-functioning and perceived parenting: relations of parental empathy and love inconsistency with narcissism, depression, and self-esteem. *J. Genet. Psychol.* 169, 51–71. doi: 10.3200/GNTP.169.1.51-71
- Vujeva, H. M., and Furman, W. (2011). Depressive symptoms and romantic relationship qualities from adolescence through emerging adulthood: a longitudinal examination of influences. *J. Clin. Child. Adolesc. Psychol.* 40, 123–135. doi: 10.1080/15374416.2011.533414
- Weinberger, M. I., Hofstein, Y., and Whitbourne, S. K. (2008). Intimacy in young adulthood as a predictor of divorce in midlife. *Pers. Relat.* 15, 551–557. doi: 10.1111/j.1475-6811.2008.00215.x
- Zayas, V., Mischel, W., Shoda, Y., and Aber, J. L. (2011). Roots of attachment: maternal caregiving at 18 months predicts adult peer and partner attachment. *Soc. Psychol. Personal. Sci.* 22, 376–383. doi: 10.1177/1948550610389822

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

Copyright © 2018 Kapusta, Jankowski, Wolf, Chéron-Le Guludec, Lopatka, Hammerer, Schnieder, Kealy, Ogrodniczuk and Blüml. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Italian Validation of the Capacity to Love Inventory: Preliminary Results

Giorgia Margherita^{1*}, Anna Gargiulo¹, Gina Troisi¹, Francesca Tessitore¹ and Nestor D. Kapusta²

¹ Department of Humanities, University of Naples Federico II, Naples, Italy, ² Department of Psychoanalysis and Psychotherapy, Medical University of Vienna, Vienna, Austria

OPEN ACCESS

Edited by:

Dorian A. Lamis,
Emory University School of Medicine,
United States

Reviewed by:

Emanuela Saita,
Università Cattolica del Sacro Cuore,
Italy
Giulio de Felice,
Sapienza Università di Roma, Italy

*Correspondence:

Giorgia Margherita
margheri@unina.it

Specialty section:

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

Received: 23 March 2018

Accepted: 23 July 2018

Published: 15 August 2018

Citation:

Margherita G, Gargiulo A, Troisi G,
Tessitore F and Kapusta ND (2018)
Italian Validation of the Capacity to
Love Inventory: Preliminary Results.
Front. Psychol. 9:1434.
doi: 10.3389/fpsyg.2018.01434

Introduction: Within a wider international research project aimed at operationalize the psychodynamic construct of capacity to love (Kernberg, 2011), the Capacity to Love Inventory (CTL-I) is a 41-items self-report questionnaire assessing six dimensions: interest in the life project of the other, basic trust, gratitude, common ego ideal, permanence of sexual passion, loss, and mourning.

Objectives: The study is aimed at validating the Italian version of the CTL-I.

Method: A total sample of 736 Italian non-clinical adults was administered a checklist assessing socio-demographic variables, and the CTL-I. A Confirmatory Factor Analyses (CFA) was conducted to examine the construct validity of the Italian version of the CTL-I. Only a part of the total sample (320 participants) was administered an additional series of concurrent measures in order to investigate the convergent validity of the CTL-I. Correlations with measures of socio-sexual orientation, quality of romance relations, and psychopathological questionnaires were examined through Pearson's correlation coefficients.

Results: CFA results suggested that the Italian CTL-I fully replicated the six-factor structure of the original CTL-I. Cronbach's alpha index provided satisfactory results for all subscales and the correlations with concurrent measures were in expected direction.

Conclusion: The results showed promising psychometric characteristics of the Italian version of CTL-I. Implications of the feasibility of the instrument in clinical and psychotherapeutic settings are discussed.

Keywords: capacity to love, capacity to love inventory, psychodynamic perspective, Italian validation, psychometric properties

INTRODUCTION

Starting from the Kernberg psychoanalytic theory (1995), the capacity to love is a human being's disposition to establish relationships with others and it is tightly connected to a person's psychic development. Thus, mature love, characterized by a set of dimensions, is considered as a theoretical frame with diagnostic potentialities. In fact, the incapacity to fall in love could be an important diagnostic marker in clinical contexts. For example, as highlighted by Kernberg (1974), narcissistic personalities have serious distortions in their internal relationships with others, demonstrating an incapacity to fall in love.

Previous studies have explored the associations between romantic relationships and depression highlighting that relationships of depressive people are characterized by greater dependency, emotional neediness and anger (Matussek et al., 1986; Fiske and Peterson, 1991). Moreover, negative qualities of romantic relationships seem have an important role in predicting depressive symptoms (La Greca and Harrison, 2005).

Building on the theory of capacity to love (Kernberg, 2011), Kapusta et al. (2018) have developed the Capacity to Love Inventory, which consists of 41 items with six dimensions (interest in the other, basic trust, gratitude, common ego ideal, permanence of sexual passion, loss, and mourning), with a good internal consistency in each subscale.

In the Italian context, instruments assessing love relationships have been developed from various theoretical perspectives: the attachment perspective (Gentili et al., 2002; Marazziti et al., 2010; Busonera et al., 2014; Carli et al., 2016), the developmental perspective (Ponti et al., 2010) and the psychosocial one (Donato et al., 2009; Boffo and Mannarini, 2015). The CTL-I adds a psychodynamic perspective into research of love relations and introduces new dimensions involved in romantic love.

The present study aims at validating the Italian version of the CTL-I (Kapusta et al., 2018) using a CFA methodology in an Italian sample. To assess the convergent validity of the CTL-I, correlations with some measures were examined through Pearson's correlation coefficients.

METHOD

Participants

The total sample was composed of 736 Italian non-clinical adults (488 women and 248 men). Mean age of the participants was 28.6 years ($SD = 9.8$).

The only inclusion criterion was the age of 18 and higher. The presence of conditions affecting the possibility of taking the assessment or the refusal of consent form was considered as exclusion criteria. Four hundred and sixteen participants were recruited through advertisements (*flyers* and newspapers distributed at the university and at public events). Before starting the surveys, participants signed a written consent form to participate the research. Instead, the remaining 320 participants were recruited through online *ads* posted at established community groups and through mailing lists. The online recruited subsample was asked to read a web page with the informed consent document and to accept it by clicking a button in order to start the research and complete the online survey. The whole study protocol was approved by the ethics committee of section of Psychology and Educational Sciences, University of Naples Federico II.

All the participants were administered the CTL-I. Out of the total of participants, only participants recruited through the online survey (320 participants: 249 women and 71 men) were administered a series of concurrent measures of depression, narcissism, socio-sexual orientation, quality of romance relations. Mean age of this subsample was 31.9 years ($SD = 11.08$). The relationship status of participants was as

TABLE 1 | Characteristics of the sample ($n = 736$).

Variables	Percentages (Mean \pm SD)
SEX	
Men	248
Women	488
Age – M(SD)	(28.6 \pm 9.8)
EDUCATION LEVEL	
Primary school	0.3%
Secondary school	5.8%
High school	37.5%
Graduated	51.3%
Post-graduate degree	0.5%
RELATIONSHIP STATUS	
In a relationship	65.7%
Single	28.2%
Married	3.7%
Divorced	0.8%
Widowed	0.7%

follows: 79.7% in a relationship, 17.8% single, 1.3% married, 1.3% divorced (Table 1).

Data Analysis Procedure

All the analyses were performed with the Statistical Package for the Social Sciences (SPSS) 19.0 and MPLUS7 software (Bentler, 1990; Schermelleh-Engel et al., 2003; Muthén and Muthén, 2012).

Confirmatory factor analysis on the Italian version of CTL-I was performed using the Maximum Likelihood (ML) as appropriate estimator.

We tested the theory driven model developed by Kapusta et al. (2018) by means of a CFA: a six-factor model with 41 items and scales being allowed to correlate with each other, in line with Kapusta et al. (2018) results. Model fit was assessed by means of the following fit indexes: (1) the chi-squared (χ^2) statistic and its degree of freedom; (2) the Standardized Root Mean Square Residuals (SRMR); and, (3) the Root Mean Square Error of Approximation (RMSEA) and its 90% confidence interval (90% CI). In line with what Schermelleh-Engel et al. (2003) affirmed, the model fit the data when χ^2/df equal or <2 , RMSEA equal or <0.05 (90% CI: the lower boundary of the CI should contain zero for exact fit and be <0.05 for close fit), although Browne and Cudek (1993) argued that values ranging from 0.05 to 0.08 are indicative of a good adequacy of the model.

Internal consistencies for the scales composing of the best fitting model were computed using Cronbach's coefficient alpha for each factor.

Measures

Capacity to Love Inventory (CTL-I)

The Capacity to Love Inventory (CTL-I) (Kapusta et al., 2018), translated through the back translation process (Van de Vijver and Leung, 1997), is a 41 items questionnaire rated on a 4 point Likert Scale and composed of six dimensions: interest in the life project of the other (INT), basic trust (BRT), gratitude

(GRT), common ego ideal (CEI), permanence of sexual passion (PSP) and loss and mourning (LOM). In the original version, Cronbach's alpha of the total scale is 0.90, whereas in the current study is 0.94; for the subscales values are 0.84 for Interest, 0.83 for Trust, 0.85 for Humility, 0.87 for Ego, 0.84 for Passion, and 0.85 for Acceptance.

Romance Qualities Scale (RQS)

The Romance Qualities Scale (RQS) (Ponti et al., 2010) is a self-report questionnaire which measures the qualitative aspects of romantic relationships. It consists of 22 items, evaluated on a 5-point Likert Scale, developed in five dimensions: Conflict, Companionship, Help, Security, Closeness. The internal consistency coefficient for the five subscales were: 0.74 and 0.75 for Conflict; 0.61 and 0.62 for Companionship; 0.82 and 0.84 for Help; 0.69 and 0.72 for Security; 0.74 and 0.78 for Closeness. In the present study, Cronbach's alphas were: 0.74 for Conflict; 0.49 for Companionship, 0.88 for Help, 0.73 for Security, 0.86 for Closeness, and 0.85 for the total scale.

Beck Depression Inventory (BDI-II)

The Beck Depression Inventory (BDI-II) is a 21 item self-report inventory that assess the presence and severity of depressive symptoms, according to DSM-IV (American Psychiatric Association, 1994) criteria. Statements regarding some feelings are rated on a 4-point Likert scale, based on the severity of depressive symptoms. The literature supported the inventory's psychometric properties in clinical and non-clinical samples (e.g., Arbisi, 2001; Balsamo and Saggino, 2007). In the current study, Cronbach's alpha was 0.88.

Socio-Sexual Orientation Inventory-Revised (SOI-R)

The Socio-sexual Orientation Inventory-Revised (SOI-R), in its Italian translation (Penke and Asendorpf, 2008), measures socio-sexual orientation assessing three facets of socio-sexuality: Past Behaviour as number of casual and changing sex partners, explicit Attitude toward uncommitted sex, sexual Desire. It is composed by nine self-report items, with a five-point Likert scale. Cronbach's alpha is high for Behaviour facet (0.84f–0.85m), Attitude (0.83f–0.87m), Desire (0.85 f–0.86m), Total (0.83). In the present study the Cronbach's alpha: 0.52 for Behaviour, 0.51 for Attitude, 0.80 for Desire, and 0.72 for the total scale.

Pathological Narcissism Inventory (PNI)

The Pathological Narcissism Inventory (PNI) (Pincus et al., 2009) is composed by 52-items rated on a six-point Likert scale. Items were translated into Italian according to procedures of back translation. It assesses the dimensions of narcissistic grandiosity and narcissistic vulnerability. In the present study the internal consistency coefficients were 0.74 for narcissistic grandiosity, 0.82 for narcissistic vulnerability, and 0.94 for the total scale.

Narcissistic Personality Inventory (NPI)

The Narcissistic Personality Inventory (NPI), in its Italian validation (Fossati et al., 2008) is characterized by 40 items and 7 sub-scales: Authorities, Exhibitionism, Superiority, Feeling in Law, Manipulation, Self-sufficiency, and Vanity. The internal consistency coefficients for the total of NPI (0.83) and for the sub-scale Authority (0.73) is good. In the present study the

Cronbach's alpha is 0.69 for Authorities, 0.67 for Exhibitionism, 0.52 for Superiority, 0.46 for Feeling in Law, 0.50 for Manipulation, 0.38 for Self-sufficiency, 0.60 for Vanity, and 0.81 for the total scale.

RESULTS

We tested a theory driven model by means of a CFA: a six-factor model with 41 items and scales being allowed to correlate with each other (Tables 2, 3).

Fit indices were: $\chi^2/\text{degrees of freedom} = 3.40$ (2598.870/764), $SRMR = 0.053$, $RMSE = 0.057$ (90% CI 0.055–0.060).

Based on the results of the χ^2 statistic, lack of overall fit for the model tested ($p < 0.001$) was shown, probably due to the sensitivity of this statistic to large sample sizes (Hu and Bentler, 1998; Kahn, 2006). In fact chi-square is highly sensitive to sample size: as the size of the sample increases, absolute differences become a smaller and smaller proportion of the expected value. The larger the sample, the larger and significant will be the chi squares, even with very small discrepancies among implied and obtained covariance matrices. On the other hand, samples of reduced size may be too prone to accept poor models (Type II error).

According to other goodness-of fit indices, that are RMSEA and SRMR, a good adequacy of the model was shown (Browne and Cudek, 1993; Hu and Bentler, 1999).

Results were very similar to those obtained by Kapusta et al. (2018) in the original version of CTL-I: $\chi^2/\text{degrees of freedom} = 3.13$ (2598.870/764), $SRMR = 0.060$, $RMSEA = 0.062$ (90% CI 0.060–0.065).

In order to investigate the potential associations among the six CTL-I subscales and other related measures, Pearson correlational analyses were computed (Table 4).

The subscales of the CTL-I show specific associations with the external correlated measures.

CTL-I *Interest* and CTL-I *Basic Trust* were positively correlated with the subscales of RQS related to Companionship, Help, Security and Closeness.

CTL-I *Gratitude* is positively correlated with the subscales of RQS Companionship, Help, Security, Closeness, and negatively correlated with the Conflict subscale of RQS and the Attitude and Desire subscales of SOI-R, and subscale Exhibitionism and Feeling in love of NPI.

CTL *Common Ego Ideal* is positively correlated with the Companionship, Help, Security, Closeness subscale of RQS and negatively correlated with Exhibitionism and Feeling in love of NPI.

CTL-I *Permanence of Sexual Passion* is negatively correlated with Companionship, Help, Security, Closeness subscales of RQS and depression (BDI-II) but positively correlates with Desire subscale of SOI-R.

CTL-I *Loss and Mourning* is positively correlated with Conflict subscale of RQS, with Narcissistic grandiosity and Narcissistic vulnerability of PNI, with the Exhibitionism, Superiority, Feeling in Law, Self-sufficiency subscale of PNI, and with depression of BDI-II. CTL-I *Loss and Mourning* is negatively

TABLE 2 | Factor loadings of 41-item CTL-I.

#	Item	INT	BTR	GRT	CEI	PSP	LOM
1	È importante conoscere il progetto di vita del mio partner.	0.702					
2	Condivido i progetti di vita con il mio partner.	0.641					
3	Sono felice di condividere i successi del mio partner.	0.875					
4	Mi arricchisce vedere la crescita personale del mio partner.	0.836					
5	Se il mio partner è infelice, mi sento triste.	0.528					
6	Capisco come essere comprensivo con il mio partner.	0.549					
7	Mi annoio spesso con il mio partner*						
8	Conto sul fatto che il mio partner sia comprensivo nei miei confronti se necessario.		0.572				
9	Il mio partner conosce le mie debolezze, i miei conflitti interiori e i miei problemi.		0.718				
10	Posso esprimere apertamente al mio partner sentimenti e bisogni personali.		0.793				
11	Conto sul fatto che il mio partner mi sostenga emotivamente nelle situazioni di incertezza.		0.751				
12	Sento di essere onesto con il mio partner.		0.591				
13	Nascondo dei segreti al mio partner*		0.452				
14	Posso confessare le mie debolezze al mio partner.		0.747				
15	A volte sento che ci sono dei limiti nei miei rapporti che non riesco a superare, nonostante gli sforzi fatti.*		0.220				
16	Sono a mio agio con il mio partner e di solito mi sento al sicuro.		0.698				
17	Avverto un senso di gratitudine per il fatto che il mio partner ci sia.			0.699			
18	Sono grato per l'amore ricevuto.			0.688			
19	Anche se siamo lontani per molto tempo continuo a sentirmi legato al mio partner.			0.698			
20	Riconosco di aver bisogno del mio partner.			0.693			
21	Mi piace confortare il mio partner.			0.640			
22	Mi piace prendermi cura del mio partner, quando ha bisogno del mio aiuto.			0.643			
23	Condivido le responsabilità del quotidiano per sollevare il mio partner.			0.679			
24	Mi impegno nelle relazioni.				0.706		
25	Con il mio partner cerchiamo di lavorare sul nostro rapporto.				0.661		
26	Ho rispetto per la personalità e i valori del mio partner.				0.748		
27	Adoro guardare i gesti e le reazioni del mio partner.				0.580		
28	Mi sento impegnato nella nostra vita comune.				0.617		
29	Cerco possibili punti di incontro nei casi di conflitti e prospettive differenti.				0.788		
30	Dico spesso al mio partner che lo amo.				0.767		
31	Mi sento profondamente legato al mio partner.				0.616		
32	In un rapporto a lungo termine ad un certo punto sopraggiunge la noia nella sfera sessuale.*					0.836	
33	Col passare del tempo diminuisce il desiderio sessuale.*					0.859	
34	È difficile accettare che la persona che amo non corrisponda il mio amore.*						0.606
35	Se un rapporto finisce, tendo ad attribuire la colpa al mio ex-partner.*						0.682
36	Provo desideri di vendetta se il mio partner mi lascia.*						0.683
37	Spesso ho difficoltà ad accettare la fine delle mie relazioni.*						0.693
38	Sono spesso geloso.*						0.577
39	Provo spesso sensi di colpa dopo una separazione.*						0.603
40	Se il partner mi ha abbandonato tendo a svalutarmi.*						0.673
41	È difficile essere coinvolto nuovamente in un rapporto dopo la fine di una relazione.*						0.581

*Reverse-coded item. INT, Interest; BTR, Basic Trust; GRT, Gratitude; CEI, Common Ego Ideal; PSP, Permanence of Sexual Passion; LOM, Loss and Mourning.

correlated with Companionship, Help, Security, of RQS and Self-sufficiency subscale of NPI.

DISCUSSION

The present study showed that the Italian version of the CTL-I exhibited a clear factor structure and good psychometric properties. Indeed, Cronbach's alphas for all subscales were good, as well as its convergent validity.

The CFS's results showed a fully satisfactory fit. The dimensions emerging from this analysis showed superimposable

to the ones obtained by Kapusta et al. (2018) in their original study. Thus, it is possible to conclude that the Italian CTL-I displayed a good psychometric functioning, similar to the one shown by the inventory's original version.

From our analysis it emerged that the CTL-I measures the capacity to love in a romantic relationship, as demonstrated by the strong positive and significant correlation with the RQS. Concerning the capability of measuring dimensions of psychopathology, the dimension of *CTL-I Loss and Mourning* strongly showed a moderate negative correlation with the subscales of the PNI and with the BDI-II, showing the possibility that capacity to love displays a protective function against the emergence of some symptomatic expressions.

Although we did not administer to a clinical sample, we believe that the instrument has some important clinical implications. First of all, the negative correlation of the CTL-I with the clinical scales should confirm the important diagnostic potentialities of the capacity to love (Gargiulo et al., 2014).

Moreover, considering the need to develop dynamic approaches to study the psychotherapeutic processes and outcomes (Bateman and Fonagy, 2012; Gelo and Salvatore, 2016; Esposito et al., 2017), the CTL-I could be a useful instrument for monitoring the efficacy of clinical intervention or their process-outcome connections. From a perspective which considers therapeutic change as an integration between the reorganization of affects and a different representation of objects relationships

TABLE 3 | Correlation between CTL-I factors.

	INT	BTR	GRT	CEI	PSP	LOM
INT	-	0.771**	0.657**	0.667**	0.151**	0.385**
BTR		-	0.689**	0.668**	0.184**	0.275**
GRT			-	0.939**	0.303**	0.116*
CEI				-	0.349**	0.098*
PSP					-	0.244**
LOM						-

***p* < 0.01, **p* < 0.05. INT, Interest; BTR, Basic Trust; GRT, Gratitude; CEI, Common Ego Ideal; PSP, Permanence of Sexual Passion; LOM, Loss and Mourning.

TABLE 4 | Pearson correlations between CTL-I Italia version and concurrent measures.

	CTL-I						
	Interest	Basic trust	Gratitude	Common ego ideal	Permanence of sexual passion	Loss and mourning	Total scale
RQS	0.154**	0.247**	0.259**	0.198**	-0.209**	-0.072	0.196**
Conflict	-0.098	-0.090	-0.126*	-0.176	0.078	0.213**	0.066
Companionship	0.174**	0.219**	0.272**	0.246**	-0.246**	-0.134*	0.188**
Help	0.165**	0.274**	0.298**	0.235**	-0.216**	-0.119*	0.214**
Security	0.165**	0.234**	0.221**	0.218**	-0.175**	-0.182**	0.160**
Closeness	0.146**	0.216**	0.277**	0.231**	-0.210**	-0.032	0.214**
SOI-R	-0.125	-0.019	-0.168**	-0.111*	0.139*	-0.011	-0.105
Behaviour facet	-0.96	-0.023	-0.100	-0.051	-0.059	-0.050	-0.050
Attitude	-0.142*	-0.035	-0.122*	-0.106	0.093	0.062	-0.118*
Desire	-0.063	-0.006	-0.164**	-0.092	0.150**	0.015	-0.070
PNI	-0.041	0.012	0.009	-0.051	0.086	0.555*	0.164**
Narcissistic grandiosity	0.014	0.094	0.041	-0.030	0.029	0.418**	0.165**
Narcissistic vulnerability	-0.060	-0.026	-0.003	-0.050	0.101	0.563**	0.150**
NPI	-0.038	-0.011	-0.075	-0.061	-0.002	0.072	-0.031
Authorities	-0.015	0.055	-0.056	-0.071	-0.009	0.028	0.020
Exhibitionism	-0.072	-0.099	-0.121*	-0.132*	0.025	0.182**	0.063
Superiority	-0.013	0.001	-0.001	-0.006	0.054	0.112*	-0.030
Feeling in Law	-0.095	0.049	-0.154**	-0.124*	0.070	0.264**	0.003
Manipulation	0.027	0.034	-0.022	-0.029	0.018	0.003	0.002
Self-sufficiency	0.043	-0.007	-0.018	0.072	-0.111	-0.195**	-0.049
Vanity	-0.030	-0.058	-0.057	-0.074	-0.035	0.056	-0.051
BDI-II	-0.074	-0.100	-0.131*	-0.096	-0.137*	0.347**	0.011

***p* < 0.01, **p* < 0.05 RQS, Romance Qualities Scale; SOI-R, Socio-sexual Orientation Inventory-Revised; PNI, Pathological Narcissism Inventory; NPI, Narcissistic Personality Inventory; BDI-II, Beck Depression Inventory.

(Kernberg, 1993; De Luca Picione et al., 2018), we believe that the scale could catch the processes of therapeutic change of a clinical intervention which should aim at reorganizing the psychic mental set of a patient as a function of new relationship modalities.

Regarding the level of the clinical relationship, we believe that the CTL-I could be a useful indicator in order to comprehend the quality of the transfert.

Finally, due to its flexibility we believe that the instrument could be also able to individuate risk and protective dimensions in non-psychotherapeutic clinical contexts (Carlino and Margherita, 2016; Margherita et al., 2017; Tessitore and Margherita, 2017).

This study is not free from limitations. One of the limitations of the study is the small number of men in the samples; this could be explained hypothesizing that men may be less interested in romantic relationships compared to women (Fralely et al., 2011). Future investigations based on gender differences and couple relationships are needed. Another limitation regards the recruitment of the sample, unbalanced between online and offline. Moreover, considering the capacity to love as a non-stable characteristic, future research also needs to consider it depending on the relationships.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the ethical guidelines for research

draft by the Italian Psychologists Association and the National Psychologists Council. The ethical guidelines of Helsinki Declaration were followed and participants were informed about the confidentiality of the data and the treatment. The protocol was approved by the Local Ethical Committee for research in Psychology.

AUTHOR CONTRIBUTIONS

All the authors listed have made a substantial contribution to the work. GM developed the theoretical framework of the present study, designed the research project and contributed to the scientific supervision of the study. AG contributed to the methodological approach, performed the translation procedures of the instrument and the data collection and wrote the manuscript. GT performed all the analysis and designed tables and figures. FT specifically wrote the Introduction and revised the manuscript. NK developed the theoretical framework of the whole project, operationalized the construct of the capacity to love and contributed to the scientific supervision of the whole work. All authors discussed the results, commented the manuscript and gave the final approval of the work.

ACKNOWLEDGMENTS

The authors warmly acknowledge Cristina Corsini for her precious help in the processes of data collection and its analysis.

REFERENCES

- American Psychiatric Association (1994). *DSM-IV (Vol. 1)*. Washington, DC: American Psychiatric Pub.
- Arbisi, P. A. (2001). "Review of the beck depression inventory-II," in *The 14th Mental Measurements Yearbook*, eds B. S. Plake and J. C. Impara (Lincoln, NE: Buros Institute of Mental Measurements), 452–468.
- Balsamo, M., and Saggino, A. (2007). Test per l'assessment della depressione nel contesto italiano: Un'analisi critica. [Tests for the assessment of depression in an Italian context: a critical review]. *Psicoter. Cogn. Comport.* 13, 167–199.
- Bateman, A. W., and Fonagy, P. (2012). *Handbook of Mentalizing in Mental Health Practice*. Arlington, TX: American Psychiatric Publishing, Inc.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychol. Bull.* 107, 238–246. doi: 10.1037/0033-2909.107.2.238
- Boffo, M., and Mannarini, S. (2015). Assessing decision-making in romantic relationships: a first italian validation of the relationship deciding scale. *TPM: Test. Psychometr. Methodol. Appl. Psychol.* 22, 415–427. doi: 10.4473/TPM22.3.7
- Browne, M. W., and Cudek, R. (1993). "Alternative ways of assessing model fit," in *Testing Structural Equation Models*, eds K.A. Bollen, and J.S. Long (Newbury Park, CA: Sage Focus Editions), 136–162.
- Busonera, A., Martini, P. S., Zavattini, G. C., and Santona, A. (2014). Psychometric properties of an Italian version of the experiences in close relationships-revised (ECR-R) scale. *Psychol. Rep.* 114, 785–801. doi: 10.2466/03.21.PR0.114k23w9
- Carli, L. L., Anzelmo, E., Gatti, E., Santona, A., Pozzi, S., and Gallucci, M. (2016). The family-couple-parenting questionnaire: development of a measure for long-term couples and young adults. *Psychol. Rep.* 118, 957–993. doi: 10.1177/0033294116639803
- Carlino, M., and Margherita, G. (2016). La maternità in donne con Sclerosi Multipla: Una ricerca esplorativa sulle possibili dimensioni di benessere nella malattia. *Psicologia Della Salute* 3, 5–25. doi: 10.3280/PDS2016-003001
- De Luca Picione, R., Martino, M. L., and Freda, M. F. (2018). Modal articulation: the psychological and semiotic functions of modalities in the sensemaking process. *Theory Psychol.* 28, 84–103. doi: 10.1177/0959354317743580
- Donato, S., Iafrate, R., Barni, D., and Bertoni, A. (2009). Measuring dyadic coping: the factorial structure of Bodenmann's "Dyadic Coping Questionnaire" in an Italian sample. *Test. Psychomet. Methodol. Appl. Psychol.* 16, 25–47.
- Espósito, G., Ribeiro, A. P., Gonçalves, M. M., and Freda, M. F. (2017). Mirroring in group counseling: analyzing narrative innovations. *Small Group Res.* 48, 391–419. doi: 10.1177/1046496417697149
- Fiske, V., and Peterson, C. (1991). Love and depression: the nature of depressive romantic relationships. *J. Soc. Clin. Psychol.* 10, 75–90.
- Fossati, A., Borroni, S., and Maffei, C. (2008). Proprietà Psicometriche della Versione Italiana del Narcissistic personality inventory [Psychometric properties of the Italian version of the Narcissistic Personality Inventory]. *Rivista Psicol. Clin.* 1, 96–115.
- Fralely, R. C., Heffernan, M. E., Vicary, A. M., and Brumbaugh, C. C. (2011). The experiences in close relationships-relationship structures questionnaire: a method for assessing attachment orientations across relationships. *Psychol. Assess.* 23:615. doi: 10.1037/a0022898
- Gargiulo, A., Plener, P. L., Baus, N., Margherita, G., Brunner, R., Kaes, M., et al. (2014). Autolesività non suicidaria (NSSI) e Disturbo da Comportamento Suicidario (SBD) nella recente pubblicazione del DSM-5. [Non-suicidal self-injury (NSSI) and suicidal behavior disorder (SBD) in DSM-5]. *M. Psichiatr.* 55, 83–90.
- Gelo, O. C., and Salvatore, S. (2016). A dynamic systems approach to psychotherapy: a meta-theoretical framework for explaining psychotherapy change processes. *J. Couns. Psychol.* 63, 379–395. doi: 10.1037/cou0000150
- Gentili, P., Contreras, L., Cassaniti, M., and D'arista, F. (2002). A measurement of dyadic adjustment: the dyadic adjustment scale. *M. Psichiatr.* 43, 107–116.

- Hu, L. T., and Bentler, P. M. (1998). Fit indices in covariance structure modeling: sensitivity to under parameterized model misspecification. *Psychol. Methods* 3, 424–453. doi: 10.1037//1082-989X.3.4.424
- Hu, L. T., and Bentler, P. M. (1999). Cut off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equation Model. Multidiscip. J.* 6, 1–55.
- Kahn, J. H. (2006). Factor analysis in counseling psychology research, training, and practice: principles, advances, and applications. *Couns. Psychol.* 34, 684–718. doi: 10.1177/0011000006286347
- Kapusta, N. D., Jankowski, K. S., Wolf, V., Chéron-Le Guludec, M., Lopatka, M., Hammerer, C., et al. (2018). Measuring the capacity to love: development of the CTL-inventory. *Front. Psychol.* 9:1115. doi: 10.3389/fpsyg.2018.01115
- Kernberg, O. F. (1974). Barriers to falling and remaining in love. *J. Am. Psychoanal. Assoc.* 22, 486–511.
- Kernberg, O. F. (1993). *Severe Personality Disorders: Psychotherapeutic Strategies*. New Haven, CT: Yale University Press.
- Kernberg, O. F. (1995). *Object Relations Theory and Clinical Psychoanalysis*. Lanham, MD: Jason Aronson.
- Kernberg, O. F. (2011). Limitations to the capacity to love. *Int. J. Psychoanal.* 92, 1501–1515. doi: 10.1111/j.1745-8315.2011.00456.x
- La Greca, A. M., and Harrison, H. M. (2005). Adolescent peer relations, friendships, and romantic relationships: do they predict social anxiety and depression?. *J. Clin. Child Adolesc. Psychol.* 34, 49–61. doi: 10.1207/s15374424jccp3401_5
- Marazziti, D., Sbrana, A., Rucci, P., Cherici, L., Mungai, F., Gonnelli, C., et al. (2010). Heterogeneity of the jealousy phenomenon in the general population: an Italian study. *CNS Spectr.* 15, 19–24. doi: 10.1017/S1092852900000262
- Margherita, G., Troisi, G., Tessitore, F., and Gargiulo A. (2017). Teen mothers who are daughters of teen mothers: Psychological intergenerational dimensions of early motherhood. *Child. Youth Serv. Res.* 83, 294–301. doi: 10.1016/j.childev.2017.10.010
- Matussek, P., Luks, O., and Seibt, G. (1986). Partner relationships of depressives. *Psychopathology* 19, 143–156.
- Muthén, L. K., and Muthén, B. O. (2012). *Mplus Statistical Modeling Software: Release 7.0*. Los Angeles, CA: Muthén and Muthén.
- Penke, L., and Asendorpf, J. B. (2008). Beyond global sociosexual orientations: a more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *J. Pers. Soc. Psychol.* 95:1113. doi: 10.1037/0022-3514.95.5.1113
- Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G., and Levy, K. N. (2009). Initial construction and validation of the pathological narcissism inventory. *Psychol. Assess.* 21, 365–379. doi: 10.1037/a0016530
- Ponti, L., Guarnieri, S., Smorti, A., and Tani, F. (2010). A measure for the study of friendship and romantic relationship quality from adolescence to early-adulthood. *Open Psychol. J.* 3, 76–87. doi: 10.2174/1874350101003010076
- 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. Online* 8, 23–74.
- Tessitore, F., and Margherita, G. (2017). A review of asylum seekers and refugees in Italy: where is the psychological research going? *Mediterr. J. Clin. Psychol.* 5. doi: 10.6092/2282-1619/2017.5.1612
- Van de Vijver, F. J., and Leung, K. (1997). *Methods and Data Analysis for Cross-Cultural Research*, Vol. 1. Thousand Oaks, CA: Sage Publications.

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

Copyright © 2018 Margherita, Gargiulo, Troisi, Tessitore and Kapusta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Intolerance of Uncertainty Inventory: Validity and Comparison of Scoring Methods to Assess Individuals Screening Positive for Anxiety and Depression

Marco Lauriola^{1*}, Oriana Mosca^{1,2}, Cristina Trentini², Renato Foschi², Renata Tambelli² and R. Nicholas Carleton³

OPEN ACCESS

Edited by:

Marco Innamorati,
Università Europea di Roma, Italy

Reviewed by:

Caterina Primi,
Università degli Studi di Firenze, Italy
Dejan Stevanovic,
Clinic for Neurology and Psychiatry
for Children and Youth, Serbia

Marta Ghisi,
Università degli Studi di Padova, Italy

*Correspondence:

Marco Lauriola
marco.lauriola@uniroma1.it

Specialty section:

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

Received: 21 January 2018

Accepted: 08 March 2018

Published: 26 March 2018

Citation:

Lauriola M, Mosca O, Trentini C,
Foschi R, Tambelli R and
Carleton RN (2018) The Intolerance
of Uncertainty Inventory: Validity and
Comparison of Scoring Methods to
Assess Individuals Screening Positive
for Anxiety and Depression.
Front. Psychol. 9:388.
doi: 10.3389/fpsyg.2018.00388

¹ Department of Social and Developmental Psychology, Sapienza University of Rome, Rome, Italy, ² Department of Dynamic and Clinical Psychology, Sapienza University of Rome, Rome, Italy, ³ Department of Psychology, University of Regina, Regina, SK, Canada

Intolerance of Uncertainty is a fundamental transdiagnostic personality construct hierarchically organized with a core general factor underlying diverse clinical manifestations. The current study evaluated the construct validity of the Intolerance of Uncertainty Inventory, a two-part scale separately assessing a unitary Intolerance of Uncertainty disposition to consider uncertainties to be unacceptable and threatening (Part A) and the consequences of such disposition, regarding experiential avoidance, chronic doubt, overestimation of threat, worrying, control of uncertain situations, and seeking reassurance (Part B). Community members ($N = 1046$; Mean age = 36.69 ± 12.31 years; 61% females) completed the Intolerance of Uncertainty Inventory with the Beck Depression Inventory-II and the State-Trait Anxiety Inventory. Part A demonstrated a robust unidimensional structure and an excellent convergent validity with Part B. A bifactor model was the best fitting model for Part B. Based on these results, we compared the hierarchical factor scores with summated ratings clinical proxy groups reporting anxiety and depression symptoms. Summated rating scores were associated with both depression and anxiety and proportionally increased with the co-occurrence of depressive and anxious symptoms. By contrast, hierarchical scores were useful to detect which facets mostly separated between for depression and anxiety groups. In sum, Part A was a reliable and valid transdiagnostic measure of Intolerance of Uncertainty. The Part B was arguably more useful for assessing clinical manifestations of Intolerance of Uncertainty for specific disorders, provided that hierarchical scores are used. Overall, our study suggest that clinical assessments might need to shift toward hierarchical factor scores.

Keywords: intolerance of uncertainty, Intolerance of Uncertainty Inventory, confirmatory factor analysis, bifactor model, clinical validity, anxiety, depression, transdiagnostic

INTRODUCTION

Uncertainty can be a significant psychological and physiological stressor. Difficulties with uncertainty have been associated with ineffective coping, neuroticism, need for predictability, and cognitive reactions to ambiguity (e.g., rigid dichotomizing into fixed categories, seeking certainty, and resorting to “black-white solutions”) (Berenbaum et al., 2008; Rosen et al., 2014; Lauriola et al., 2015; McEvoy and Erceg-Hurn, 2015; Carleton, 2016b). Intolerance of Uncertainty (IU) is an “individual’s dispositional incapacity to endure the aversive response triggered by the perceived absence of salient, key, or sufficient information, and sustained by the associated perception of uncertainty” (Carleton, 2016b, p. 31). IU is a latent multidimensional construct, reflecting fear of the unknown (Hong and Cheung, 2015; Carleton, 2016a). Substantial evidence indicates IU is a transdiagnostic factor for diverse psychopathology (Carleton et al., 2012; Mahoney and McEvoy, 2012; Einstein, 2014; Carleton, 2016a), with higher scores in clinical populations across disorders (Holaway et al., 2006; Gentes and Ruscio, 2011; Sternheim et al., 2011) and proportionate increases with comorbidity (Holaway et al., 2006; Yook et al., 2010; McEvoy and Mahoney, 2011).

The *Intolerance of Uncertainty Inventory* (IUI; Gosselin et al., 2008; Carleton et al., 2010) is a new comprehensive IU scale. Different from other IU scales, the IUI is comprised of two sets of items that can be administered together or separately. The first set (IUI-A; General Unacceptability of Uncertainty) was developed to assess core beliefs about IU as currently defined (Carleton, 2016b). Accordingly, IUI-A items were devised as a coherent set of statements tapping into the tendency for the person to consider uncertainties in life to be unacceptable and threatening (e.g., “Not knowing what will happen in advance is often unacceptable for me”). Importantly, these beliefs were added later to the theoretical definition of the IU construct and were not specifically addressed in the classic IUS scales (Carleton, 2016b). The second set of items (IUI-B; Negative Manifestations of Uncertainty) was devised to cover six specific consequences of IU, which are common to observe in clinical patients, across different affective disorders.

Worrying may be the most common IU consequence included in the IUI-B (e.g., “Uncertain situations worry me”). Patients with GAD report ongoing worry helps them prepare to cope with unpredictable negative events (Dugas et al., 1998; Newman et al., 2013). High IU potentiates *overestimation of threat* operatively defined in the IUI-B as the tendency to exaggerate the probability that a negative event will occur (e.g., “In an uncertain situation, I tend to exaggerate the chances that things may go badly”). Chronic IU is associated with doubt, a hallmark feature of Obsessive Compulsive Disorder (OCD) (Nikodijevic et al., 2015; Samuels et al., 2017); accordingly, the IUI-B includes *doubting* items to assess absent confidence in thoughts, judgments, actions, and feelings (e.g., “When I am uncertain, I tend to doubt my capabilities”). Patients with GAD and OCD report desires to control uncertainty and therein defuse short-term anxiety and discomfort (e.g., compulsions in OCD, safety behaviors in GAD); as such, the

IUI-B includes items assessing *need for control* (e.g., “I prefer to control everything in order to decrease uncertainties”). When worrying and control are insufficient, high IU may cause *reassurance seeking* from others or authoritative sources, as measured by the IUI (e.g., “When I am uncertain, I need to be reassured by others”); paradoxically, seeking reassurance can maintain anxiety symptoms over time (e.g., Kobori and Salkovskis, 2013). Finally, patients with high IU may engage in *avoidance* to cope, which typically produces only short-term reductions in anxiety (Sexton and Dugas, 2008; Mahoney et al., 2016). The IUI-B avoidance items assess attempts to escape uncertainty (e.g., “I tend not to engage in activities involving some uncertainty”).

Sound methods for separately assessing IU core beliefs and the clinical consequences of IU are useful for ascribing the positive consequences of clinical interventions to changes in beliefs as well as identifying specific targets to prioritize in clinical practice. Nevertheless, the IUI has not been extensively used in clinical research, nor has the IUI factorial structure been cross validated beyond north-American borders. Existing evidence generally supported a unidimensional structure for the IUI-A, and a six-factor structure for the IUI-B reflecting the aforementioned consequences of IU (Gosselin et al., 2008; Carleton et al., 2010). However, these findings were not unequivocal. Although a unidimensional structure was acceptable, the first study (Gosselin et al., 2008) concluded that a three-factor structure [(I) intolerance of uncertainty and uncertain situations; (II) intolerance of the unexpected; (III) difficulty waiting in an uncertain situation] best represented the IUI-A. Regarding IUI-B, the same study showed that the hypothesized six-factor structure [(I) avoidance; (II) doubt; (III) overestimation; (IV) worry; (V) control; and (VI) reassurance] was an excellent fit to the data. The second study (Gosselin et al., 2008) showed that the fit indices for the IUI-A were unacceptable both for the unidimensional factor model and for the multifactor model. The unitary factor model was trimmed based on the modification indices, and the atheoretical removal of items #2, #9, and #13 improved the model fit. The same study also showed that the fit indices supported the six-factor model for the IUI-B, but did not meet the acceptable standards (Carleton et al., 2010). As a whole, these results underscore the need for a cross-validation study of IUI factors on independent samples in different languages as well as for some psychometric refinements of the IUI scoring system.

The current study was primarily designed to assess the factor structure of the IUI-A and IUI-B using the models proposed in the extant literature as well as testing new hierarchical models for the IUI-B. For the IUI-A, we started with testing a unidimensional model to replicate the overall tendency for IU core beliefs items to reflect a unitary core dimension, and then followed up this analysis to assess the impact of removing critical items, as proposed in the literature (Carleton et al., 2010). For the IUI-B, previous research did not find recognizable solutions within modification indices. Nevertheless, hierarchical factor models were not fitted to the IUI-B item set, although this class of models is more appropriate to represent multifaceted personality constructs. First, we proposed a second-order factor

model in which a general IU factor influences item responses through the six IUI-B first-order factors. Theoretically, the second-order model assumes that general IU (i.e., “a latent fear of the unknown”; Carleton, 2016a) will not directly influence the behavioral manifestations of IU; instead, general IU effects are expected to be mediated by more proximal first-order factors (i.e., avoidance, doubting, overestimation, worrying, control, and reassurance). Second, we assessed a bifactor model in which a general IU factor does directly influence IUI-B items above and beyond the more proximal more proximal first-order factors.

Multifaceted scales should also be assessed for the relative utility of the total and subscale scores in clinical assessments. General and specific variance proportions are variably entangled, complicating the extent to which clinical groups may differ on a general trait (e.g., ‘a latent fear of the unknown’ for IU multidimensional assessment scales) or on a specific manifestation of that trait (e.g., ‘overestimation of threat,’ ‘need for control’). Hierarchical factor models offer clinical researchers an opportunity to derive factor scores that parse general and specific variance (Reise et al., 2010; Chen et al., 2012). The current study compares aggregated IUI scores and hierarchical factor scores for assessing individuals screening positive for anxiety and depressive disorders, which are highly comorbid and critically associated with IU (Miranda et al., 2008; Carleton et al., 2012; Mahoney and McEvoy, 2012; Carleton, 2016a). Comparing scoring methods may provide insights for the co-occurrence of depressive and anxious symptoms. In clinical groups, elevated subscale scores may be due to higher general distress rather than IU-specific mechanisms. Profile elevations across subscales may be due to entanglement with general IU factor variance. Accordingly, we hypothesized that IUI summated ratings might produce divergent response patterns between participants who were above the clinical cut-offs for anxiety and depression and those who were not (Holaway et al., 2006; Gentes and Ruscio, 2011; Sternheim et al., 2011), and aggregated ratings proportionally increase with the co-occurrence of depressive and anxious symptoms (Holaway et al., 2006; Yook et al., 2010; McEvoy and Mahoney, 2011). By contrast, we expect some divergent response patterns between the two groups using hierarchical factor scores. For example, some scores (e.g., worry, doubting) might best characterize individuals screening positive for anxiety disorders, while other scores (e.g., overestimation of threat) might best characterize those individuals screening positive for depression.

MATERIALS AND METHODS

Participants and Procedures

The sample was based on convenience rather than randomly drawn from a target population; nevertheless, approximate quotas were set for age, gender and education to ensure heterogeneous sampling. Participants included 1046 community members (414 men, 627 women, 5 undisclosed gender) who completed a series of self-report measures as part of a larger

study approved by the local ethical review board for psychological research. Participant ages ranged from 20 to 76 years ($M = 36.69$; $SD = 12.31$). Completed education levels were distributed as follows: senior high school ($N = 454$; 43.5%), junior high school ($N = 464$; 44.5%), and elementary school ($N = 125$; 12.0%). Eighty-nine undergraduate psychology students attending an advanced clinical assessment class were asked to recruit research participants among their acquaintances and to serve as interviewers. The third author of this paper trained all the students for standardization of questionnaire administration in small group sessions. Before data entry, the third author debriefed the students and verified the accuracy of the collected data. No special problems were encountered but sporadic missing data. Other psychology students or close family members of the recruiter were excluded from the study. The questionnaires were administered at home in a quiet and comfortable room. Each interviewer acquainted potential participants with the study goals, the voluntary nature of participation, the right to withdraw from the study at any moment, and that responses would be kept anonymous once submitted. Verbal consent was obtained from each participant before data collection. The data were collected over a 3-week period, and each interviewer collected a variable number of cases (ranging from 6 to 31) on a voluntary base.

Measures

Intolerance of Uncertainty Inventory

Participants completed the 45-item version of the IUI (Gosselin et al., 2008), containing 15 items for IUI-A and 30 items for IUI-B. The IUI items were translated into Italian by the first and the second author for use in the current study. Then, a bilingual professional translator, without reference to the original text, back-translated the IUI into English to verify linguistic equivalence. Minor discrepancies between translations were resolved through discussion. Following Gosselin et al. (2008), the items were administered using a 5-point Likert scale ranging from 1 (‘not at all characteristic of me’) to 5 (‘entirely characteristic of me’). The Italian version of the IUI items is reported in the Supplementary Table 1. Scoring key and descriptive statistics are reported in the Supplementary Table 2.

Beck Depression Inventory II

The Beck Depression Inventory II (BDI-II; Beck et al., 1996; Italian version, Ghisi et al., 2006) is a 21-item multiple-choice self-report scale. The BDI-II was designed to assess affective, somatic, or cognitive symptoms of depression. Respondents task was to rate the severity of each symptom using a 4-point Likert scale ranging from 0 to 3 (higher numbers indicated greater severity). The total score ($\alpha = 0.89$, in the present study) is a valid measure of the severity of depression. Total scores of 0–13 indicates ‘minimal or no depression.’ Total scores ranging from 14–19, 20–28, and 29–63 are used to classify participants as reporting ‘mild,’ ‘moderate,’ and ‘severe’ depression levels, respectively.

State-Trait Anxiety Inventory

State-Trait Anxiety Inventory (STAI-Y; Spielberger, 1983; Italian version, Pedrabissi and Santinello, 1989) is a 40-item self-report

measure designed to assess transitory and chronic anxiety symptoms. The 20-item trait subscale was used in the present study. The total score ($\alpha = 0.93$, in the present study) is considered a valid measure of trait neuroticism, that is the tendency to chronically experience a wide range of negative affect states (e.g., fear, worry, autonomic nervous system somatic symptoms). The recommended clinical cut-off score for the A-trait total score is > 46 (Fisher and Durham, 1999), which is how the A-trait scale was used to separate high trait anxious individuals from the rest of the sample.

Data Analysis

Missing Values

Sporadic missing values were imputed using a random Hot Deck (Andridge and Little, 2010). Accordingly, we replaced each missing value in an item with an individual response from a similar case picked at random from those in the dataset (i.e., same age, gender, and education).

CFA Models

Structural equation modeling (EQS 6.2; Bentler, 2004) was used to assess the factorial structure of the IUI. Separate analyses were carried out for IUI-A and IUI-B. The data deviated from the assumptions of multivariate normality (i.e., Mardia's normalized coefficient = 46.18 and 113.85, respectively, for the IUI-A and IUI-B datasets); accordingly, the Maximum Likelihood Robust method (MLR) was used to adjust model parameters and fit. In line with previous research (Gosselin et al., 2008; Carleton et al., 2010), we tested single-factor, two-factor, and three-factor models for the IUI-A, as well as three-factor and six-factor models for the IUI-B. Recent evidence and theory have suggested that IU was best modeled as hierarchical multifaceted construct (Hale et al., 2016; Lauriola et al., 2016); accordingly, we also tested second-order and bifactor models for the IUI-B, in which a general IU factor loaded all items, while six independent group factors loaded on avoidance, doubt, overestimation, worry, control and reassurance items, respectively.

Assessment of Model Fit

Model fit was assessed using the following indices: Satorra-Bentler scaled χ^2 ($SB\chi^2$), robust versions of Comparative Fit Index (CFI), Bentler-Bonnett Non-Normed Fit Index (NNFI), Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR). According to Hu and Bentler (1999), cutoff values close to 0.95 for NNFI and CFI, close to 0.06 for RMSEA, and close to 0.08 for SRMR are needed to conclude that there is a relatively good fit between the factor model and the data.

Model Comparisons

Nested factor models are models that can be derived one from the other by estimating fewer parameters. For example, a single factor model is nested in a two-factor model that insists on the same observed variables, so that the former can be obtained from the latter by constraining the correlation

between the two latent variables to 1.00 (i.e., the single factor model has one parameter less than the two-factor model). Nested models can be compared statistically with a chi-square difference test to assess whether the model restrictions significantly impacted fit. For comparisons that are not statistically significant the more restrictive model is preferred. Conversely, the less restrictive model is preferred for statistically significant comparisons. In using MLR for the current study the chi-square difference test was corrected per Satorra and Bentler (2001) formula.

Non-nested models that insist on different subsets of observed variables can also be compared (e.g., dropping items with poor fit from subsequent CFA analyses) using 'information criteria' indices that adjust the ML fit functions based on the number of parameters. The Consistent Akaike Information Criterion (CAIC; Bozdogan, 1987) is considered the preferred index for such analyses but has no intuitive value for interpretation and no recommended cut-off scores. Lower CAICs are associated with a higher likelihood that the tested model approximates the 'true' model, thereby having greater chances to be replicated in subsequent cross-validation studies.

Reliability Analyses

For standard factor models, like the IUI-A single-factor model or the IUI-B six-factor model, the coefficient omega (ω) was used to assess the proportion of reliable variance in the set of observed variables that was accounted for by each latent variable in the model. For hierarchical models, in which each observed variable reflects both common and unique amounts of reliable variance, measurement equations were used to assess the relative contribution of each amount. The reliability coefficient omega was computed for the total score that, in second-order or bifactor models, reflects the proportion reliable variance that was accounted for by both the general and the group factors. The omega hierarchical coefficient (ω_h) was used to assess the proportion of variance accounted for by the general factor only in the total score. Where ω_h is appreciably different from ω , the reliable variance in the total score reflects the general factor as well as the group factors (Reise, 2012). The omega (ω) and omega scale (ω_s) coefficients can also be compared to assess the viability of subscale scores with group-factor items (Reise, 2012). Whereas ω reflects a mixture of general and unique variance for any specific subscale, ω_s is a measure of subscale reliability after the general factor variance has been partialled out. If ω_s is as large as ω , then the subscale score reflected mostly the group factor reliable variance. Most commonly ω_s tend to be smaller than ω as the common variance is greater.

Validity Analyses

Participant responses on the clinical scales for anxiety and depression were screened into positive and negative groups. The screening was based on internationally established cut-offs (i.e., BDI-II and STAI scores greater than 13 and 46, respectively). The delineation allowed for comparisons of IUI responses patterns for the IUI-A and IUI-B between groups, either using summated

ratings or bifactor model scores. Hierarchical factor scores were computed using the Anderson-Rubin method, which ensures the orthogonality of the estimated factors and produces scores that have a mean of 0, and a standard deviation of 1 (Distefano et al., 2009; Revelle, 2017).

RESULTS

CFA Results of IUI-A

We first examined the fit indices for factor models proposed elsewhere for the IUI-A (Gosselin et al., 2008). As detailed in **Table 1**, the single factor model was statistically significant and the fit indices were inconsistent with the recommended standards (Hu and Bentler, 1999); nevertheless, all items loaded onto the latent factor significantly and the composite reliability coefficient for the total score was high ($\omega = 0.92$). Similarly, the two-factor model with “intolerance of uncertainty and of uncertain situations” (i.e., items 1, 2, 3, 4, 5, 8, 9, 11, 15) and “intolerance of the unexpected and difficulty waiting in an uncertain situation” (i.e., items 6, 7, 10, 12, 13, 14) (Gosselin et al., 2008, p. 1434) were inconsistent with the recommended standards (**Table 1**); moreover, the two latent variables were too highly inter-correlated ($\phi = 0.98$) to support meaningful distinctions. The three-factor model with “intolerance of uncertainty and of uncertain situations” (i.e., items 1, 2, 3, 4, 5, 8, 9, 11, 15), “intolerance of the unexpected” (i.e., items 7, 14), and “difficulty waiting in an uncertain situation” (i.e., items 6, 10, 12, 13)” (Gosselin et al., 2008, p. 1434) were also inconsistent with the recommended standards (**Table 1**). The latent variables for the three-factor model were again highly inter-correlated (ϕ -s > 0.91), suggesting against that solution for the IUI-A.

For the standardized factor loadings for the single factor IUI-A model, ten items had coefficients greater than 0.60 and five items (i.e., 1, 2, 3, 6, 15) had relatively lower loadings (i.e., .50, .59, .59, .42, respectively). We tested a new model with a second latent variable using these five items. The results were statistically significant and significantly improved Model’s fit relative to the single factor model, $\Delta SB\chi^2 = 186.19$ ($df = 1$; $p < 0.001$) and the two-factor model proposed by Gosselin et al.

(2008, p. 1434), $\Delta SB\chi^2 = 63.83$, ($\Delta df = 0$); however, the fit indices still were inconsistent with the recommended standards (**Table 1**). The two latent variables were again very highly inter-correlated ($\phi = 0.83$).

Overall, the IUI-A results supported a unitary factor structure consistent with previous research, but also advised to optimize the scale. The inspection of the standardized factor loading matrix suggested that one might remove the five items with the lower commonality (i.e., $h^2 < 0.36$). The revised IUI-A single factor model after item removal was consistent with most of the recommended standards for all indices (**Table 1**). All items loaded significantly on the latent factor (all λ s > 0.60) and the reliability coefficient omega for the total score with ten items was about as large as that assessed in the previous analysis ($\omega = 0.91$). Accordingly, we used the ten-item IUI-A factor score in subsequent validity analyses ($M = 27.78$; $SD = 9.38$, in the present study).

CFA Results of IUI-B

The IUI-B was designed to have a six-factor structure reflecting clinical manifestations of core IU beliefs, like ‘avoidance’ (i.e., Items 1, 8, 12, 22, 26), ‘doubt’ (i.e., Items 2, 7, 13, 21, 30), ‘overestimation’ (i.e., Items 3, 14, 19, 23, 29), ‘worry’ (i.e., Items 6, 15, 17, 20, 28), ‘control’ (i.e., Items 4, 10, 18, 24, 27), and ‘reassurance’ (i.e., Items 5, 9, 11, 16, 25). Accordingly, we started by testing that six-factor model with correlated latent variables. The resulting fit indices were consistent with the recommended standards for all indices (**Table 1**), and CAIC was -376.05 ; however, an alternative three-factor model has been proposed (Carleton et al., 2010), with the original ‘control’ and ‘overestimation’ factors plus a ‘manifestations of anxious thought’ broad factor subsuming ten items selected from the original doubt, reassurance, and worry factors (i.e., Items 2, 5, 6, 7, 9, 11, 13, 17, 21, and 30). The three-factor model also produced fit indices consistent with the recommended standards for all indices (**Table 1**), but the CAIC was -1684.97 . Since smaller CAIC values indicate better fit, the three-factor model based on lesser items was preferred.

The six-factor model inter-factor correlations were high with ϕ -s ranging from 0.68 to 0.83, except for ‘doubt’ with ‘control’ factors ($\phi = 0.52$). The IUI-B was designed as a

TABLE 1 | Fit indices for the confirmatory factor analytic models of IUI-A and IUI-B.

Model	$SB\chi^2$ (df)	NNFI	CFI	SRMR	RMSEA (95% CI)
IUI-A, Single Factor	947.36** (92)	0.874	0.890	0.067	0.094 (0.089–0.100)
IUI-A, Two Factor	922.28** (89)	0.873	0.893	0.063	0.095 (0.089–0.100)
IUI-A, Three Factor	865.28** (87)	0.879	0.900	0.062	0.093 (0.087–0.098)
IUI-A, Two Factor, 5-item Latent Variable	858.45** (89)	0.883	0.901	0.068	0.091 (0.085–0.096)
IUI-A, Single Factor, 10-items	338.14** (35)	0.932	0.947	0.064	0.091 (0.082–0.100)
IUI-B, Six Factor	1405.24** (390)	0.932	0.939	0.043	0.051 (0.043–0.053)
IUI-B, Three Factor	948.18** (167)	0.921	0.931	0.050	0.068 (0.063–0.072)
IUI-B, Second Order Model	1520.03** (399)	0.927	0.933	0.069	0.053 (0.050–0.055)
IUI-B, Bifactor Model	1331.22** (375)	0.934	0.943	0.041	0.050 (0.047–0.053)

NNFI, robust version of non-normed fit index; CFI, robust version of comparative fit index; RMSEA, robust version of root mean square error of approximation; SRMR, standardized root mean square residual; IUI-A, intolerance of uncertainty index part A; IUI-B, intolerance of uncertainty index part B; ** $p < 0.001$.

TABLE 2 | Standardized factor loadings for the bifactor confirmatory factor analysis model of IUI-B.

Item	F1 Avoidance	F2 Doubting	F3 Overestimation	F4 Worrying	F5 Control	F6 Reassurance	F7 General
IUI-B 12	0.50	–	–	–	–	–	0.60
IUI-B 22	0.37	–	–	–	–	–	0.57
IUI-B 26	0.34	–	–	–	–	–	0.49
IUI-B 1	0.32	–	–	–	–	–	0.46
IUI-B 8	0.31	–	–	–	–	–	0.63
IUI-B 21	–	0.47	–	–	–	–	0.66
IUI-B 7	–	0.39	–	–	–	–	0.68
IUI-B 13	–	0.29	–	–	–	–	0.63
IUI-B 2	–	0.27	–	–	–	–	0.64
IUI-B 30	–	0.21	–	–	–	–	0.71
IUI-B 29	–	–	0.44	–	–	–	0.76
IUI-B 14	–	–	0.41	–	–	–	0.71
IUI-B 19	–	–	0.41	–	–	–	0.77
IUI-B 3	–	–	0.34	–	–	–	0.71
IUI-B 23	–	–	0.46	–	–	–	0.75
IUI-B 28	–	–	–	0.45	–	–	0.65
IUI-B 6	–	–	–	0.39	–	–	0.69
IUI-B 17	–	–	–	0.23	–	–	0.78
IUI-B 15	–	–	–	0.21	–	–	0.76
IUI-B 20	–	–	–	0.16	–	–	0.79
IUI-B 27	–	–	–	–	0.60	–	0.58
IUI-B 10	–	–	–	–	0.58	–	0.58
IUI-B 18	–	–	–	–	0.56	–	0.65
IUI-B 4	–	–	–	–	0.46	–	0.43
IUI-B 24	–	–	–	–	0.43	–	0.48
IUI-B 9	–	–	–	–	–	0.65	0.56
IUI-B 5	–	–	–	–	–	0.56	0.48
IUI-B 11	–	–	–	–	–	0.40	0.55
IUI-B 25	–	–	–	–	–	0.20	0.66
IUI-B 16	–	–	–	–	–	0.19	0.65

multidimensional clinical tool and the current results support notions that IU clinical manifestations represent lower order facets of a multifaceted hierarchical model. We tested a second-order factor model in which a General IUI-B factor was posited to affect the various clinical manifestations or consequences of IU through the six first-order factors. The second-order factor model produced fit indices consistent with the recommended standards for all indices (**Table 1**); however, the model fitted significantly worse than the six-factor model, $\Delta SB\chi^2 = 120.56$ ($df = 9$; $p < 0.001$). A less constrained bifactor model, in which a common IUI-B factor was posited to affect the clinical manifestations or consequences of IU directly and independently from the six group factors, produced fit indices consistent with the recommended standards for all indices (**Table 1**). This model fitted significantly better than the six-factor model, $\Delta SB\chi^2 = 75.38$ ($df = 15$; $p < 0.001$), and appeared to be the most accurate IUI-B factorial structure representation for the current data.

The IUI-B can be scored by deriving a single total score for the general factor or six subscale scores for each of the group factors. Based on standardized factor loadings (**Table 2**),

the reliability analyses described by Reise (2012) for bifactor model scores were used to assess the viability of total and sub-scale scores for IUI-B. First, we assessed the proportion of reliable variance in the total score accounted for by the general factor ($\omega_h = 0.70$) and compared that to total proportion of reliable variance ($\omega = 0.96$). The general factor accounted for about 70% of the total score reliable variance, whereas the total score reliability was lower for the portion of reliable variance accounted for by group factors (i.e., ~26%). We then compared the standard omega assessed for each group-factor items (ω) and the omega scale hierarchical (ω_s) to assess the unique information conveyed by the IUI-B subscales. The ω_s provided a measure of reliability after partialing out the general factor variance was from the sub-scale scores. The standard omega coefficients for the six subscales were all fairly high for five-item scales (i.e., $\omega = 0.83$ for avoidance, doubt, overestimation, worry, control; $\omega = 0.82$ for reassurance). In contrast, ω_s coefficients fell – often substantially – for worry ($\omega_s = 0.11$), doubt ($\omega_s = 0.16$), overestimation ($\omega_s = 0.20$), avoidance ($\omega_s = 0.25$), reassurance ($\omega_s = 0.27$), and control ($\omega_s = 0.40$). Overall, the IUI-B subscale scores reliably measured common variance in IU, but also maintained some

specific amount of information relative to using the IUI-B total score, particularly for avoidance, reassurance, and control.

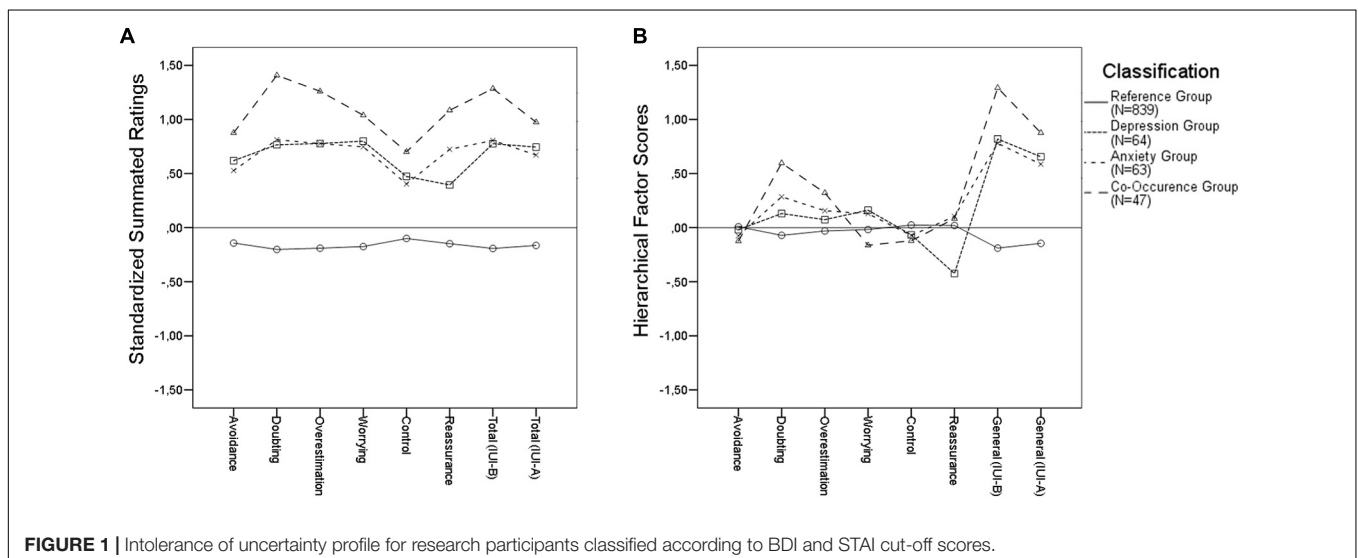
Comparison of Summated Ratings and Hierarchical Factor Scores

Using the established cut-offs for anxiety and depression scales, we identified $N = 112$ (10.7%) and $N = 114$ (10.9%) participants screening positive for chronic anxiety (STAI A-trait > 46) and moderate depression (BDI-II > 20), respectively. The STAI A-trait and BDI-II classifications were positively correlated (Spearman's $Rho = 0.36$; $p < 0.001$). Accordingly, we reclassified the research participants into three clinical proxy groups: 63 cases (6.1% of the sample) scoring above the cut-off for chronic anxiety on the STAI A-trait, only; 66 cases (6.3%) scoring above the cut-off for moderate depression on the BDI-II, only; 48 cases (4.6%) scoring above the cut-off on both the STAI A-trait and the BDI-II. A reference group of 868 cases (83%) participants who scored below the clinical cut-off for both anxiety and depression, and were also identified for comparisons in data analyses. For simplicity, hereafter we refer to these groups as “anxiety,” “depression,” “co-occurrence,” and “reference” group, respectively. This classification was used as a between-subjects factor in two multivariate analyses of variance under the hypotheses that greater IU is associated with greater co-occurrence of depressive and anxious symptoms, and that depression and/or anxiety is associated only with specific clinical manifestations of IU. The first analysis compared the groups on the IUI summated ratings (Figure 1A). The second analysis compared the groups on the IUI-A standard factor score and on the IUI-B hierarchical factor scores estimated from the best fitting CFA models for each part of the inventory (Figure 1B). Divergent results between the analysis might reveal the extent to which group differences could be biased, and potentially misleading, when summated ratings are used to make inferences at the facet level for multifaceted hierarchical constructs.

The IUI-A and the IUI-B total scores were highly correlated both using summated ratings and factor scores (r -s = 0.76 and 0.74, respectively). Using summated ratings, the IUI-B total score were highly correlated with IUI-B subscale scores (r -s range 0.75–0.86); the coefficients were somewhat lower for IUI-A with IUI-B subscale scores (r -s range 0.75–0.86). Using factor scores, the IUI-B general factor was uncorrelated with IUI-B factor scores; specifically, the coefficients were significant only for the IUI general factor with worry ($r = 0.20$) and need for control ($r = 0.07$). This correlation analysis indicated that, using hierarchical factor scores, respondents can have IUI-B scores that parse specific and common sources of variance in ratings.

The analysis of summated ratings (Figure 1A) indicated a significant multivariate effect of the classification variable (Roy's root = 0.298; $F = 42.82$; df -s = 8,1006; $p < 0.001$; $\eta_p^2 = 0.23$). Follow-up contrast analyses indicated that, when combined, the three clinical proxy groups were significantly higher than the reference group on all IUI-B subscales, as well as on the IUI-B and the IUI-A total scores ($\Psi_{\text{avoidance}} = 2.45$; $\Psi_{\text{doubting}} = 3.57$; $\Psi_{\text{overestimation}} = 3.39$; $\Psi_{\text{worrying}} = 3.11$; $\Psi_{\text{control}} = 1.86$; $\Psi_{\text{reassurance}} = 2.64$; $\Psi_{\text{IUI-B total}} = 3.47$; $\Psi_{\text{IUI-A total}} = 2.89$; all p -s < 0.001). The anxiety and depression groups were significantly lower than the co-occurrence group on some IUI-B subscales, and on both the IUI-B and IUI-A total score ($\Psi_{\text{doubting}} = -1.26$; $\Psi_{\text{overestimation}} = -0.96$; $\Psi_{\text{reassurance}} = -1.06$; $\Psi_{\text{IUI-B total}} = -0.99$; all p -s < 0.001); however, the anxiety and depression groups were not significantly different on any of the summated ratings scores.

The analysis of hierarchical scores (Figure 1B) also indicated a significant multivariate effect of the classification variable (Roy's root = 0.293; $F = 36.43$; df s = 8,1006; $p < 0.001$; $\eta_p^2 = 0.22$). As in the analysis of summated ratings, the follow-up contrasts revealed that the three clinical proxy groups combined were significantly higher than the reference group on both the IUI-B and IUI-A general factor scores ($\Psi_{\text{IUI-B general}} = 3.45$; $\Psi_{\text{IUI-A general}} = 2.55$; both p -s < 0.001); however, only some of the IUI-B factor scores yielded significant differences between



clinical proxy groups combined and the reference group ($\Psi_{\text{doubting}} = 1.22, p < 0.001$; $\Psi_{\text{overestimation}} = 0.64, p < 0.05$). Regarding doubting scores, a follow up analysis indicated that the co-occurrence group was significantly higher than the depression and anxiety groups combined ($\Psi_{\text{doubting}} = 0.78, p < 0.05$), while these latter groups did not differ significantly. Instead, no combination of clinical proxy groups yielded statistically significant comparisons on overestimation of threat factor scores.

The anxiety and depression groups were also significantly lower than the co-occurrence group on the IUI-B general factor ($\Psi_{\text{IUI-B general}} = -0.99; p < 0.01$), and marginally on the IUI-A general factor ($\Psi_{\text{IUI-A general}} = 0.51; p = 0.07$). The two clinical proxy groups of participants scoring above the cut-off either on anxiety or depression were statistically different on the reassurance group factor ($\Psi_{\text{reassurance}} = 0.54; p < 0.01$), but not on the doubting and overestimation scores. In particular, as detailed in **Figure 1B**, participants in the depression group were less apt than other clinical proxy groups to seek reassurance from other people or presumed authoritative sources in order to cope with feared unknowns, a result that would be overlooked using summated ratings instead of hierarchical scores.

DISCUSSION

The current study evaluated the validity of the IUI, a two-part scale separately assessing IU core beliefs (IUI-A) and the clinical consequences of these beliefs in diverse clinical disorders (IUI-B). The IUI-A was best explained by a unidimensional structure. Alternative multiple factor models proposed in the extant literature for French and English versions of the scale were not supported. Indeed, the hypothesis that items poorly loading on the single latent variable could give rise to a theoretically meaningful second latent factor was rejected due to the large empirical overlapping of the two latent variables in the models tested. Despite a unidimensional structure, however, the IUI-A produced the most robust fit indices for a single factor model in which the latent variable used only a 10-item subset of the original 15 items. Previous research also showed that an atheoretical removal of three items from the English language version improved the fit of a unitary solution for the IUI-A (Carleton et al., 2010). Nevertheless, the subset of items used in the present study was different from that used in previous studies. Previous research with Italian and English speaking participants has pointed out some caveats related to the use of IU scales across countries (Bottesi et al., 2015, 2016). Because the IUI-A factor structure was problematic in two different languages (i.e., French and English), as well as in the present study, while the IUI-B seemed more robust to cultural and translational issues, we believe that translation bias was not a significant problem in this study. We speculate that people with different cultural background may differ in how the cultures engage with uncertainty at the level of IU core beliefs (e.g., appraisal and acceptance of uncertain situations, discomfort with the unexpected, or difficulty waiting in an uncertain situation). By contrast, the structure of the

clinical consequences of IU was approximately the same in French, English, and Italian studies, showing that reactions to uncertainty were comparable across cultures. The present findings add to the extant literature (Gosselin et al., 2008; Carleton et al., 2010) in that they reinforce the need for refining the assessment of IU core beliefs for the use of the IUI-A in cross-cultural research. The unitary factor structure of the IUI-A was supported overall, but the impact of removing items from the original set remains to be reassessed. In the present study, we proposed a 10-item version that calls for a cross-validation across languages (e.g., French and English) and cultures (e.g., North American and European countries). It is noteworthy, however, that the IUI-A total score with ten items was highly reliable and had fair criterion validity with BDI-II and STAI classifications as well as high convergent validity with the IUI-B.

Regarding the IUI-B, the intended six-factor structure produced robust fit indices in all countries and languages with avoidance, doubting, overestimation, worrying, seeking reassurance, and need for control factors (Gosselin et al., 2008; Carleton et al., 2010). Moreover, the factor analytic results supported the view that IU was best modeled as hierarchical multifaceted construct (Hale et al., 2016; Lauriola et al., 2016). A bifactor model with one general factor common to all the items, as well as the six factors common to specific groups of items, was evidenced as producing superior model fit indices relative to the standard six-factor model. In other words, the general factor captured the variance common to all items describing diverse clinical manifestations of IU in GAD, Depression and OCD patients, but each specific manifestation was also affected by a unique source of variance associated with specific groups of items. This result implies that the general IU factor may be contributing to a transdiagnostic range of disorders whereas the group factors may be contributing to specific disorders, or patients (Carleton et al., 2012; Mahoney and McEvoy, 2012; Einstein, 2014; Carleton, 2016b).

According to the view that IU is higher in clinical groups than in control groups across several disorders (Holaway et al., 2006; Gentes and Ruscio, 2011; Sternheim et al., 2011; Carleton et al., 2012), our study showed that both the IUI-A and IUI-B summated rating scores discriminated between clinical proxy groups and a reference group. The IUI-A and IUI-B summated rating scores were both higher among participants scoring above the cut-off on the two proxy measures of anxiety and depression, relative to those scoring above the cut-off on only one of the proxy measures; as such, the results were consistent with the view that IU proportionally increases with co-occurrence of depressive and anxious symptoms (Holaway et al., 2006; Yook et al., 2010; McEvoy and Mahoney, 2011). The overall results were confirmed with a parallel analysis using hierarchical factor scores, in which the IUI-A and the IUI-B general factors reproduced quite well the expected divergent response patterns between clinical proxy groups and the reference group (Holaway et al., 2006; Gentes and Ruscio, 2011; Sternheim et al., 2011).

The current results support important avenues for future research regarding the interrelationships between IU, anxiety,

depression, and comorbidity. The pattern suggests that targeting IU as a general risk factor may be beneficial at a global level, but when engaging treatment for a specific disorder (e.g., GAD, OCD) there may be benefits from targeting specific manifestations of IU. The contemporary transdiagnostic treatment models (e.g., the Unified Protocol; Ellard et al., 2010) may therefore be particularly well-suited as initial interventions, followed thereafter as necessary by disorder-specific modules (Grayson, 2010). The reverse order for treatment, starting with disorder-specific modules and then engaging transdiagnostic modules, may also be appropriate. In either case, the areas warrant additional research.

The current results also offer preliminary proof-of-concept evidence that using hierarchical factor scores to disentangle general and unique variance components could be useful to highlight common and specific characteristics of clinical-proxy samples (Reise, 2012). Nevertheless, the presence of a general IU factor represents a challenge for future research. On the one hand, the general factor might genuinely reflect IU-specific mechanisms that might account for diverse clinical manifestations of IU in a transdiagnostic framework. On the other hand, the general factor could merely represent a common method factor or some response set biases. Whatever the source of the common variance, group differences in the clinical manifestations of IU were overestimated when using summated ratings to assess non-clinical participants.

Different groups screening positive for anxiety and/or depression were not actually statistically different in some of the factor scores, as it was observed for “avoidance” and “need for control,” after controlling for the effect of the general IU variance. The result suggests that experiential avoidance and attempts to control uncertainty in anxious and depressed patients might be due to generalized IU core beliefs. If confirmed with clinical patients, these findings might suggest that IU core beliefs should be prioritized when treating patients reporting these specific clinical consequences of IU. By contrast, factor scores like “doubting” and “overestimation of threat” were still significant after controlling for the effect of the general IU variance. Not only the three clinical proxy groups were significantly higher than the reference group on “doubting” scores, but a follow up analysis revealed that the co-occurrence group was significantly higher than the depression and anxiety groups combined. The current results, if confirmed with clinical patients, might suggest that both IU core beliefs and doubting should be prioritized when treating patients reporting this specific clinical consequence of IU.

The current study has limitations that also provide important directions for future research. First, the current study used established clinical tools and applied internationally valid cut-offs to identify participants reporting clinically significant symptoms. Nevertheless, a major constraint is the lack of clinical interviews, which would have provided more accurate information concerning the clinical status of the research participants. Therefore, it is not warranted that the findings of the study could be generalized to clinical patients. Indeed, future investigations should attempt to replicate the current results with data gathered from formally diagnosed participants, or

adding clinical interviews to the research design. If replicated, the results would support more nuanced clinical utility for total, subscale, and factor scores. Second, despite robust psychometrics, the application of IUI-B subscale scores was undermined by the relatively low unique variance. The current results support deriving a total score through simple aggregation of items for each of the six subscales mostly to reflect general factor variance. Accordingly, use of the subscale scores as reliable indicators of specific constructs currently warrants caution (Chen et al., 2012). The factor scores from the bifactor model may be more reliable (Reise et al., 2010; Chen et al., 2012), but present challenges for practicality. Future researchers should consider developing applications to facilitate the practical utility of clinical factor scores for identifying general and specific (i.e., IUI-B) sources of variance in IU. Third, the incremental value of the IUI-B hierarchical factor scores over standard assessments of needs to be addressed in rigorous empirical investigations before the clinical implementation of this scoring method. Future researchers should consider developing larger and more diverse assessments of general and specific clinical manifestations to strengthen the incremental utility of specific IU sources (e.g., Thibodeau et al., 2015).

Notwithstanding the limitations, the current study contributes to cross-validation of the IUI beyond use with French and English Canadian samples and beyond North America. Our study provided psychometric support for the Italian version of the IUI scales and preliminary normative data for international clinical research on IU. Previous cross-validation efforts worldwide supported contemporary refinements for defining (see Carleton, 2016b) and assessing IU (e.g., use of the IUS-12; Helsen et al., 2013). Similarly, the current results suggest (1) an abridged ten-item version of the IUI-A as a promising candidate for transdiagnostic measurement of IU core beliefs in large assessment batteries; and (2) using factor scores may be appreciably more defensible than simple aggregates for measuring general and specific IU for clinical and experimental methods.

ETHICS STATEMENT

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study

AUTHOR CONTRIBUTIONS

The authors discussed the contents of this article together. ML, OM, RF, CT and RT conceived the study. ML, OM and RNC elaborated on the theoretical framework and the research hypotheses. CT and RF collected the data. ML and RNC analyzed the data. The final version of the manuscript was written by ML, OM, CT and RNC.

ACKNOWLEDGMENTS

The authors wish to thank Prof. William Revelle for helpful advice concerning the computation of hierarchical factor scores.

REFERENCES

- Andridge, R. R., and Little, R. J. A. (2010). A review of hot deck imputation for survey non-response. *Int. Stat. Rev.* 78, 40–64. doi: 10.1111/j.1751-5823.2010.00103.x
- Beck, A. T., Steer, R. A., and Brown, G. K. (1996). *Beck Depression Inventory-ii (bdi-ii)*. San Antonio, TX: Psychological Corporation.
- Bentler, P. M. (2004). *EQS 6: Structural Equation Program Manual*. Encino, CA: Multivariate Software, Inc.
- Berenbaum, H., Bredemeier, K., and Thompson, R. J. (2008). Intolerance of uncertainty: Exploring its dimensionality and associations with need for cognitive closure, psychopathology, and personality. *J. Anxiety Disord.* 22, 117–125. doi: 10.1016/j.janxdis.2007.01.004
- Bottesi, G., Ghisi, M., Carraro, E., Barclay, N., Payne, R., and Freeston, M. H. (2016). Revising the intolerance of uncertainty model of generalized anxiety disorder: evidence from UK and Italian undergraduate samples. *Front. Psychol.* 7:1723. doi: 10.3389/fpsyg.2016.01723
- Bottesi, G., Ghisi, M., Novara, C., Bertocchi, J., Boido, M., De Dominicis, I., et al. (2015). Intolerance of uncertainty scale (IUS-27 e IUS-12): Due studi preliminari. *Psicoter. Cogn. Compont.* 21, 345–365.
- Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions. *Psychometrika* 52, 345–370. doi: 10.1007/BF02294361
- Carleton, R. N. (2016a). Fear of the unknown: One fear to rule them all? *J. Anxiety Disord.* 41, 5–21. doi: 10.1016/j.janxdis.2016.03.011
- Carleton, R. N. (2016b). Into the unknown: a review and synthesis of contemporary models involving uncertainty. *J. Anxiety Disord.* 39, 30–43. doi: 10.1016/j.janxdis.2016.02.007
- Carleton, R. N., Gosselin, P., and Asmundson, G. J. G. (2010). The intolerance of uncertainty index: replication and extension with an English sample. *Psychol. Assess.* 22, 396–406. doi: 10.1037/a0019230
- Carleton, R. N., Mulvogue, M. K., Thibodeau, M. A., McCabe, R. E., Antony, M. M., and Asmundson, G. J. G. (2012). Increasingly certain about uncertainty: intolerance of uncertainty across anxiety and depression. *J. Anxiety Disord.* 26, 468–479. doi: 10.1016/j.janxdis.2012.01.011
- Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., and Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: a comparison of the bifactor model to other approaches. *J. Pers.* 80, 219–251. doi: 10.1111/j.1467-6494.2011.00739.x
- Distefano, C., Zhu, M., and Mindrila, D. (2009). Understanding and using factor scores: considerations for the applied researcher. *Pract. Assess. Res. Eval.* 14, 1–11.
- Dugas, M. J., Gagnon, F., Ladouceur, R., and Freeston, M. H. (1998). Generalized anxiety disorder: a preliminary test of a conceptual model. *Behav. Res. Ther.* 36, 215–226. doi: 10.1016/S0005-7967(97)00070-3
- Einstein, D. A. (2014). Extension of the transdiagnostic model to focus on intolerance of uncertainty: a review of the literature and implications for treatment. *Clin. Psychol. Sci. Pract.* 21, 280–300. doi: 10.1111/cpsp.12077
- Ellard, K. K., Fairholme, C. P., Boisseau, C. L., Farchione, T. J., and Barlow, D. H. (2010). Unified protocol for the transdiagnostic treatment of emotional disorders: protocol development and initial outcome data. *Cogn. Behav. Pract.* 17, 88–101. doi: 10.1016/j.cbpra.2009.06.002
- Fisher, P. L., and Durham, R. C. (1999). Recovery rates in generalized anxiety disorder following psychological therapy: an analysis of clinically significant change in the {STAI}-{T} across outcome studies since 1990. *Psychol. Med.* 29, 1425–1434. doi: 10.1017/S0033291799001336
- Gentes, E. L., and Ruscio, A. M. (2011). A meta-analysis of the relation of intolerance of uncertainty to symptoms of generalized anxiety disorder, major depressive disorder, and obsessive-compulsive disorder. *Clin. Psychol. Rev.* 31, 923–933. doi: 10.1016/j.cpr.2011.05.001

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00388/full#supplementary-material>

- Ghisi, M., Flebus, G. B., Montano, A., Sanavio, E., and Sica, C. (2006). *Beck Depression Inventory-II. Adattamento italiano: Manuale*. Florence: Giunti Organizzazioni Speciali.
- Gosselin, P., Ladouceur, R., Evers, A., Laverdière, A., Routhier, S., and Tremblay-Picard, M. (2008). Evaluation of intolerance of uncertainty: development and validation of a new self-report measure. *J. Anxiety Disord.* 22, 1427–1439. doi: 10.1016/j.janxdis.2008.02.005
- Grayson, J. B. (2010). OCD and intolerance of uncertainty: treatment issues. *J. Cogn. Psychother.* 24, 3–15. doi: 10.1891/0889-8391.24.1.3
- Hale, W., Richmond, M., Bennett, J., Berzins, T., Fields, A., Weber, D., et al. (2016). Resolving uncertainty about the intolerance of uncertainty scale-12: application of modern psychometric strategies. *J. Pers. Assess.* 98, 200–208. doi: 10.1080/00223891.2015.1070355
- Helsen, K., Van den Bussche, E., Vlaeyen, J. W., and Goubert, L. (2013). Confirmatory factor analysis of the Dutch intolerance of uncertainty scale: comparison of the full and short version. *J. Behav. Ther. Exp. Psychiatry* 44, 21–29. doi: 10.1016/j.jbtep.2012.07.004
- Holaway, R. M., Heimberg, R. G., and Coles, M. E. (2006). A comparison of intolerance of uncertainty in analogue obsessive-compulsive disorder and generalized anxiety disorder. *J. Anxiety Disord.* 20, 158–174. doi: 10.1016/j.janxdis.2005.01.002
- Hong, R. Y., and Cheung, M. W.-L. (2015). The structure of cognitive vulnerabilities to depression and anxiety. *Clin. Psychol. Sci.* 3, 892–912. doi: 10.1177/2167702614553789
- 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
- Kobori, O., and Salkovskis, P. M. (2013). Patterns of reassurance seeking and reassurance-related behaviours in OCD and anxiety disorders. *Behav. Cogn. Psychother.* 41, 1–23. doi: 10.1017/S1352465812000665
- Lauriola, M., Foschi, R., Mosca, O., and Weller, J. (2015). Attitude toward ambiguity: empirically robust factors in self-report personality scales. *Assessment* 23, 353–373. doi: 10.1177/1073191115577188
- Lauriola, M., Mosca, O., and Carleton, R. N. (2016). Hierarchical factor structure of the intolerance of uncertainty scale short form (IUS-12) in the Italian version. *Test. Psychom. Methodol. Appl. Psychol.* 23, 377–394. doi: 10.4473/TPM23.3.8
- Mahoney, A. E. J., Hobbs, M. J., Newby, J. M., Williams, A. D., Sunderland, M., and Andrews, G. (2016). The worry behaviors inventory: assessing the behavioral avoidance associated with generalized anxiety disorder. *J. Affect. Disord.* 203, 256–264. doi: 10.1016/j.jad.2016.06.020
- Mahoney, A. E. J., and McEvoy, P. M. (2012). A transdiagnostic examination of intolerance of uncertainty across anxiety and depressive disorders. *Cogn. Behav. Ther.* 41, 212–222. doi: 10.1080/16506073.2011.622130
- McEvoy, P. M., and Erceg-Hurn, D. M. (2015). The search for universal transdiagnostic and trans-therapy change processes: evidence for intolerance of uncertainty. *J. Anxiety Disord.* 41, 96–107. doi: 10.1016/j.janxdis.2016.02.002
- McEvoy, P. M., and Mahoney, A. E. J. (2011). Achieving certainty about the structure of intolerance of uncertainty in a treatment-seeking sample with anxiety and depression. *J. Anxiety Disord.* 25, 112–122. doi: 10.1016/j.janxdis.2010.08.010
- Miranda, R., Fontes, M., and Marroquín, B. (2008). Cognitive content-specificity in future expectancies: role of hopelessness and intolerance of uncertainty in depression and GAD symptoms. *Behav. Res. Ther.* 46, 1151–1159. doi: 10.1016/j.brat.2008.05.009
- Newman, M. G., Llera, S. J., Erickson, T. M., Przeworski, A., and Castonguay, L. G. (2013). Worry and generalized anxiety disorder: a review and theoretical synthesis of evidence on nature, etiology, mechanisms, and treatment. *Annu. Rev. Clin. Psychol.* 9, 275–297. doi: 10.1146/annurev-clinpsy-050212-185544

- Nikodijevic, A., Moulding, R., Anglim, J., Aardema, F., and Nedeljkovic, M. (2015). Fear of self, doubt and obsessive compulsive symptoms. *J. Behav. Ther. Exp. Psychiatry* 49, 164–172. doi: 10.1016/j.jbtep.2015.02.005
- Pedrabissi, L., and Santinello, M. (1989). *Inventario per l'ansia di stato e di tratto (forma Y) [State-Trait Anxiety Inventory (Form Y)]*. Florence: Organizzazioni Speciali.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivar. Behav. Res.* 47, 667–696. doi: 10.1080/00273171.2012.715555
- Reise, S. P., Moore, T. M., and Haviland, M. G. (2010). Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *J. Pers. Assess.* 92, 544–559. doi: 10.1080/00223891.2010.496477
- Revelle, M. W. (2017). *psych: Procedures for Personality And Psychological Research (R package)*. Evanston, IL: Northwestern University, 1–250.
- Rosen, N. O., Ivanova, E., and Knäuper, B. (2014). Differentiating intolerance of uncertainty from three related but distinct constructs. *Anxiety Stress Coping* 27, 55–73. doi: 10.1080/10615806.2013.815743
- Samuels, J., Bienvenu, O. J., Krasnow, J., Wang, Y., Grados, M. A., Cullen, B., et al. (2017). An investigation of doubt in obsessive-compulsive disorder. *Compr. Psychiatry* 75, 117–124. doi: 10.1016/j.comppsy.2017.03.004
- 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
- Sexton, K. A., and Dugas, M. J. (2008). The cognitive avoidance questionnaire: validation of the English translation. *J. Anxiety Disord.* 22, 355–370. doi: 10.1016/j.janxdis.2007.04.005
- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory STAI (Form Y) ("Self-Evaluation Questionnaire")*. Palo Alto, CA: Consulting Psychology Press.
- Sternheim, L., Konstantellou, A., Startup, H., and Schmidt, U. (2011). What does uncertainty mean to women with anorexia nervosa? An interpretative phenomenological analysis. *Eur. Eat. Disord. Rev.* 19, 12–24. doi: 10.1002/erv.1029
- Thibodeau, M. A., Carleton, R. N., McEvoy, P. M., Zvolensky, M. J., Brandt, C. P., Boelen, P. A., et al. (2015). Developing scales measuring disorder-specific intolerance of uncertainty (DSIU): a new perspective on transdiagnostic. *J. Anxiety Disord.* 31, 49–57. doi: 10.1016/j.janxdis.2015.01.006
- Yook, K., Kim, K. H., Suh, S. Y., and Lee, K. S. (2010). Intolerance of uncertainty, worry, and rumination in major depressive disorder and generalized anxiety disorder. *J. Anxiety Disord.* 24, 623–628. doi: 10.1016/j.janxdis.2010.04.003

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

Copyright © 2018 Lauriola, Mosca, Trentini, Foschi, Tambelli and Carleton. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Selfie Expectancies Among Adolescents: Construction and Validation of an Instrument to Assess Expectancies Toward Selfies Among Boys and Girls

Valentina Boursier^{1*} and Valentina Manna²

¹ Department of Humanities, University of Naples Federico II, Naples, Italy, ² Association for Social Promotion Roots in Action, Naples, Italy

OPEN ACCESS

Edited by:

Michela Balsamo,
Università degli Studi "G. d'Annunzio"
Chieti – Pescara, Italy

Reviewed by:

Silvia Casale,
Università degli Studi di Firenze, Italy
Adriano Schimmenti,
Kore University of Enna, Italy

*Correspondence:

Valentina Boursier
valentina.boursier@unina.it

Specialty section:

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

Received: 07 March 2018

Accepted: 09 May 2018

Published: 29 May 2018

Citation:

Boursier V and Manna V (2018)
Selfie Expectancies Among
Adolescents: Construction
and Validation of an Instrument
to Assess Expectancies Toward
Selfies Among Boys and Girls.
Front. Psychol. 9:839.
doi: 10.3389/fpsyg.2018.00839

Selfie-taking and posting is one of the most popular activities among teenagers, an important part of online self-presentation that is related to identity issues and peer relations. The scholarly literature emphasizes different yet conflicting motivations for selfie-behavior, stressing deeper analysis of psychological factors and the influence of gender and age. Expectancies are “explanatory device[s]” that can help us study adolescent behavior. However, no instruments have been devised that specifically explore the expectations teenagers have about selfies and their influence on selfie-frequency. The current study proposes a short and reliable instrument to identify teen expectancies about selfie-behavior. This instrument was validated using a sample of 646 Italian adolescents (14 to 19 years old) by means of Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). We also explore the relationship between selfie expectancies and selfie-frequency, as well as the role of gender in shaping selfies. Our results point toward a 7-factor model that characterizes expectations toward selfies as a multi-dimensional construct linked to both positive and negative perceptions of the nature and consequences of selfies. The overall model fitted the data sufficiently ($\chi^2 = 5067.051$, $p = 0.0000$; CFI = 0.962; TLI = 0.954; RMSEA \leq 0.05; 0.035; SRMR = 0.046), showing an adequate reliability of the scale ($\alpha = 0.830$). Bivariate correlations between selfie expectancies and selfie-frequency ($r = 0.338$, $p < 0.001$) confirmed the convergent validity of the tool. Selfie-sharing is a common practice that is widespread among the participants in this study. Self-promotion represents a positive function of selfies. Selfies promote self-presentation and self-confidence, both in boys and girls. Moreover, selfie expectancies address sexual self-attractiveness, especially among boys. Despite the positive aspects of selfies, our results stress adolescent awareness of the negative consequences of this type of web-exposure. This is especially true among girls, whose selfie-behavior is, paradoxically, more frequent than boys. Self-management through selfie-posting is a positive outcome of selfie-behavior that plays a

key role among adolescents, even though the dangers of manipulating selfies in order to garner approval from one's peers need to be considered. The positive psychometric properties of the measure point toward the need for further research on both generalized and specific selfie-behaviors.

Keywords: selfie, expectancies, adolescents, gender, assessment, measure, validation

INTRODUCTION

The neologism “selfie” was the Oxford Dictionary’s Word of the Year in 2013 (Oxford Dictionaries, 2013). It commonly refers to a photograph of oneself (alone or with other people) that is taken with a camera, camera phone, or some other hand-held device. Even though the selfie concept addresses several self-portrayal issues (Kiprin, 2013), selfies are typically shared through social media (Sorokowski et al., 2015). Indeed, self-portrayal is one of the most widespread online activities, particularly among adolescents (Lenhart et al., 2010; Kiprin, 2013; Senft and Baym, 2015) and college-age young adults (Katz and Crocker, 2015). According to Lee and Sung (2016), smartphone users take approximately 93 million selfies each day, and approximately 880 billion online photos were shared in 2014. Moreover, 30% of the total photos shared on social networking sites (SNS) in 2014 were selfies posted by adolescents (Locateadoc.com, 2014). It has been estimated that Instagram users alone have shared 238 million photos with the hashtag #selfie, and 128 million photos with the hashtag #me (Weiser, 2015). A recent study in the United States showed that 98% of participants (aged 18 to 24) took selfies, and 69% tended to share selfies 3 to 20 times daily (Katz and Crocker, 2015).

Selfie-taking/sharing certainly represents “one of the dominant forms of content shared in the computer-mediated communication platforms” (Dhir et al., 2016; *p* 0.549). The selfie craze has encouraged greater interest in examining the psychological and psychosocial aspects of this phenomenon, thus feeding the significant debate on both the psychopathological facets of this type of behavior and the growing risks of hyper-pathological conceptualization of common media use (Billieux et al., 2015; Kardefelt-Winther et al., 2017).

According to Nadkarni and Hofmann (2012), social media use fulfills two social needs: self-presentation and the need to belong. Selfie-sharing on SNS improves one’s self-esteem/mood through “likes” (Reich et al., 2018), and seems to be especially related to self-presentation behaviors and relationship construction (Sorokowska et al., 2016; Taylor et al., 2017).

Even though posting selfies allows people to express their own identity and social relationships, other psychological factors might produce different types of selfie behaviors (Albury, 2015). Attitudes toward selfie-taking have been analyzed in three countries by Katz and Crocker (2015). Their study demonstrated the importance of self-presentation and identification in selfie production, as well as the need to receive feedback from one’s peers. Moreover, taking selfies helps people experiment with their appearance, their accessories, and their environment

(Kiprin, 2013). Young women declare that selfie-taking helps them to feel authentic (Warfield, 2014). Nguyen and Barbour (2017) recently found that young women consider selfies to be authentic expressions of identity. By contrast, Christoforakos and Diefenbach’s (2016) study found that selfies are associated with a lack of authenticity. They also concluded that young men and women identified both positive aspects (e.g. independence, memory/documentation, relatedness, and control/self-staging) and negative aspects (e.g., illusion/fake, threat to self-esteem, and negative impression on others, and bad picture quality) of selfies.

Recent studies point toward different/conflicting motivations for selfie-taking. For instance, Sung et al. (2016) have shown that attention seeking, archiving, communication, and entertainment motivates selfie-posting on SNS, while also arguing that narcissism considerably predicts selfie-posting frequency. An Italian study, moreover, suggests that various personality traits can predict dissimilar selfie posting behaviors in adolescents and young adults (Baiocco et al., 2016). Other scholars, however, have suggested that narcissism significantly predicts selfie-posting frequency, especially among women (Fox and Rooney, 2015; Sorokowski et al., 2015; Weiser, 2015, 2018; Lee and Sung, 2016; McCain et al., 2016; Barry et al., 2017). Halpern et al. (2016) similarly suggest that selfies have a self-reinforcement effect - that narcissists frequently take selfies in order to maintain positive views of themselves, which in turn increases their narcissism levels.

Etgar and Amichai-Hamburger (2017) have identified three principal motivations behind selfie-taking: selfie-approval, belonging, and documentation. They also suggest that each motivation can be connected to various personality traits. However, unlike previous studies in this area, they did not find a connection between these motivations and narcissism. This somewhat contradictory finding demonstrates that selfies are a multidimensional phenomenon that requires further research. Some research has emphasized the analysis of psychopathological (obsessive) traits among selfie-taking adolescents, oftentimes treating it as a potentially addictive behavior (Balakrishnan and Griffiths, 2017; Griffiths and Balakrishnan, 2018). However, a recent study on the positive psychological effects selfies have on self-presentation strategies has been conducted on young European men and women (Diefenbach and Christoforakos, 2017). The authors’ findings showed that who’s more engaged in selfie-taking considers selfies a good possibility for a selective self-presentation. Strategies associated with self-promotion and/or self-disclosure play an especially important role in supporting various selfie behaviors.

Age and Gender Differences in Selfie Behavior

Both age and gender influences SNS use, as well as the user's attitudes and perceptions of Internet-based activities (Dhir et al., 2016). Posting selfies is typically assumed to be a gendered process (Albury, 2015), one that varies according to the type of selfie, selfie frequency, selfie attitudes, and motivations. Males and females tend to use selfies for self-presentation (Katz and Crocker, 2015), however, it has been observed that males and females tend to post different selfies (Sorokowski et al., 2015; Dhir, 2016) and that women are more inclined to post selfies than men (Qiu et al., 2015; Sorokowska et al., 2015, 2016).

Nguyen (2014) has observed that young women (18 to 29 years old) share selfies on Instagram in order to accumulate "likes," and that the quality of a selfie depends on lighting, scenography, and posture. Nguyen (2014) also found that selfies allow young women to experiment with new and different looks. Recently, Chae (2017) concluded that selfie-editing on social media is related to the average young woman's attempts to cultivate an ideal form of online self-presentation. Similarly, Nelson (2013) argues that young women share selfies in order to receive positive feedback. For this reason, a selfie code of conduct seems to be especially popular among young women (Warfield, 2014).

Adolescents suggest that selfie-posting could have a negative impact on their self-presentation and social capital (Gibbs et al., 2014). Indeed, they are more likely than adults to engage in a "selfie policy" that emphasizes selecting the ideal photo (Senft and Baym, 2015).

Among young women, selfie posts seem to produce higher self-esteem (Poe, 2015). However, Sorokowska et al. (2016) found that there is no firm relationship between self-esteem and selfie-posting behavior, even though social exhibitionism and extraversion can predict the frequency of selfie-posting among both men and women.

Kim and Chock's (2015) study states that gender isn't a significant predictor of selfie behaviors, but it does moderate the relationship between the need for popularity and posting selfies. Indeed, they found that the need for popularity significantly predicts selfie behavior among men, but not women. Meanwhile, Weiser (2015) observes that selfie-posting among women shows a stronger association with leadership and/or authority, while men's use of selfies seems to be linked primarily to ideas on entitlement and exploitation.

Unfortunately, the scholarly literature on selfies has tended to focus on one gender (Nelson, 2013; Nguyen, 2014; Warfield, 2014), thereby increasing the need to examine selfie behavior among mixed-sex and mixed-age groups (Albury, 2015). Dhir's (2016) work is one of the few studies to analyze age and gender differences in selfie production and posting. His findings suggest that exploring and building one's online identity plays a key role in shaping the selfie behavior of both adolescents and young adults. Females and adolescents were found to be more active than males and adults in terms of selfie-taking and posting, collecting photos, and photo-editing. However, male adolescents tend to be influenced by photo-tagging gratifications more than

girls, oftentimes using this part of the SNS experience to gain popularity, likes, and comments. Overall, photo-tagging activities tend to satisfy the adolescent's need for self-construction, identity development, and peer approval (Dhir and Torsheim, 2016).

Young adults seem to have little concern about the risks and consequences of selfie-taking/posting (Katz and Crocker, 2015). Young men and women seem to be conscious of their own privacy, as they tend to be aware that not all selfies should be shared with the general public. People might share their own private images without fully realizing it, which suggests that it is necessary to discriminate between private/personal and public/communicative selfies (Albury, 2015). Moreover, boys seem to have more freedom to exhibit their bodies without risk of disapproval. By contrast, young women's pictures (and bodies) are subject to a specific kind of surveillance and criticism (Burns, 2014; Albury, 2015). This suggests that culture and gender needs to be evaluated when considering various aspects of selfie behavior (Doring et al., 2016). Furthermore, gender differences often shape the self-presentation strategies of teens who regularly post selfies.

Expectancies of Internet-Related Behaviors

Expectancies are conscious or unconscious beliefs or thoughts (Goldman, 1994) that reflect the personal beliefs or perceptions about the effect or consequences of a certain behavior (Jung, 2010). The scholarly literature on this topic suggests that personal expectancies influence decisions and behaviors by estimating the consequences of, say, drinking alcohol or engaging in various sexual activities (Dermen and Cooper, 1994; Reich et al., 2010). Indeed, positive outcome expectancies often address and reinforce people's behavior (Patrick and Maggs, 2009).

Addiction research often sees expectancies as "explanatory device[s]" that can analyze the various decision-making processes that often characterize many addictive behaviors (Reich et al., 2010). Debates on Internet addiction have focused on how estimating positive and negative outcomes can impact one's behavior. The influence of expectancies on SNS use has been analyzed in young adults (Turel and Serenko, 2012). Dir et al. (2013) have similarly introduced a measure for sexting expectancies and tested its validity on the development of sexting behaviors among undergraduate students. Finally, Brand et al. (2014) have examined the mediating role of cognitive expectations for Internet use and coping styles in the growth and reinforcement of a Generalized Internet Addiction (GIA). By assuming that addictive Internet use is influenced by Internet-related cognitions (Turel et al., 2011; Xu et al., 2012; Lee et al., 2014), several scholars have stressed that Internet-related expectancies play a significant role in the development of GIA in young adults, males and females alike (Brand et al., 2014). In other words, expectancies mediate between specific personality characteristics and the development of Internet addiction. Indeed, the predictive role of expectancies associated with frequent Internet use on various Internet

communication disorders has been confirmed in young adults (Wegmann and Brand, 2016; Wegmann et al., 2017). However, no specific gender differences have been analyzed in this area.

The Present Study

Despite the popularity of selfies among adolescents, there are few instruments and studies that specifically explore teenage beliefs and expectations about selfies and their consequences. We are unaware of any studies that look at how selfie expectancies and gender guide the selfie-behavior of teenagers. Thus, little is known about the quality of the selfie experience among adolescents. Very little information is available about what boys and girls expect from selfies, and the potential correlations between these expectancies and selfie frequency.

The current study aims to validate a reliable instrument that can identify teenage expectancies about selfie production. This involves:

- evaluating the psychometric properties of a selfie expectancies measure;
- exploring the connections between selfie expectancies and how often individuals create selfies;
- analyzing differences that emerge due to gender.

According to the expectancy theory perspective introduced by Dir et al. (2013) and Brand et al. (2014), we assumed that expectancies regarding the consequences of selfies influences selfie practice, which in turn influences future expectancies.

MATERIALS AND METHODS

Participants

According to Gudmundsson (2009), an instrument must be administered to a fairly large sample to be accurately adapted. Brown (2006) and Kline (2011) suggest using at least 10 subjects per item in order to obtain an adequate sample size for Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). Following these suggestions, our convenience sample was composed of 646 adolescents aged 14–19 years ($M = 16$ years; $SD = 2.519$), all of whom were recruited in six secondary schools (I and II grade) from culturally diverse areas of Naples in Southern Italy. The sample was 58.5% male and 41.5% female, and 97.8% of participants have a smartphone. Of this total, 91.4% use it to make phone calls; 94.8% use it to send messages; 81% use it to exchange photos/videos; and 93.5% use it to surf the Internet. Facebook (77%) and WhatsApp (80%) are the two most popular sites for exchanging messages and photos/videos.

All participants were Caucasians from Italian families. All of them participated in this study on a voluntary basis and were informed about the confidentiality/anonymity of the data. There were no incentives for participation and ethical guidelines from the Helsinki Declaration were followed. In accordance with ethical guidelines that are used by the Italian Psychologists Association and the National Psychologists Council, we asked for consent from both the parents of the participants and the

relevant school boards. Individual consent was considered when the students voluntarily completed the questionnaire. The Local Ethical Committee approved the study.

Measures

Participants answered to a self-report anonymous questionnaire during the school hours. It was comprised of four sections: (1) socio-demographic information, (2) mobile phone/social networks/app usage patterns, (3) the Selfie Frequency Scale (SFS), and (4) a newly developed scale to assess selfie expectancies. Four socio-demographic categories were used: gender, age, school year, and school location (town borough).

Within the second section we asked the participants to refer (1) if they have a smartphone; (2) purpose of using smartphones (for calling, to send messages, share photos/videos, surf the Internet); and (3) which apps and social networks they prefer to use for sharing messages and photos/videos.

The *Selfie Frequency Scale* (SFS) (Manna and Boursier, 2017) is an original 19-item tool that was developed to quantify how often adolescents share selfies ($\alpha = 0.880$). Its structure and dimensions were obtained through a factorial analysis. The measure is based on the assumption that frequency (i.e., the number of times an event occurs) plays a crucial role in determining how adolescents approach the production of selfies. Frequency may provide a consistent measure of problematic selfie behavior from a quantitative point of view. Indeed, frequency may be an indicator of excessive engagement, thus revealing risky behavior. The SFS is a 5-point Likert scale, ranging from 1 (never) to 5 (always), under the query “how often do you. . .” (e.g., take selfies alone, with a friend, etc.; see **Table 1**). The Selfie Frequency Scale includes three items that refer to both the type and frequency of selfies:

- Standard Selfie (F1), which includes eight items regarding the “ordinary” nature of the practice, done in everyday situations with familiar people ($\alpha F1 = 0.838$);
- Sexual Selfie (F2), which is made up of seven items that refer to the tendency to create selfies that feature provocative or sexualized content, alone or with others ($\alpha F2 = 0.839$);
- Friendship Selfie (F3), which includes four items that address companionship and amusement selfies, shared with others and in specific situations ($\alpha F3 = 0.833$).

In the newly developed *Selfie Expectancies Scale* (SES) – the first version of which consisted of 54 items – participants had to state “how much selfie-taking. . .?” by using a 5-point Likert scale. Subsequent statements referred to perceptions about selfies and their possible effects. The scale was developed to both fill a void in the scholarly literature and reinforce the importance of adopting a non-addictive perspective. This measure was based on:

- the various selfie-related behaviors that have been previously addressed in the scholarly literature, with special reference to existing qualitative and quantitative data about selfie outcomes and reasons for taking selfies;
- existing expectancies measures that were developed in previous studies, most notably sexting expectancies

TABLE 1 | Selfie Frequency Scale.**F1 – Standard Selfie** (How often do you...)

1. Take Selfies
2. Take Selfies alone
3. Take Selfies with members of your family
4. Take Selfies with your boyfriend/girlfriend
5. Take Selfies in daily situations (e.g. schools, home...)
6. Take care of your look (hairs, makeup, dresses) when you take Selfies
7. Share Selfies on Facebook or other SNS
8. Share Selfies via smartphone or apps

F2 – Sexual Selfie

9. Take Transgressive Selfies
10. Take Selfies with sexually provocative attitudes alone
11. Take Selfies with sexually provocative attitudes with a friend
12. Take Selfies with sexually provocative attitudes with a group of friends
13. Take Selfies with sexually provocative attitudes with family members
14. Take Selfies with sexually provocative attitudes with your boyfriend/girlfriend
15. Take Selfies showing your body (or part of it) naked/almost naked

F3 – Friendship Selfie

16. Take Funny Selfies
17. Take Selfies with a friend
18. Take Selfies with a group of friends
19. Take Selfies during particular situations (e.g. parties, events, celebrations...)

Cronbach's α :.880

(Weisskirch and Delevi, 2011; Dir et al., 2013) and internet use expectancies (Brand et al., 2014);

- focus groups – carried out in various high schools – that featured adolescents who regularly take and post selfies.

Three core points emerged in the focus groups:

- *Worries*: the perceived effects (i.e., risks and benefits) of selfie production on reputation, relationships, etc.;
- *Attitudes*: treating selfies as a way to satisfy various needs, including sexual fantasies, projecting confidence, etc.;
- *Feelings*: how selfies make someone feel (e.g., excited, anxious, guilty, stupid, dirty, sexy, confident, etc.).

We hypothesized that there would be two overarching types of selfie expectancies:

- *positive expectancies*, which encompasses both positive feelings associated with selfie taking and expectations of positive individual/relational behaviors;
- *negative expectancies*, which entails negative feelings or outcomes that could result from selfie production.

We introduced items referring to negative, positive, and neutral domains. We did not hypothesize, *a priori*, the number of dimensions associated with our expectancies, all of which were based on a factorial analysis. Examples of positive expectancies are included in items stating that selfies might feel participants more popular, more self-confident, or more desired. On the other hand, negative expectancies are expressed under items like “selfie might ruin your relationship/damage your reputation/cause you problems in the future”. Finally, the neutral domain of selfie expectancies is covered by items referring to the widespread use

of selfies (e.g., selfie perceived as a habit or a part of current relationships).

Data Analysis

In order to test the construct validity of the measure, we adopted a random split sample method that divided the overall sample in half. We conducted an EFA on the first half-sample, and then a CFA was performed on the second half-sample in order to confirm the findings from the EFA. This procedure has been adopted in studies that similarly attempted to validate measures for analyzing attitudes (Judd et al., 2014; Martínez et al., 2017). First, we explored the structure of the SES by means of EFA using the software Mplus 6.11 (Muthén and Muthén, 2010). A Robust Maximum Likelihood with oblique Geomin rotation was employed because the sample showed a non-normal distribution. Criteria for identifying the factorial solutions were: (1) a factorial saturation of at least 0.30, (2) the analysis of residuals, and (3) the attempt to avoid elevated cross-loadings (Fabrigar et al., 1999). The scree-plot analysis, the Bartlett's test of sphericity, and KMO measure of sampling adequacy supported the factorial solution. A Confirmatory Factor Analysis (CFA) with Robust Maximum Likelihood was employed to verify the identified factorial solution of the SES and its dimensionality. CFI, RMSEA (90% CI), TLI, and SRMR were used as indexes to evaluate the model fit to the data. We also carried on a second order CFA to test the presence of a single implicit psychological construct and to supplementary verify the construct validity. Cronbach's α , item-total correlations, and factor correlations were adopted to calculate the internal reliability and to examine the internal coherence of the subscales. Bivariate correlations between SES and SFS were conducted to assess the convergent validity and with the purpose of examining the mutual influence of the two measures. A one-sample *t*-test (t ; $p < 0.005$) was calculated with mean values to compare motivations and draw conclusions about the strongest/less strong reasons to selfie practice. The test value referred to the mean of all motivations on the whole sample. Finally, we evaluated the role of gender by means of one-way ANOVAs (F ; $p < 0.005$).

RESULTS

Exploratory and Confirmatory Factor Analysis

During the exploratory analysis, 31 items were removed because of low saturation or high cross-loading. As a result, the final version of the SES consisted of 23 items. EFA on these items yielded all factor loadings greater than 0.3. Both the scree-plot and the eigenvalue suggested a 7-factor solution which explains the 51.26% of variance (Bartlett's test of the sphericity: 0.828) (Table 2). The solution was then verified by means of CFA. The overall model fitted the data adequately ($\chi^2 = 5067.051$ $p = 0.0000$; CFI = 0.962; TLI = 0.954; RMSEA < = 00.05: 0.035; SRMR = 0.046) (Figure 1).

TABLE 2 | Results from Exploratory Factor Analysis.

	F1	F2	F3	F4	F5	F6	F7
F1 – Relational Worries (How much selfie-taking...)							
1. Might damage your reputation	0.745	–0.028	0.081	0.003	–0.049	–0.009	0.0135
2. Might cause you school problems	0.707	0.010	–0.003	–0.056	–0.025	0.114	0.037
3. Might disappoint your parents	0.776	0.054	–0.059	0.007	0.003	0.042	–0.046
4. Might ruin your romantic relationship or the chance to have a relation with someone you are in	0.573	0.067	–0.099	0.093	0.005	–0.236	–0.173
F2 – Web-related anxieties							
5. Might worry you because your photos could end up in the hands of other people who could use them in an unapproved manner	0.172	0.745	0.049	–0.064	0.002	–0.017	–0.002
6. Might worry you because your photos could be retouched	0.024	0.825	0.002	–0.092	–0.018	0.012	–0.086
7. Might worry you because your photos/identity could be stolen	–0.101	0.888	–0.020	0.088	–0.009	0.007	0.078
F3 – Sexual desire							
8. Is exciting	0.064	–0.006	0.560	0.029	0.215	0.070	0.004
9. Improves your sexual fantasies	0.011	–0.009	0.902	0.001	–0.008	–0.061	–0.014
10. Is something your partner expects/would expect from you	–0.116	0.082	0.576	–0.001	–0.093	0.135	–0.030
F4 – Ordinary practice							
11. Is cool	0.034	0.110	–0.017	0.416	0.121	0.109	–0.023
12. Is a part of current relationships	–0.075	0.033	0.108	0.526	0.021	0.182	0.013
13. Is a habit	0.028	–0.070	–0.024	0.725	–0.048	–0.039	0.019
F5 – Self-confidence							
14. Improves your self-esteem	0.014	0.034	0.007	0.137	0.526	0.161	–0.323
15. Makes you feel more popular	–0.025	–0.050	0.051	–0.008	0.533	0.301	0.229
16. Makes you feel more self-confident	–0.020	–0.021	–0.011	0.012	0.914	–0.060	–0.027
17. Makes you feel desired	–0.063	–0.022	0.184	–0.018	0.620	0.122	0.188
F6 – Self-presentation							
18. Is a way to show you off	0.029	0.029	0.045	0.095	–0.088	0.750	0.045
19. Is a way to show to the others the best part of you	0.032	–0.036	0.060	0.045	0.208	0.572	–0.150
20. Is a way to introduce you to the others	–0.027	0.010	–0.052	0.067	0.279	0.522	–0.097
F7 – Generalized risks							
21. Might cause you problems in the future	0.240	0.152	–0.187	0.003	0.099	–0.076	0.384
22. Is risky	0.251	0.287	–0.016	0.042	0.006	0.033	0.408
23. Need to be careful	0.101	0.143	0.058	0.067	0.180	–0.202	0.417
Cronbach's α	0.755	0.861	0.673	0.600	0.837	0.737	0.621

The emerged structure shows that the various expectancies toward selfies suggest the presence of a multicomponent construct that includes references to several different dimensions: the Self, sexual issues, the relational component of identity, and positive or negative perceptions of selfie-behavior. Seven factors were considered:

- *Relational Worries* (F1): this includes the negative consequences of selfie-behavior. They reflect both the individual and relational characterization of selfie-sharing, placing special emphasis on the effects on the self, the family, and personal relationships.
- *Web-related anxieties* (F2): features items that refer to the negative consequences related to the online nature of selfie-behavior. This factor suggests that selfie sharing is perceived by adolescents as being potentially dangerous.
- *Sexual desire* (F3): this factor reinforces the idea that sexuality is an important component of selfies, oftentimes spurring fantasies and feelings of excitement that can be shared with a partner.
- *Ordinary practice* (F4): this factor focuses on items that emphasize the ordinary (and ubiquitous) nature of selfies among adolescents. This dimension might reduce the adolescent's ability to identify the risks connected to behavior that is often considered to be "normal."
- *Self-confidence* (F5): this factor highlights the reinforcing nature of selfies, with special emphasis placed on the extent to which adolescents expect selfies to increase their self-esteem and improve their status among others, thereby increasing their confidence.
- *Self-presentation* (F6): this factor focuses on how adolescents use selfies to show parts of themselves or a specific aspect of the self to the world. The risk of self-manipulation should be considered here, due to a powerful need to be accepted and "liked" by others during this stage of the life cycle.
- *Generalized risks* (F7): this factor keys in on the idea that selfies are dangerous. On the one hand, this factor shows that adolescents understand that there are risks associated with selfie-taking. On the other hand,

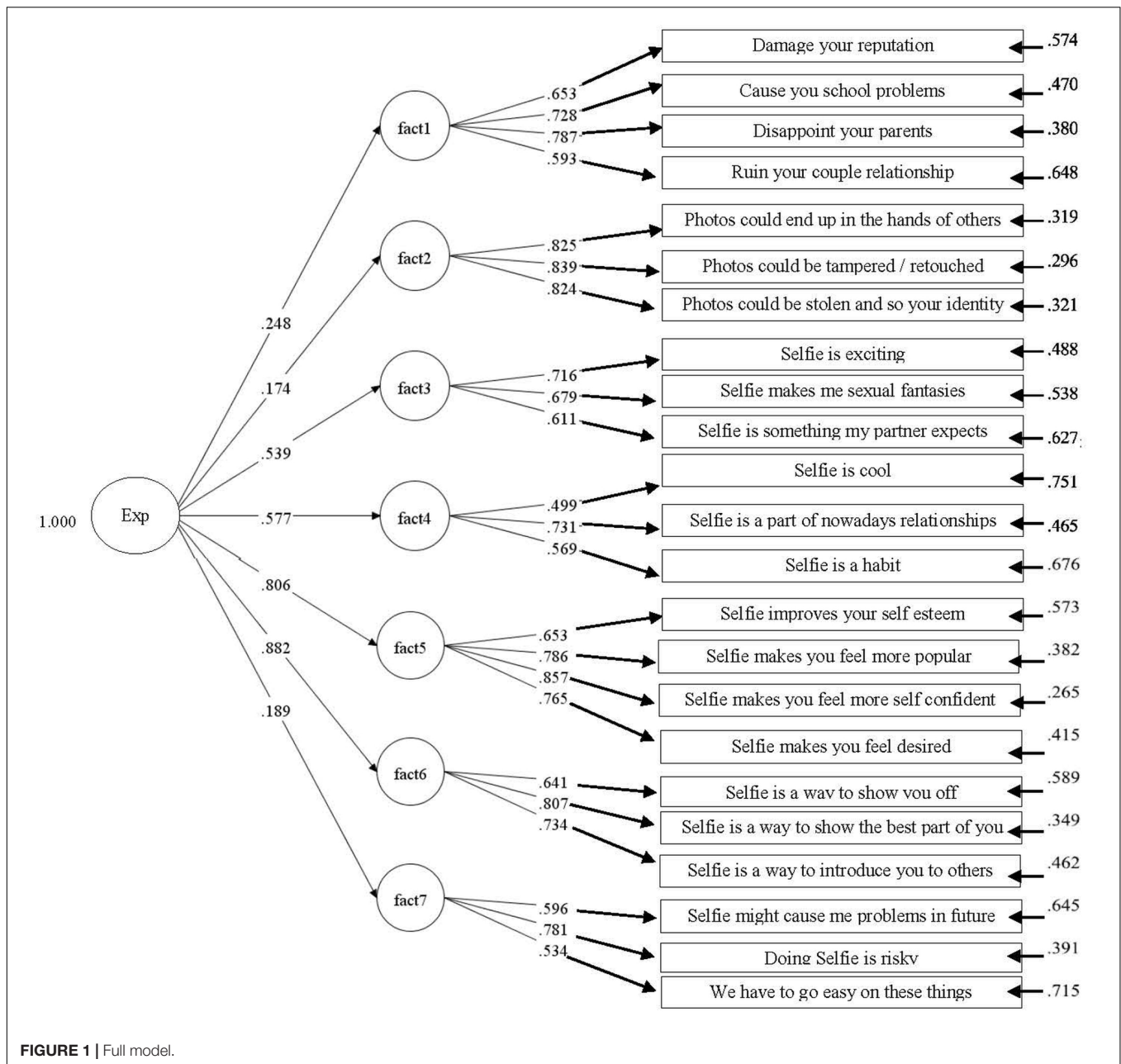


FIGURE 1 | Full model.

these risks might be protective in nature, a means of encouraging adolescents to adopt a safer approach to selfie use.

Convergent Validity

Bivariate correlations showed that Selfie Expectancies and Selfie Frequency assess distinct constructs strongly interrelated ($r = 0.338$; $p < 0.001$), thus confirming the convergent validity of the tool (Table 3). *Self-confidence* is strongly correlated with *Selfie frequency* ($r = 0.413$; $p < 0.001$). Moreover, *Self-presentation* correlates most with *Standard selfie* ($r = 0.415$; $p < 0.001$). One of the strongest correlations emerges between *Sexual desire* and *Sexual selfie* ($r = 0.474$; $p < 0.001$).

Positive expectancies are most negatively correlated with *Web-related anxieties* ($r = -0.512$; $p < 0.001$), and are most positively correlated with *Self-confidence* ($r = 0.582$; $p < 0.001$) and *Self-presentation* ($r = 0.558$; $p < 0.001$). These last two factors also produce the highest levels of inter-correlation ($r = 0.611$; $p < 0.001$) and have the strongest correlation in terms of both frequency and expectancies about selfies. A strong correlation has also been found among *Relational Worries* and *Web-related anxieties* ($r = 0.442$; $p < 0.001$).

Reliability

Cronbach's alphas showed an adequate reliability of the scale ($\alpha = 0.830$) and an acceptable internal consistency for the

TABLE 3 | Correlations among scales and factors.

	SFS	SFS-F1	SFS-F2	SFS-F3	SES	SES-F1	SES-F2	SES-F3	SES-F4	SES-F5	SES-F6	SES-F7
Correlations												
Selfie Frequency Scale	1											
SFS-F1 – Standard Selfie	0.927**	1										
SFS-F2 – Sexual Selfie	0.603**	0.349**	1									
SFS-F3 – Friendship Selfie	0.777**	0.683**	0.218**	1								
Selfie Expectancies Scale	0.338**	0.330**	0.269**	0.158**	1							
SES-F1 – Relational Worries	-0.064	-0.035	-0.142**	-0.030	-0.378**	1						
SES-F2 – Web-related anxieties	-0.067	-0.089	-0.016**	-0.084*	-0.512**	0.442**	1					
SES-F3 – Sexual desire	0.261**	0.157**	0.474**	0.062	0.420**	-0.181**	-0.009	1				
SES-F4 – Ordinary practice	0.130**	0.124**	0.069	0.095*	0.450**	0.003	0.009	0.196**	1			
SES-F5 – Self-confidence	0.413**	0.261**	0.290**	0.211**	0.582**	-0.162**	-0.107**	0.358**	0.120**	1		
SES-F6 – Self-presentation	0.311**	0.415**	0.221**	0.180**	0.558**	-0.145**	-0.051	0.333**	0.168**	0.611**	1	
SES-F7 – Generalized risks	-0.111**	-0.143**	-0.033	-0.043	-0.398**	0.329**	0.370**	-0.096*	-0.008	-0.047	-0.114**	1

**The correlation is significant at level 0.01 (2-codes).

*The correlation is significant at level 0.05 (2-codes).

subscales ($\alpha_{F1} = 0.755$; $\alpha_{F2} = 0.861$; $\alpha_{F3} = 0.673$; $\alpha_{F4} = 0.600$; $\alpha_{F5} = 0.837$; $\alpha_{F6} = 0.737$; $\alpha_{F7} = 0.621$). The solution revealed sufficient inter-item correlations (from 0.255 to 0.742) and significant inter-correlations among its factors ($p < 0.001$).

In terms of the correlations between SES and SFS, *Self-confidence* is strongly correlated with Selfie frequency ($r = 0.413$; $p < 0.001$). Moreover, *Self-presentation* produces the highest correlation with *Standard selfie* ($r = 0.415$; $p < 0.001$). One of the strongest correlations emerges between *Sexual desire* and *Sexual selfie* ($r = 0.474$; $p < 0.001$).

Descriptives and Results From t-Test

Data from the SFS revealed that selfies are a widespread practice: only 3.6% of our sample have never taken a selfie. They are a ubiquitous feature of contemporary youth culture, oftentimes being created during special events ($M = 3.54$; $SD = 1.109$) and in daily situations ($M = 2.81$; $SD = 1.145$). The selfie is a tool for socialization. It is usually taken 2–4 times a week with a boyfriend/girlfriend (84%) or friends (87%), and feature humorous content (64.9%). Selfies are also shared with others by 82% of participants, especially on SNS (59.3%) or WhatsApp groups (60.2%).

Descriptives from the SES and results from the one-sample *t*-test (Table 4) reveal that selfies have a reinforcement function. Indeed, our findings show that selfies are used as a tool to manage self-confidence (F5: $M = 2.45$; $SD = 1.055$), increase self-esteem ($M = 2.42$; $SD = 1.254$), make adolescents feel more self-confident ($M = 2.52$; $SD = 1.298$), and desired ($M = 2.45$; $SD = 1.302$). Secondly, we found that selfies were often used as an instrument to present oneself (F6: $M = 2.40$; $SD = 1.036$), allowing our participants to show off ($M = 2.53$; $SD = 1.229$), introduce themselves to others ($M = 2.46$; $SD = 1.224$), and reveal the best part of themselves to others ($M = 2.22$; $SD = 1.249$).

In terms of negative expectancies, our participants appear particularly worried about web-related anxieties (F2: $M = 2.60$; $SD = 1.279$) and their relationship to various identity issues. They seem especially worried that their photos may end up in the hands of other people who could use them in an unapproved manner ($M = 2.83$; $SD = 1.440$); that their own photos/identity could be stolen ($M = 2.57$; $SD = 1.498$); and that their photos could be tampered with or retouched ($M = 2.41$; $SD = 1.368$). Interestingly enough, web-related anxieties tend to overshadow the positive expectancies (F5 and F6) mentioned earlier.

Our participants are less likely to think that selfies are dangerous (F7: $M = 2.36$; $SD = 0.893$), as many of them refuse to believe that future problems could arise from taking selfies ($M = 1.63$; $SD = 0.950$). However, they are more likely to recognize the necessity to be careful with selfies ($M = 3.00$; $SD = 1.265$), considered as a risky practice in general ($M = 2.46$; $SD = 1.208$). In a similar vein, our participants are not especially concerned about the negative consequences selfies might have on one's self, one's family, or one's personal relationships (F1: $M = 1.65$; $SD = 0.772$). Furthermore, they do not think that selfies are capable of ruining romantic relationships ($M = 1.94$; $SD = 1.207$), damaging one's reputation ($M = 1.74$; $SD = 0.973$), disappointing parents ($M = 1.55$; $SD = 0.968$), or causing school problems ($M = 1.36$; $SD = 0.815$).

TABLE 4 | Descriptives and results from one-sample *t*-test.

	<i>M</i>	<i>SD</i>	<i>t</i> (1.966)	<i>Sig.</i> (2-tailed)
F1 – Relational Worries	1.65	0.772	–23.255	0.000
Might ruin romantic relationship	1.94	1.207	–8.708	0.000
Might damage reputation	1.74	0.973	–16.171	0.000
Might disappoint parents	1.55	0.968	–20.933	0.000
Might cause school problems	1.36	0.815	–30.923	0.000
F2 – Web-related anxieties	2.60	1.279	4.636	0.000
Photos could end up in the hands of other people	2.84	1.440	8.111	0.000
Photos/identity could be stolen	2.57	1.498	3.599	0.000
Photos could be retouched	2.41	1.368	0.803	0.422
F3 – Sexual desire	1.64	0.803	–22.661	0.000
Is exciting	1.80	1.014	–13.853	0.000
Improves your sexual fantasies	1.47	0.978	–22.936	0.000
Is something your partner expects/would expect from you	1.65	1.040	–17.152	0.000
F4 – Ordinary practice	3.58	0.931	33.147	0.000
Is cool	3.86	1.203	31.530	0.000
Is a habit	3.78	1.161	30.905	0.000
Is a part of current relationships	3.10	1.309	14.303	0.000
F5 – Self-confidence	2.45	1.055	2.168	0.031
Increases self-esteem	2.42	1.254	1.210	0.027
Makes feel more self-confident	2.52	1.298	3.166	0.002
Makes feel desired	2.45	1.302	1.773	0.017
Makes popular	2.41	1.257	1.093	0.275
F6 – Self-presentation	2.40	1.036	1.004	0.036
Is a way to show you off	2.53	1.229	3.234	0.001
Is a way to introduce you to the others	2.46	1.224	2.124	0.034
Is a way to show to the others the best part of you	2.22	1.249	–2.869	0.004
F7 – Perceived risks	2.36	0.893	0.174	0.862
Might cause future problems	1.63	0.950	–19.360	0.000
Need to be careful	3.00	1.265	12.770	0.000
Is risky	2.46	1.208	2.200	0.028

Overall, the highest scores were registered in the selfie as an ordinary practice concept (F4: $M = 3.58$; $SD = 0.931$). This suggests that our participants see selfies as a common feature of adolescence – a cool trend ($M = 3.86$; $SD = 1.203$), a habit ($M = 3.78$; $SD = 1.161$) or a key part of contemporary relationships ($M = 3.10$; $SD = 1.309$).

Finally, the sexual aspects of selfies received the lowest scores (F3: $M = 1.64$; $SD = 0.803$). Items from this dimension include: selfies are exciting ($M = 1.80$; $SD = 1.014$); selfies promotes sexual fantasies ($M = 1.47$; $SD = 0.978$); and selfies are something my partner expects/would expect from me ($M = 1.65$; $SD = 1.040$). These results align with the findings from the SFS. Indeed, only 15.9% of participants claimed to have taken transgressive selfies, while only 11.1% claimed to have taken provocative selfies. As a result, it is safe to say that although selfies have a sexual component, adolescents don't consider this a major feature of the selfie-taking process.

Gender Differences

Our findings suggest that a moderate role is played by gender. The SFS found that although selfies, in general, are more common among females ($M_F = 3.79$; $SD_F = 0.912$; $M_M = 3.12$;

$SD_M = 0.959$), selfies with sexual content are more common among males ($M_F = 1.21$; $SD_F = 0.628$; $M_M = 1.35$; $SD_M = 0.778$). Indeed, males registered a higher prevalence on all items related to the sexual, provocative, and transgressive nature of selfies. No gender differences were found in items that focused on friends, SNS use, and apps, thus confirming that selfies are used primarily as a tool for managing and sharing information about relationships.

Nonetheless, some gender differences were found in several factors. ANOVAs performed on the SES, for instance, revealed significant preoccupation levels among girls. As shown in **Table 5**, girls report more web-related anxieties (F2: $M_F = 2.86$; $SD_F = 1.337$; $M_M = 2.40$; $SD_M = 1.201$) and perceived risks (F7: $M_F = 2.46$; $SD_F = 0.911$; $M_M = 2.30$; $SD_M = 0.875$). The only concern that is greater among males than among females is the fear that selfies might ruin a personal relationship ($M_F = 1.73$; $SD_F = 1.095$; $M_M = 2.09$; $SD_M = 1.261$).

Boys are more likely to see selfies in a sexual light, placing special emphasis on self-attractiveness (F3: $M_F = 1.37$; $SD_F = 0.559$; $M_M = 1.83$; $SD_M = 0.890$). Selfies are exciting to boys; they contribute to their sexual fantasies and often lead to expectations that their partners should create similarly explicit

TABLE 5 | One-way ANOVAs by gender with means and standard deviations for gender variables.

	<i>M (SD)</i>		<i>F</i>	<i>Sig.</i>	η^2
	<i>Males</i>	<i>Females</i>			
F1 - Relational Worries	1.67 (0.732)	1.61 (0.824)	1.132	0.288	0.002
Might ruin romantic relationship	2.09 (1.26)	1.73 (1.09)	14.026	0.000*	0.132
Might damage reputation	1.74 (0.973)	1.76 (1.04)	0.246	0.620	0.000
Might disappoint parents	1.72 (0.920)	1.56 (1.06)	0.003	0.959	0.000
Might cause school problems	1.55 (0.896)	1.34 (0.741)	0.739	0.390	0.001
F2 - Web-related anxieties	2.40 (1.20)	2.86 (1.33)	20.622	0.000*	0.132
Photos could end up in the hands of other people	2.63 (1.39)	3.10 (1.45)	16.617	0.000*	0.126
Photos/identity could be stolen	2.36 (1.43)	2.88 (1.53)	18.752	0.000*	0.129
Photos could be retouched	2.23 (1.29)	2.65 (1.43)	14.515	0.000*	0.133
F3 - Sexual desire	1.83 (0.890)	1.36 (0.559)	56.248	0.000*	0.181
Is exciting	1.99 (1.09)	1.53 (0.813)	33.583	0.000*	0.150
Improves your sexual fantasies	1.69 (1.11)	1.16 (0.632)	49.846	0.000*	0.173
Is something your partner expects/would expect from you	1.82 (1.11)	1.42 (0.871)	24.108	0.000*	0.012
F4 - Ordinary practice	3.55 (0.944)	3.61 (0.913)	0.467	0.494	0.001
Is cool	3.82 (1.21)	3.92 (1.19)	0.995	0.319	0.002
Is a habit	3.69 (1.21)	3.91 (1.06)	5.727	0.017	0.009
Is a part of current relationships	3.17 (1.29)	3.01 (1.33)	2.368	0.124	0.004
F5 - Self-confidence	2.48 (1.06)	2.39 (1.04)	1.098	0.295	0.002
Improves self-esteem	2.33 (1.21)	2.55 (1.29)	4.685	0.031	0.007
Makes feel more self-confident	2.50 (1.30)	2.56 (1.28)	0.383	0.536	0.001
Makes feel desired	2.60 (1.32)	2.24 (1.23)	11.798	0.001*	0.018
Makes popular	2.53 (1.28)	2.25 (1.19)	8.025	0.005*	0.012
F6 - Self-presentation	2.47 (1.05)	2.29 (1.00)	5.070	0.025	0.008
Is a way to show you off	2.68 (1.30)	2.31 (1.25)	13.231	0.000*	0.021
Is a way to introduce you to the others	2.53 (1.24)	2.36 (1.16)	3.223	0.073	0.005
Is a way to show to the others the best part of you	2.23 (1.25)	2.20 (1.24)	0.091	0.763	0.000
F7 - Perceived risks	2.29 (0.875)	2.46 (0.910)	5.341	0.021	0.008
Might cause future problems	1.63 (0.950)	1.63 (0.952)	0.007	0.934	0.000
Need to be careful	2.96 (1.24)	3.05 (1.29)	0.749	0.387	0.001
Is risky	2.30 (1.18)	2.70 (1.20)	17.696	0.000*	0.027

* $p < 0.005$.

content. Boys also have greater positive expectancies, as they tend to consider selfies as self-presentation tools (F6: $M_F = 2.29$; $SD_F = 1.006$; $M_M = 2.47$; $SD_M = 1.051$) that are connected to their sexual desires.

Since girls are more likely to regard selfie-taking as a risky practice ($M_F = 2.70$; $SD_F = 1.209$; $M_M = 2.30$; $SD_M = 1.188$), they might be more cognizant of the negative consequences of posting selfies. Among boys, by contrast, selfies are tied to excitement, sexual desire, and managing their self-image. Selfies, in short, help boys feel more desired ($M_F = 2.24$; $SD_F = 1.237$; $M_M = 2.60$; $SD_M = 1.328$), providing them with a venue in which they can show off to their friends ($M_F = 2.31$; $SD_F = 1.256$; $M_M = 2.68$; $SD_M = 1.308$).

These findings should consider the magnitude of effect size, as given by the η^2 . According to Pierce et al. (2004), a η^2 value lower than 0.13 is considered small, a value from 0.13 to 0.23 is moderate, and values higher than 0.23 are considered large. Using this criterion as a guide, our data set revealed moderate effects of gender on Sexual desire and Web-related anxieties. In fact, 18.1% of the variance found in the Sexual

desire dimension can be attributed to gender, especially items pertaining to excitement ($\eta^2 = 0.150$) and sexual fantasies ($\eta^2 = 0.173$). Moreover, 13.2% of the variance in Web-related anxieties is due to gender, as a moderate effect has been found in all of the items (selfie practice may ruin a personal relationship: $\eta^2 = 0.132$; photos could end up in the hands of other people: $\eta^2 = 0.126$; photos could be tampered with or retouched: $\eta^2 = 0.133$; and photos/identity could be stolen: $\eta^2 = 0.129$). All the other differences that arose due to gender are significant, but not to the same extent as the items discussed above. Nonetheless, the idea that boys are more involved in the sexualized aspects of selfie-behavior, and that girls are more worried about the negative consequences of selfies, requires further research.

DISCUSSION

Unfortunately, the scholarly literature that has emerged in recent years on selfie culture doesn't address age and gender differences.

Scholars have shown that both age and gender affect the way the Internet and SNSs are utilized (Albury, 2015), and yet few studies have investigated social media use and selfie practices among people of different age and gender (Dhir et al., 2016, 2017).

This study contributes to the ongoing scientific debate on the psychological functions and attitudes implied in selfie-behavior, as well as the motivations behind this practice. Moreover, the trend to medicalize everyday behavior has influenced this study by allowing us explore selfie production among adolescents without adopting an addiction/medicalized perspective (Starcevic et al., 2018).

Furthermore, this study has a unique age/gender viewpoint. Indeed, these themes were explored with special reference to selfie diffusion among adolescents, many of whom are engaged in self-definition, identity construction, and relational interactions. In fact, selfies may help individuals express and fortify their own identity in an online context. According to some scholars (Nadkarni and Hofmann, 2012; Nguyen, 2014; Katz and Crocker, 2015; Sorokowska et al., 2016; Diefenbach and Christoforakos, 2017; Etgar and Amichai-Hamburger, 2017; Taylor et al., 2017; Reich et al., 2018), self-presentation, self-promotion, and self-approval are prominent features of selfie experience.

If we assume that expectations play a key role in determining people's behavior, then it is safe to say that a measure that is specifically oriented to assess selfie expectations could be especially valuable to both scholars and practitioners. This study aimed to validate a psychometric tool that can be used to assess expectations toward selfies among adolescents. This tool overcomes the shortcomings of extant instruments, and allows us to better recognize what motivates adolescents to create selfies, without necessarily treating it as symptomatic behavior or a unique psychiatric issue.

The proposed 7-factor model fitted the data adequately, while also highlighting that positive, negative, and neutral consequences need to be considered. Our sample showed that selfies were most often created via smartphones, and that selfies are a key component of contemporary adolescence. Selfie creation is neither positive nor negative, but strongly related to the customs and habits of millennials.

Positive expectations toward selfies are related to the idea that selfies are a tool for self-presentation and self-promotion, which in turn are related to self-disclosure and self-management strategies. The use of selfies to garner approval (and feelings of gratification) from one's peers and improve one's self-esteem, self-confidence, and popularity has been confirmed by previous research in this area (Etgar and Amichai-Hamburger, 2017). According to Diefenbach and Christoforakos (2017), selfie-taking may play a key role in self-presentation and self-promotion. Moreover, our study found that the process of taking selfies among adolescents often focuses on choosing what to show others, which suggests that adolescents fear having their images tampered with or manipulated (McLean et al., 2015; Chae, 2017). Additionally, the sexual aspects of selfies emerged as a constitutive dimension of selfie expectations, especially among boys who were concerned with self-attractiveness issues. In other words, selfies are often

used by our participants to manage a host of identity-related issues.

Differently from Diefenbach and Christoforakos' (2017) study on young adults, neither positive aspects due to the authentic expression of oneself, nor concerns about the illusory dimension of selfies emerged in our results. However, common risks related to the general consequences of selfies are considered here, even though these concerns don't weigh as heavily among our participants as web-related anxieties. Our participants were worried about losing control of their self-images – for example, that their selfies may end up in the hands of other people who could use them for unapproved purposes; that their photos could be tampered with or retouched by others; or that their photos/identities could be stolen – especially among girls. Privacy concerns (Livingstone, 2008) tend to overshadow the positive expectations related to self-confidence and self-presentation. Indeed, self-disclosure can often result in criticism and negative opinions from others, including hostile assessments from total strangers, which explains why the adolescents in our study were well aware of the negative consequences of web-exposure. As we know, privacy disturb online self-presentation (Wang et al., 2011; Kaur et al., 2016), however, Dhir et al. (2017) recently analyzed the “privacy paradox” (Barnes, 2006), a concept that addresses privacy concerns and online self-disclosure through selfies. Privacy concerns seem to affect women more than men, and young adults more than adolescents and adults. Regardless, this doesn't necessarily result in lower selfie activity, as privacy concerns seem to be inversely related to selfie taking/posting (Dhir et al., 2017).

The results from our sample confirm this paradox. Even though girls are more likely than boys to see selfies as a somewhat risky practice and worry about the consequences of posting selfies, this activity is more common among girls. By contrast, boys tend to see selfies (and web exposure in general) as a form of self-promotion. This is in line with Kim and Chock's (2015) findings on the importance of popularity in shaping selfie behavior among males – a notion that was similarly confirmed in Dhir and Torsheim's (2016) work on photo-tagging among boys. Furthermore, our study shows that the appeal of selfies among boys is also tied to ideas about excitement and sexual desire.

Our findings suggest that selfie expectations among boys and girls are quite different, and that selfie-behavior is a decidedly gendered phenomenon. As Doring et al. (2016) have noted, cultural stereotypes and social differences between boys and girls should be considered when studying the importance of selfies among adolescents and young adults.

The measure presented in this study can reliably assess adolescent expectations toward selfies and ought to be used in further research on generalized or specific selfie behavior. For instance, using selfies as both a self-promotion tool and as a means of improving one's self-confidence needs to be considered. The tendency to show only the best part of oneself, or to present a modified representation of oneself via photos, is another aspect of selfie culture that needs to be evaluated. Moreover, if we assume that selfies can be used for self-support and aid in self-construction, then it makes sense that creating selfies

in hopes of receiving the approval of others should be analyzed. Our study found that although being aware of the consequences of web-exposure encouraged a host of anxieties, it didn't necessarily lower the frequency of selfie production among adolescents. This is probably a product of the ubiquitous nature of selfie culture nowadays, as well as the influence of one's personality, impulsivity, emotional state, and unconscious motivations. Since identity, body-image, and related factors play significant roles in selfie behaviors, our findings point toward the necessity of promoting preventive programs that are differentiated by gender and take into account a wide array of dimensions.

Limitations and Suggestions for Future Research

The reported findings should be interpreted by taking in account some limitations of the study.

For starters, the external validity of the findings may be limited by the sampling technique, which was based on a non-probability procedure of recruitment of the participants (see, for example, Mann, 2003; Balsamo et al., 2013). Anyway, we haven't been able to find any other research that adequately discusses this specific topic.

Potential biases (e.g., social desirability biases) due to a self-report questionnaire are well known. However, we considered the relevant advantages provided by this kind of tool, such as: the possibility to collect a rich amount of information, the interpretability, the practicality of the administration and the participants' motivation to share their opinions (Paulhus and Vazire, 2007).

Even though this study featured a large sample of adolescents, our research was limited to one specific geographic area. Future research should include different regions of Italy in order to compare findings from, say, Northern and Southern Italy. The findings of any study often depend on cultural aspects that should be addressed in future research. Indeed, a cross-cultural perspective could shed light on our own findings in interesting and provocative ways.

Exploratory and Confirmatory Factorial Analysis have been conducted on our sample, even though our sample was split into two half-samples. This approach was chosen due to the difficulties in tracking down a large group of participants. However, this strategy is largely adopted to validate new measures for analyzing attitudes. Generally speaking, conducting a new CFA on different

samples could help us better confirm the dimensionality and validity of the measure.

The present study also has some key strengths that are worth noting. For instance, our research represents an important step in examining selfie behaviors among adolescents, providing a short and psychometrically valid measure to assess the expectations of teenagers who take part in selfie practice. Moreover, given the strong psychometrics of the instrument, researchers are encouraged to consider using this tool to assess the quality selfie-related behavior in samples of adolescents.

This study also complements previous qualitative and quantitative findings on how age and gender often shapes (and predicts) selfie behaviors (Nelson, 2013; Nguyen, 2014; Warfield, 2014; Christoforakos and Diefenbach, 2016; Dhir et al., 2017; Diefenbach and Christoforakos, 2017). It also provides a new understanding of selfie culture by engaging with a demographic that hasn't been studied much in Italy.

Lastly, this study has some important clinical implications. Chief among them is the tendency among girls to use selfies as a means of managing various identity issues, as well as the tendency among boys to focus on sexual matters, most notably self-attractiveness issues.

CONCLUSION

This study provides a new means of analyzing selfie behavior among adolescents. It examines seven important motivations and expectations that often shape the production of selfies. Our findings build on previous research on selfie behavior among millennials, while also highlighting the importance of studying the influence of age and gender on selfie-related behavior. Indeed, our selfie expectations scale should be seen as a useful tool that can help scholars and practitioners alike better understand a multifaceted and widespread phenomenon.

AUTHOR CONTRIBUTIONS

VB and VM both designed and conducted the study. VB led the literature search. VM analyzed the data. Both authors contribute to the interpretation and discussion of data and approved the final version of the manuscript for submission and agreed to be accountable for all aspects of the work.

REFERENCES

- Albury, K. (2015). Selfies, sexts, and sneaky hats: young people's understandings of gendered practices of self-representation. *Int. J. Commun.* 9, 1734–1745.
- Baiocco, R., Chirumbolo, A., Bianchi, D., Ioverno, S., Morelli, M., and Nappa, M. R. (2016). How HEXACO personality traits predict different selfie-posting behaviors among adolescents and young adults. *Front. Psychol.* 7:2080. doi: 10.3389/fpsyg.2016.02080
- Balakrishnan, J., and Griffiths, M. D. (2017). An exploratory study of 'selfitis' and the development of the Selfitis Behavior Scale. *Int. J. Ment. Health. Ad.* 1–15. doi: 10.1007/s11469-017-9844-x
- Balsamo, M., Lauriola, M., and Saggino, A. (2013). Work values and college major choice. *Learn. Individ. Differ.* 24, 110–116. doi: 10.1016/j.lindif.2012.12.022
- Barnes, S. B. (2006). *A Privacy Paradox: Social Networking in the United States*. Bridgman, MI: First Monday, 11.
- Barry, C. T., Doucette, H., Loflin, D. C., Rivera-Hudson, N., and Herrington, L. L. (2017). Let me take a selfie: associations between self-photography, narcissism, and self-esteem. *Psychol. Pop. Media. Cult.* 6, 48–60. doi: 10.1037/ppm0000089
- Billieux, J., Schimmenti, A., Khazaal, Y., Maurage, P., and Heeren, A. (2015). Are we overpathologizing everyday life? A tenable blueprint for behavioral addiction research. *J. Behav. Addict.* 4, 119–123. doi: 10.1556/2006.4.2015.009
- Brand, M., Laier, C., and Young, K. S. (2014). Internet addiction: coping styles, expectancies, and treatment implications. *Front. Psychol.* 5:1256. doi: 10.3389/fpsyg.2014.01256
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Press.

- Burns, A. (2014). *A Handy Summary of What We Know So Far*. Available at: <https://thecarceralnet.wordpress.com/2014/07/11/a-handy-summary-of-what-we-know-so-far/>
- Chae, J. (2017). Virtual makeover: selfie-taking and social media use increase selfie-editing frequency through social comparison. *Comput. Hum. Behav.* 66, 370–376. doi: 10.1016/j.chb.2016.10.007
- Christoforakos, L., and Diefenbach, S. (2016). “Das Selfie-Paradoxon – Kaum einer mag sie, alle tun es,” in *Poster Presentation at the German Psychological Society (DGPs) Congress*, Leipzig.
- Dermen, K. H., and Cooper, M. L. (1994). Sex-related alcohol expectancies among adolescents: I. Scale development. *Psychol. Addict. Behav.* 8, 152–160. doi: 10.1037/0893-164X.8.3.152
- Dhir, A. (2016). Why do young people avoid photo tagging? A new service avoidance scale. *Soc. Sci. Comput. Rev.* 35, 480–497. doi: 10.1177/0894439316653636
- Dhir, A., Ståle Pallesen, S., Torsheim, T., and Schou Andreassen, C. (2016). Do age and gender differences exist in selfie-related behaviours? *Comput. Hum. Behav.* 63, 549–555. doi: 10.1016/j.chb.2016.05.053
- Dhir, A., and Torsheim, T. (2016). Age and gender differences in photo tagging gratifications. *Comput. Hum. Behav.* 63, 630–638. doi: 10.1016/j.chb.2016.05.044
- Dhir, A., Torsheim, T., Pallesen, S., and Andreassen, C. S. (2017). Do online privacy concerns predict selfie behavior among adolescents, young adults and adults? *Front. Psychol.* 8:815. doi: 10.3389/fpsyg.2017.00815
- Diefenbach, S., and Christoforakos, L. (2017). The selfie paradox: nobody seems to like them et everyone has reasons to take them. An exploration of psychological functions of selfies in self-presentation. *Front. Psychol.* 8:7. doi: 10.3389/fpsyg.2017.00007
- Dir, A. L., Coskunpinar, A., Steiner, J. L., and Cyders, M. A. (2013). Understanding differences in sexting behaviors across gender, relationship status, and sexual identity, and the role of expectancies in sexting. *Cyberpsychol. Behav. Soc. Netw.* 16, 568–574. doi: 10.1089/cyber.2012.0545
- Doring, N., Reif, A., and Poeschl, A. (2016). How gender-stereotypical are selfies? A content analysis and comparison with magazine adverts. *Comput. Hum. Behav.* 55, 955–962. doi: 10.1016/j.chb.2015.10.001
- Etgar, S., and Amichai-Hamburger, Y. (2017). Not all selfies took alike: distinct selfie motivations are related to different personality characteristics. *Front. Psychol.* 8:842. doi: 10.3389/fpsyg.2017.00842
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., and Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychol. Methods.* 4, 272–299. doi: 10.1037/1082-989X.4.3.272
- Fox, J., and Rooney, M. C. (2015). The Dark Triad and trait self-objectification as predictors of men’s use and self-presentation behaviors on social networking sites. *Pers. Ind. Dif.* 76, 161–165. doi: 10.1016/j.paid.2014.12.017
- Gibbs, M., Carter, M., Nansen, B., and Kohn, T. (2014). Selfies at funerals:remediating rituals of mourning. *Sel. Papers Int. Res.* 4.
- Goldman, M. S. (1994). The alcohol expectancy concept: applications to assessment, prevention, and treatment of alcohol abuse. *Appl. Prev. Psychol.* 3, 131–144. doi: 10.1016/S0962-1849(05)80066-6
- Griffiths, M. D., and Balakrishnan, J. (2018). The psychosocial impact of excessive selfie-taking in youth: a brief overview. *Educ. Health* 36, 3–6.
- Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nord. Psychol.* 61, 29–45.
- Halpern, D., Valenzuela, S., and Katz, J. E. (2016). “Selfie-ists” or “Narci-selfiers”? a cross-lagged panel analysis of selfie taking and narcissism. *Pers. Individ. Differ.* 97, 98–101. doi: 10.1016/j.paid.2016.03.019
- Judd, S. E., Letter, A. J., Shikany, J. M., Roth, D. L., and Newby, P. K. (2014). Dietary patterns derived using exploratory and confirmatory factor analysis are stable and generalizable across race, region, and gender subgroups in the REGARDS study. *Front. Nutr.* 1:29. doi: 10.3389/fnut.2014.00029
- Jung, J. (2010). *Alcohol, Other Drugs, and Behavior: Psychological Research Perspectives*, 2nd edn. Los Angeles: SAGE.
- Kardefelt-Winther, D., Heeren, A., Schimmenti, A., Rooij, A., Maurage, P., Carras, M., et al. (2017). How can we conceptualize behavioural addiction without pathologizing common behaviours? *Addiction* 112, 1709–1715. doi: 10.1111/add.13763
- Katz, J. E., and Crocker, E. T. (2015). Selfies and photo messaging as visual conversation: reports from the United States, United Kingdom, and China. *Int. J. Commun.* 9, 1861–1872.
- Kaur, P., Dhir, A., Chen, S., and Rajala, R. (2016). Understanding online regret experience using the theoretical lens of flow experience. *Comput. Hum. Behav.* 57, 230–239. doi: 10.1016/j.chb.2015.12.041
- Kim, J. W., and Chock, M. T. (2015). Body image 2.0: Associations between social grooming on Facebook and body image concerns. *Comput. Hum. Behav.* 48, 331–339. doi: 10.1016/j.chb.2015.01.009
- Kiprin, B. (2013). *Go Selfie Yourself*. Available at: <https://borislavkiprin.com/2013/12/13/go-selfie-yourself/>
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modelling*. London: The Guilford Press.
- Lee, J. A., and Sung, Y. (2016). Hide-and-seek: Narcissism and selfie-related behavior. *Cyberpsychol. Behav. Soc. Netw.* 19, 347–351. doi: 10.1089/cyber.2015.0486
- Lee, Y. H., Ko, C. H., and Chou, C. (2014). Re-visiting internet addiction among Taiwanese students: a cross-sectional comparison of students’ expectations, online gaming, and online social interaction. *J. Abnorm. Child Psychol.* 43, 589–599. doi: 10.1007/s10802-014-9915-4 [Epub ahead of print].
- Lenhart, A., Purcell, K., Smith, A., and Zickuhr, K. (2010). *Social Media and Mobile Internet Use Among Teens and Young Adults*. *Pew Internet and American Life Project*. Available at: http://pewinternet.org/~{}media/Files/Reports/2010/PIP_Social_Media_and_Young_Adults_Report_Final_with_toplines.pdf [accessed November 12, 2012].
- Livingstone, S. (2008). Taking risky opportunities in youthful content creation: teenagers’ use of social networking sites for intimacy, privacy and selfexpression. *New Media Soc.* 10, 393–411. doi: 10.1177/1461444808089415
- Locateadoc.com. (2014). *Infographic: Selfie Psychology Stats and Info*. Available at: <http://www.locateadoc.com/article/infographic-selfiepsychology-stats-and-info> [accessed October 10, 2015].
- Mann, C. (2003). Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emerg. Med. J.* 20, 54–60. doi: 10.1136/emj.20.1.54
- Manna, V., and Boursier, V. (2017). “How often do you...? A study of sexting and selfie prevalence among adolescents,” in *Paper Presented at the 15th European Congress of Psychology*, Amsterdam. July 11-14 (Amsterdam).
- Martínez, I., Cruise, E., García, Ó. F., and Murgui, S. (2017). English validation of the parental socialization scale—ESPA29. *Front. Psychol.* 8:865. doi: 10.3389/fpsyg.2017.00865
- McCain, J. L., Borg, Z. G., Rothenberg, A. H., Churillo, K. M., Weiler, P., and Campbell, W. K. (2016). Personality and selfies: Narcissism and the Dark Triad. *Comput. Hum. Behav.* 64, 126–133. doi: 10.1016/j.chb.2016.06.050
- McLean, S. A., Paxton, S. J., Wertheim, E. H., and Masters, J. (2015). Photoshopping the selfie: self photo editing and photo investment are associated with body dissatisfaction in adolescent girls. *Int. J. Eat. Disord.* 48, 1132–1140. doi: 10.1002/eat.22449
- Muthén, L. K., and Muthén, B. O. (2010). *Mplus User’s Guide: Statistical Analysis with Latent Variables: User’s Guide*. Los Angeles, CA: Muthén & Muthén.
- Nadkarni, A., and Hofmann, S. G. (2012). Why do people use Facebook? *Pers. Individ. Differ.* 52, 243–249. doi: 10.1016/j.paid.2011.11.007
- Nelson, O. (2013). *Dark Undercurrents of Teenage Girls’ Selfies*. Melbourne, VIC: The Age.
- Nguyen, A. J. (2014). *Exploring The Selfie Phenomenon: The Idea of Self-Presentation and Its Implications Among Young Women*. Master’s thesis. Northampton: Smith College.
- Nguyen, A. J., and Barbour, K. (2017). *Selfies as Expressively Authentic Identity Performance*, Vol. 22. Bridgman, MI: First Monday, 11. doi: 10.5210/fm.v22i11.7745
- Oxford Dictionaries (2013). *Word of the Year 2013*. Available at: <https://en.oxforddictionaries.com/word-of-the-year/word-of-the-year-2013>
- Patrick, M. E., and Maggs, J. L. (2009). Does drinking lead to sex? Daily alcohol–sex behaviors and expectancies among college students. *Psychol. Addict. Behav.* 23, 472–481. doi: 10.1037/a0016097
- Paulhus, D. L., and Vazire, S. (2007). “The self-report method,” *Handbook of Research Methods in Personality Psychology*, Vol. 1, eds R. W. Robins, R. C. Fraley, and R. F. Krueger (New York, NY: Guilford Press), 224–239.

- Pierce, C. A., Block, R. A., and Aguinis, H. (2004). Cautionary note on reporting Eta-squared values from multifactor ANOVA designs. *Educ. Psych. Meas.* 64, 916–924. doi: 10.1177/0013164404264848
- Poe, J. (2015). “The link between “likes” and self-worth: how women use selfies on Instagram for self-presentation and the effects of social comparison,” in *14th Annual Celebration for Undergraduate Research and Creative Performance*, 24.
- Qiu, L., Lu, J., Yang, S., Qu, W., and Zhu, T. (2015). What does your selfie say about you? *Comput. Hum. Behav.* 52, 443–449. doi: 10.1016/j.chb.2015.06.032
- Reich, R. R., Below, M. C., and Goldman, M. S. (2010). Explicit and implicit measures of expectancy and related alcohol cognitions: a meta-analytic comparison. *Psychol. Addict. Behav.* 24, 13–25. doi: 10.1037/a0016556
- Reich, S., Schneider, F. M., and Heling, L. (2018). Zero Likes – Symbolic interactions and need satisfaction online. *Comput. Hum. Behav.* 80, 97–102. doi: 10.1016/j.chb.2017.10.043
- Senft, T. M., and Baym, N. K. (2015). What does the selfie say? investigating a global phenomenon. *Int. J. Commun.* 9, 1588–1606.
- Sorokowska, A., Oleszkiewicz, A., Frackowiak, T., Pisanski, K., Chmiel, A., and Sorokowski, P. (2016). Selfies and personality: who posts self-portrait photographs? *Pers. Individ. Differ.* 90, 119–123. doi: 10.1016/j.paid.2015.10.037
- Sorokowski, P., Sorokowska, A., Oleszkiewicz, A., Frackowiak, T., Huk, A., and Pisanski, K. (2015). Selfie posting behaviors are associated with narcissism among men. *Pers. Ind. Differ.* 85, 123–127. doi: 10.1016/j.paid.2015.05.004
- Starcevic, V., Billieux, J., and Schimmenti, A. (2018). Selfitis, selfie addiction, twitteritis: irresistible appeal of medical terminology for problematic behaviours in the digital age. *Aust. N. Z. J. Psychiatry* 52, 408–409. doi: 10.1177/0004867418763532
- Sung, Y., Lee, J. A., Kim, E., and Choi, S. M. (2016). Why we post selfies: understanding motivations for posting pictures of oneself. *Pers. Ind. Differ.* 97, 260–265. doi: 10.1016/j.paid.2016.03.032
- Taylor, S. A., Hinck, A. S., and Lim, H. (2017). An experimental test of how selfies change social judgments on facebook. *Cyberpsych. Beh. Soc. N.* 20, 610–614. doi: 10.1089/cyber.2016.0759
- Turel, O., Serenko, A., and Giles, P. (2011). Integrating technology addiction and use: an empirical investigation of online auction users. *MIS Q.* 35, 1043–1061. doi: 10.2307/41409972
- Turel, O., and Serenko, A. (2012). The benefits and dangers of enjoyment with social networking websites. *Eur. J. Inf. Syst.* 21, 512–528. doi: 10.1057/ejis.2012.1
- Wang, Y., Komanduri, S., Leon, P., Norcie, G., Acquisti, A., and Cranor, L. (2011). “I regretted the minute I pressed share: a qualitative study of regrets on Facebook,” in *Proceedings of the 7th Symposium on Usable Privacy and Security*, Pittsburgh, PA, 1–16. doi: 10.1145/2078827.2078841
- Warfield, K. (2014). *Making Selfies/Making Self: Digital Subjectivities in the Selfie*. Berlin: Image Conference.
- Wegmann, E., and Brand, M. (2016). Internet-communication disorder: it’s a matter of social aspects, coping, and internet-use expectancies. *Front. Psych.* 7:1747. doi: 10.3389/fpsyg.2016.01747
- Wegmann, E., Oberst, U., Stodt, B., and Brand, M. (2017). Online-specific fear of missing out and internet-use expectancies contribute to symptoms of Internet-communication disorder. *Addict. Behav.* 5, 33–42. doi: 10.1016/j.abrep.2017.04.001
- Weiser, E. B. (2015). #Me: narcissism and its facets as predictors of selfie posting frequency. *Pers. Ind. Dif.* 86, 477–481. doi: 10.1016/j.paid.2015
- Weiser, E. B. (2018). “Shameless selfie promotion: narcissism and its association with selfie-posting behavior,” in *Selfies as a Mode of Social Media and Work Space Research*, ed. S. Hai-Jew (Hershey, PA: Igi Global), 1–27. doi: 10.4018/978-1-5225-3373-3.ch001
- Weisskirch, R. S., and Delevi, R. (2011). “Sexting” and adult romantic attachment. *Comput. Hum. Behav.* 27, 1697–1701. doi: 10.1016/j.chb.2011.02.008
- Xu, Z. C., Turel, O., and Yuan, Y. F. (2012). Online game addiction among adolescents: motivation and prevention factors. *Eur. J. Inf. Syst.* 21, 321–340. doi: 10.1057/ejis.2011.56
- Yongjun, S., Lee, J., Kim, E., and Sejung Marina Choi, S. M. (2016). Why we post selfies: understanding motivations for posting pictures of oneself. *Pers. Ind. Differ.* 97, 260–265. doi: 10.1016/j.paid.2016.03.032

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

Copyright © 2018 Boursier and Manna. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Psychometric Properties of the Italian Version of the Young Schema Questionnaire L-3: Preliminary Results

Aristide Saggino^{1,2*}, Michela Balsamo¹, Leonardo Carlucci¹, Veronica Cavalletti³, Maria R. Sergi¹, Giorgio da Fermo^{4,5}, Davide Dèttore⁶, Nicola Marsigli³, Irene Petrucci⁷, Susanna Pizzo⁸ and Marco Tommasi^{1,2}

¹ School of Medicine and Health Sciences, Università degli Studi 'G. d'Annunzio' Chieti - Pescara, Chieti, Italy, ² Center for the Study of Personality, Napoli, Italy, ³ IPSICO - Istituto di Psicologia e Psicoterapia Comportamentale e Cognitiva, Firenze, Italy, ⁴ Azienda USL di Pescara, Pescara, Italy, ⁵ Centro di Psicologia Clinica, Pescara, Italy, ⁶ Department of Health Sciences, Florence University, Florence, Italy, ⁷ Department of Human Sciences and Society, Enna "Kore" University, Enna, Italy, ⁸ Cognitive and Behavioral Therapy Institute, Padua, Italy

OPEN ACCESS

Edited by:

Elisa Pedrolì,
Istituto Auxologico Italiano (IRCCS),
Italy

Reviewed by:

Michele Settanni,
Università degli Studi di Torino, Italy
Anne Chatton,
Geneva University Hospitals (HUG),
Switzerland

*Correspondence:

Aristide Saggino
aristide.saggino@unich.it

Specialty section:

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

Received: 03 November 2017

Accepted: 26 February 2018

Published: 27 March 2018

Citation:

Saggino A, Balsamo M, Carlucci L, Cavalletti V, Sergi MR, da Fermo G, Dèttore D, Marsigli N, Petrucci I, Pizzo S and Tommasi M (2018) Psychometric Properties of the Italian Version of the Young Schema Questionnaire L-3: Preliminary Results. *Front. Psychol.* 9:312. doi: 10.3389/fpsyg.2018.00312

Schema Therapy (ST) is a well-known approach for the treatment of personality disorders. This therapy integrates different theories and techniques into an original and systematic treatment model. The Young Schema Questionnaire L-3 (YSQ-L3) is a self-report instrument, based on the ST model, designed to assess 18 Early Maladaptive Schemas (EMSs). During the last decade, it has been translated and validated in different countries and languages. This study aims to establish the psychometric properties of the Italian Version of the YSQ-L3. We enrolled two groups: a clinical ($n = 148$) and a non-clinical one ($n = 918$). We investigated the factor structure, reliability and convergent validity with anxiety and depression between clinical and non-clinical groups. The results highlighted a few relevant findings. Cronbach's alpha showed significant values for all the schemas. All of the factor models do not seem highly adequate, even if the hierarchical model has proven to be the most significant one. Furthermore, the questionnaire confirms the ability to discriminate between clinical and non-clinical groups and could represent a useful tool in the clinical practice. Limitations and future directions are discussed.

Keywords: Young Schema Questionnaire L3, reliability, validity, schema therapy, factor analysis, statistical

INTRODUCTION

Schema Therapy (ST; Young, 1994; Young et al., 2003) provided an innovative approach to psychotherapy aiming to treat patients with chronic psychological problems. Several studies showed that ST is an evidence-based treatment for personality disorders (e.g., Giesen-Bloo et al., 2006; Gude and Hoffart, 2008; Farrell et al., 2009; Nadort et al., 2009; Sempertegui et al., 2013; Bamelis et al., 2014), as well as for anxiety and depressive disorders (Balsamo, 2010, 2013; Renner et al., 2013; Malogiannis et al., 2014; Balsamo et al., 2015c; for a review, Hawke et al., 2011) and eating disorders (Waller et al., 2007). ST is currently being implemented within the mental health services of several nations, such as Denmark (Bach et al., 2015).

ST was developed as the clinical implication of Young (1994) schema theory. It is an integrative therapy, mixing elements of different approaches such as Cognitive-Behavioral Therapy, Gestalt

therapy, Attachment Theory, Object Relations Theory and emotional-focused models (Young, 1994). Influenced by these theories, Young and colleagues (Young, 1994; Young et al., 2003) developed the “Early Maladaptive Schemas” (EMSs) concept, as a broad, pervasive, trait-like, cognitive and emotional self-defeating pattern, concerning beliefs about the self, others and the future. According to the ST model, EMSs derive from early childhood noxious experiences with primary caregivers and are established by unmet core emotional needs (Young et al., 2003), as well as from peer relations during childhood and adolescence (Mash and Dozois, 2003; Renner et al., 2013). Little evidence seemed to support the association between early relational experiences and EMSs (e.g., Muris, 2006; Wright, 2007) as well as between schemas and psychopathology symptoms such as depression and anxiety in adulthood (Halvorsen et al., 2009; Hawke et al., 2011; Renner et al., 2012; Riso et al., 2017), or in youth (Van Vlierberghe et al., 2010; Balsamo et al., 2015c), even though some authors maintained that infant attachment may be an overrated predictor (e.g., Meins, 2017).

The current list of EMSs consists of 18 schemas, which have been identified in the general populations, as well as in clinical groups (Young, 1994). The 18 EMSs have been grouped into five broad categories of unmet emotional needs called “schema domains.” These broad categories are: disconnection and rejection, impaired autonomy and performance, other directedness, over-vigilance and inhibition and impaired limits (Young et al., 2003).

The *Young Schema Questionnaire* (YSQ; Young and Brown, 1994) is a self-report measure developed to assess EMSs within the ST. It is used as a clinical instrument in psychotherapy and as a research measure in developmental psychopathology studies. The first YSQ-Long Form consisted of 205 items, representing the 16 EMSs listed by the authors. After a psychometric revision of the EMSs (Schmidt et al., 1995), Young et al. (2003) 18 EMS were operationally defined and a new YSQ-Long Form was developed. This Third Edition (YSQ-L3; Young and Brown, 1994), consisted of 232 items. According to a literature review (Oei and Baranoff, 2007), although the Third Edition underwent many revisions, no consistent factor structures emerged for the YSQ-L3.

Whereas the psychometric properties of the YSQ were tested in different languages and groups (clinical and non-clinical participants), almost all of the studies employed the short form or the previous forms, which are not comparable with the YSQ L3 form. Furthermore, to the best of our knowledge, this is the first study in Italy that explores the YSQ-L3 structural validity by means of Confirmatory Factor Analysis.

In this study, we examined the reliability and structural validity of the 18 schema scales, as measured by the YSQ-L3. We specifically tested its structural validity by investigating whether the five correlated first-order factor structure, proposed by the test developers (Young et al., 2003), could be replicated in two Italian groups (clinical and non-clinical subjects) by Confirmatory Factor Analysis, as well as the one-factor model, recently found in the Italian version of the YSQ-L3 via Exploratory Factor Analysis (see Saggino et al., 2017). Since the findings resulting from current literature on the YSQ-L3’s latent factor structure were inconclusive (Oei and Baranoff, 2007),

we also tested a bi-factor model, strongly suggested by Kriston et al. (2012) for the YSQ-SF3, in which all the 18 schemas loaded each on own domain and on one global factor, called “Psychopathology.”

Finally, we tested the second-order model with five first-order factors according to Young’s model as well as a general second-order factor.

We also investigated the reliability of the YSQ-L3, as well as its convergent validity by computing associations between the YSQ-L3 and concurrent measures of anxiety and depression. In addition, we carried out a Multigroup Confirmatory Factor Analysis (MG-CFA) to test measurement invariance of the YSQ-L3 with respect to groups of subjects with and without psychological syndromes. Furthermore, false positive (FP) risk values were calculated to discriminate between non-clinical and clinical subjects.

MATERIALS AND METHODS

Participants

Participants ranged between the ages of 18 and 89 and had the capacity to complete self-administered questionnaires. This group was the same used for the Italian norms in a previous study (Saggino et al., 2017). Inclusion criteria for the clinical group were: existence of a psychiatric diagnosis and age = or > 17 years old. Exclusion criteria included ongoing psychotic symptoms, serious physical illnesses and central nervous system major disorders (e.g., Alzheimer’s disease and Parkinson’s disease). Participants were 1,112 Italian subjects: 157 clinical and 955 community participants. Forty-six were excluded from the analyses: 9 clinical and 37 non-clinical subjects were removed because they had missing values $\geq 10\%$ at EMSs. Missing values rated below 10%, were replaced with the average values of each schema.

The *clinical group* was formed by 148 outpatients of which 52 females (35.1%) and 96 males (64.9%). The group’s mean age was 37.92 ($SD = 10.43$; range = 18–64 years). The mean age for men was 38.28 years ($SD = 9.96$), and 37.25 years for women ($SD = 11.80$). No significant age by gender interaction effect was found [$F_{(1, 146)} = 0.328, p = 0.568$]. The mean years of education was 12.47 ($SD = 3.23$; range = 8–20 years): 11.89 ($SD = 3.08$) for males and 13.60 years ($SD = 3.22$) for females. A significant years of education by gender interaction effect was found [$F_{(1, 136)} = 9.17, p = 0.003$].

The *non-clinical group* was formed by 918 subjects of which 522 females (56.9%) and 396 males (43.1%). The group’s mean age was 29.85 years ($SD = 12.56$; range = 18–89 years): 31.09 years ($SD = 13.09$) for males, and 28.92 years ($SD = 12.35$) for females. There was a statistically significant difference in age between males and females [$F_{(1, 912)} = 6.58, p = 0.010$]. The mean years of education was 13.63 years ($SD = 3.36$; range 5–25 years): 13.45 years ($SD = 3.34$) for males and 13.77 years ($SD = 3.38$) for females. No statistically significant difference was found in years of education between males and females [$F_{(1, 892)} = 1.89, p = 0.169$]. All subjects were white.

The clinical group was recruited through private practice ($N = 49$; 33.1%), private psychiatric hospitals ($N = 13$;

8.8%), public psychiatric hospital ($N = 23$; 15.5%) and mental health departments ($N = 63$; 42.6%). Diagnoses were conducted according to the Diagnostic and Statistical Manual of Mental Disorders standards (DSM-IV-TR; American Psychiatric Association, 2000) by accredited psychiatrists and psychologists. The patients included in this group were diagnosed as follows: 56.8% ($N = 84$) received a diagnosis of a disorder on DSM-IV-TR Axis I, 15.5% ($N = 23$) received a diagnosis of a disorder on DSM-IV-TR Axis II and 20.9% ($N = 31$) received a comorbid diagnosis Axis I/Axis II. For 6.8% ($N = 10$) of the clinical group there was no information available about the diagnosis.

The non-clinical group was recruited through advertisements posted in established community groups (e.g., youth centers, church groups, university student associations). Study participants contributed voluntarily and anonymously. Each participant anonymously completed the questionnaire packet and gave informed consent prior to being included in the study.

131 non-clinical participants (94 females and 37 males, mean age = 22.15 and $SD = 4.37$) filled out the YSQ-L3 again after 1 month (T0); 72 non-clinical participants (57 females and 15 males, mean age = 20.86 $SD = 2.97$) filled out the YSQ-L3 again 1 month after the first retest (T1); 40 non-clinical participants (28 females and 12 males, mean age = 21.75 $SD = 3.71$) filled out the YSQ-L3 1 month after the second retest (T2).

Instruments

All participants were administered the Italian versions of the Young Schema Questionnaire Long Form, Third Edition (YSQ-L3), the Teate Depression Inventory (TDI), the State-Trait Inventory for Cognitive and Somatic Anxiety Trait Scale (STICSA). All respondents completed paper-and-pencil versions of the questionnaires in a fixed order (a socio-demographic checklist, the YSQ L3, the TDI, and the STICSA) on site at established community groups. The protocol was administered by licensed psychologists who received a brief training wherein the objectives of the research, characteristics of the instruments administered and information about common issues in the psychological assessment of adults were explained. Informed consent was obtained from every single participant included in the study, in accordance with the Ethical Standards of the Helsinki Declaration.

Young Schema Questionnaire-Long Form, Third Edition

The YSQ-L3 (Young et al., 2003) is a 232-item self-report tool developed to assess 18 EMSs. The Italian version of the questionnaire is in the Appendix of the Young et al. (2003)'s Italian book. Participants are asked to rate each statement on a 6-point Likert scale ranging from 1 ("it is completely untrue for me") to 6 ("it describes me perfectly"). Items are clustered by 18 scales and grouped into five domains, bringing together the EMSs that tend to develop together: Disconnection/Rejection (Abandonment, Mistrust/ Abuse, Emotional Deprivation, Defectiveness/Shame, Social Isolation/Alienation); Impaired Autonomy/Performance (Dependence/Incompetence, Vulnerability to Harm or Illness, Enmeshment/Undeveloped Self, Failure); Impaired Limits (Entitlement/Grandiosity,

Insufficient Self-Control/Self-Discipline); Other-Directedness (Subjugation, Self-Sacrifice, Approval-Seeking/Recognition-Seeking); and Overvigilance/Inhibition (Negativity/Pessimism, Emotional Inhibition, Unrelenting Standards/Hypercriticalness, Punitiveness). A sum or a mean score is calculated for each EMS, a higher score representing a higher endorsement of the EMS in question. YSQ has demonstrated adequate test-retest reliability and internal consistency, as well as convergent and discriminant validity (Young et al., 2003). Results attained from several YSQ studies support its validity as an EMS measure (Lee et al., 1999; Stopa et al., 2001; Hoffart et al., 2005). Cronbach's α coefficients for this current study are reported in **Table 2**. All the statistical analyses in this research were based on the mean score of each EMS.

State-Trait Inventory for Cognitive and Somatic Anxiety

The STICSA (Ree et al., 2008; Italian version see Balsamo et al., 2015a, 2016) is a 21-item measure designed to assess cognitive and somatic symptoms, both on Trait and State variations. In the trait anxiety subscale, the subject rates how often a statement is true *in general* (on a four-point Likert-type scale from "1-almost never at all" to "4-almost always"), whereas she/he rates how she/he feels *at the moment of assessment* (on a four-point Likert-type scale from "1-not at all" to "4-very much") in the state anxiety subscale. In total, the overall scale is made up of four subscales: State-Somatic (SS), Trait-Somatic (TS), State-Cognitive (SC), and Trait-Cognitive (TC).

The STICSA was developed to address the psychometric limitations of existing anxiety measures, especially as far as their extensive overlapping depression (Caci et al., 2003; Balsamo et al., 2013a; Roberts et al., 2016). The factor structure showed strong support and the total scale and subscales exhibited high internal consistencies, as well as construct consistent correlations in patients, controls, and community groups (Grös et al., 2007; Ree et al., 2008; Van Dam et al., 2013; Saggino et al., 2017). Cronbach's α coefficients for this current study are from 0.812 (State-Somatic) to 0.926 (State).

Teate Depression Inventory

The TDI (Balsamo and Saggino, 2013, 2014; Balsamo et al., 2014) is a 21-item self-report instrument designed to assess Major Depressive Disorder as specified by the latest edition of the DSM (American Psychiatric Association, 2013). It was developed via Rasch logistic analysis of responses (Rasch, 1960), within the framework of Item Response Theory, in order to overcome inherent psychometric weaknesses of existing depression measures, including the BDI-II (Balsamo and Saggino, 2007). Each item is rated on a 5-point Likert-type scale, ranging from 0 (always) to 4 (never). Growing literature suggests that the TDI has strong psychometric properties in both clinical and non-clinical groups, including an excellent Person Separation Index, no evidence of bias due to item-trait interaction, good discriminant and convergent validity and control of major response sets (Balsamo et al., 2013b, 2015a,b,c; Innamorati et al., 2013). In a recent study, three cut-off scores were recommended in terms of sensitivity, specificity

and classification accuracy to screen for varying levels (minimal, mild, moderate and severe) of depression severity in a group of patients diagnosed with Major Depressive Disorder (Balsamo and Saggino, 2014). In our groups, Cronbach's alpha was 0.943 for the clinical participants and 0.917 for the non-clinical group.

Data Analysis

Descriptive Statistics

The 18 EMSs were preliminarily submitted to analyses in order to check the normal distribution by computing means, standard deviations and indices of skewness and kurtosis. Inspection of skewness and kurtosis indices indicated that departures from normality were not severe according to West et al. (1995) with only a few exceptions. Thus, no variable transformations were deemed necessary. Statistical analyses were performed with IBM SPSS.

Reliability, and Convergent Validity Analysis of the YSQ-L3

In order to investigate the psychometric properties of the YSQ-L3, we assessed internal consistency of its scales using Cronbach's alphas indices separately for the two groups. The two-way mixed effects ICC (Intraclass-Correlation; Shrout and Fleiss, 1979; McGraw and Wong, 1996) was used to assess the 3-month test-retest stability (T0, T1, T2) of each EMS' schema on a group formed by 40 non-clinical subjects. The strong reduction of subjects is due to mortality or to the fact that many subjects refused to repeat test administration. Since the Shrout and Fleiss' (1979) ICC rules of thumb were criticized (Hopkins, 2000), we considered the following values as a general rule: ≥ 0.90 high, between 0.80 and 0.90 moderate, and ≤ 0.80 insufficient (Vincent, 1999).

The convergent validity of the YSQ-L3 schemas was investigated by computing Pearson's r correlation coefficients with well-established depression and anxiety measures (TDI and STICSA, respectively). Error α was adjusted with Bonferroni's correction. These statistical analyses were performed with IBM SPSS.

Confirmatory Factor Analyses of the YSQ-L3

Different Confirmatory Factor Analyses (CFAs) were performed separately for the clinical and non-clinical participants. Due to a slight deviation from multivariate normality all analyses were carried out using robust maximum-likelihood estimation methods. Given the heterogeneity of the results reported in literature regarding the latent factor structure of Young's EMSs (for a review, see Kriston et al., 2012), most of which referred to the different YSQ versions, we compared five alternative factor models for the Italian version of the YSQ-L3. These versions were: (1) the one-factor model (1F model), in which all 18 schemas were forced to load on a single higher order factor (Saggino et al., 2017); (2) the five correlated first-order factors model (5F-correlated model), based on Young's original theoretical model (Young et al., 2003); (3) the five not correlated first-order factors model, according to Young's model, without correlations between factors (5F-not correlated model); (4) the bi-factor model (bi-factor model), strongly suggested by Kriston

et al. (2012), in which all of the 18 EMS schemas loaded each on own domain and on one global factor, called "Psychopathology"; (5) finally, the second-order model, with the five first-order factors model, according to Young's model, and a general second-order factor.

The goodness-of-fit indices to test model validity were the Satorra-Bentler χ^2 , the ratio χ^2/df , the Comparative Fit Index (CFI), the Tucker-Lewis fit index (TLI), the Root Mean Square Error of Approximation (RMSEA) and the corresponding confidence interval (90% RMSEA). Models with an acceptable fit should have $\chi^2/df < 3$, RMSEA < 0.08 , and CFI and TLI > 0.95 (Hu and Bentler, 1999; Schermelleh-Engel et al., 2003).

Measurement Invariance of the YSQ-L3 Between Non-clinical and Clinical Groups

We performed a Multigroup Confirmatory Factor Analysis (MG-CFA) to test measurement invariance of the YSQ-L3 with respect to groups of subjects with and without psychological syndromes on a set of nested models (Meredith, 1993; Saggino et al., 2017):

1. The baseline configural invariance model (M1) in which the same factorial pattern was specified for each group, but with loadings and intercepts free to vary across groups;
2. The metric invariance model (M2), wherein loadings were constrained to be equal across groups;
3. The scalar invariance model (M3), wherein factor loadings and intercepts were constrained to be equal across groups;

There is also the model for testing strict invariance (loadings, intercepts and residual variances were constrained to be equal across groups), but strict invariance is not fundamental for the validity of the model. Model fit was assessed using the χ^2 statistical test, the χ^2/df , the RMSEA, the 90% CI of RMSEA, the SRMR, the TLI and the CFI.

Difference between CFIs (Δ CFI) of nested models was estimated for testing measurement invariance. A value of Δ CFI smaller than or equal to $|0.01|$ (in absolute values) indicates that the null hypothesis of invariance should not be rejected (Cheung and Rensvold, 2002). Tests which have scalar invariance are considered consistent tests, because unaffected by group characteristics (Meredith, 1993). If multigroup invariance is confirmed with models M2 or M3, we also tested if factor means are different across groups by setting a model wherein the factor means are zero in all groups (M4). We estimated the difference between the chi-square value of M4 and that of model M2 or M3. If the value of the difference is not significant, factor means can be considered equal across groups. CFAs and MG-CFA were performed using M-Plus 7.0 (Muthén and Muthén, 2012).

Furthermore, false positive (FP) risk values were calculated for each YSQ-L3 schema and domain. FP risks are determined by the False Positive Rate (FPR), which is the ratio between the probability of False Positives (FPs) and the sum of FPs and True Positives (TPs). Because a clinical test such as the YSQ-L3 has to discriminate between non-clinical and clinical subjects, we must estimate FPR ratio, instead of using the criterion of rejecting the null hypothesis with a first-type error probability value of 0.05, in order to attain the correct percentage of risk to make FPs using test scores (Colquhoun, 2014). All of the analyses were based on

the standardized scores for any schema and on the factor scores, for any latent domain.

All missing data were substituted by the serial mean. The work of Chen et al. (2012) showed that with a percentage of missing data below 20% there is no reduction of fit indices. The model fit decreases as the number of missing data gets larger. The authors suggest that when the percentage of missing data is higher than 30%, both the serial mean and the trend missing imputation methods offer a better model fit than the other available methods. Because the missingness in our data was always below 10%, we therefore used the Serial Mean method.

RESULTS

Descriptive Statistics of the YSQ-L3

Descriptive statistics of the 18 EMS, the TDI and the STICSA State-Trait; somatic and cognitive scales in the Italian clinical and non-clinical groups are displayed in **Table 1**.

As shown in **Table 1**, in our sample all the EMS schemas exhibited no absolute value of skewness larger than 2, neither absolute values of kurtosis larger than 7, in both groups, excepting for Defectiveness which presented a skewness corresponding to 2.030 in the non-clinical group, according to the guidelines recommended by West et al. (1995). A similar

trend of normality distribution was observed for the TDI and the STICSA scales and subscales.

Reliability, and Convergent Validity Analysis of the YSQ-L3

As shown in **Table 2**, internal consistency reliability of the 18 EMS was high (range $\alpha_{\text{clinical}} = 0.804\text{--}0.921$ and $\alpha_{\text{non-clinical}} = 0.834\text{--}0.941$).

As shown in **Table 2**, the ICC estimates were similar in value, for each Young's schema. The Emotional Deprivation schema ICC was 0.925, with 95% confidence interval from 0.878 to 0.957 [$F_{(39, 78)} = 38.148, p < 0.001$]. A moderate reliability degree was also found for Social Isolation [ICC = 0.869; 95%, CI = 0.792–0.924; $F_{(39, 78)} = 20.947, p < 0.001$], Defectiveness [ICC = 0.889; 95%, CI = 0.822–0.936; $F_{(39, 78)} = 25.087, p < 0.001$], Vulnerability [ICC = 0.856; 95% CI = 0.770–0.916; $F_{(39, 78)} = 18.785, p < 0.001$], Self-Sacrifice [ICC = 0.854; 95%, CI = 0.769–0.914; $F_{(39, 78)} = 18.596, p < 0.001$], and Unrelenting Standards [ICC = 0.802; 95% CI = 0.694–0.882; $F_{(39, 78)} = 13.185, p < 0.001$]. The remaining EMS schemas showed ICC values considered as insufficient (cut-off ≤ 0.80 ; Vincent, 1999), ranging from 0.703 (Failure to Achieve) to 0.791 (Insufficient Self-control).

Table 3 shows the correlations among the 18 EMS, measures of depression (TDI) and trait and state anxiety (STICSA,

TABLE 1 | Descriptive Statistics of the EMS, TDI, and STICSA for non-clinical ($n = 918$) and clinical sample ($n = 148$).

EMS	Non-clinical (N = 918)				Clinical (N = 148)			
	Mean*	SD*	Skewness	Kurtosis	Mean*	SD*	Skewness	Kurtosis
Emotional deprivation	2.10	0.92	1.133	1.018	2.78	1.16	0.583	-0.240
Abandonment	2.37	0.84	0.741	0.337	2.87	1.07	0.507	-0.279
Mistrust/Abused	2.28	0.75	0.787	0.653	2.59	0.94	0.472	-0.212
Social isolation	1.92	0.85	1.346	1.976	2.43	1.12	0.608	-0.532
Defectiveness	1.68	0.67	2.030	6.202	2.11	0.92	1.175	1.185
Failure to achieve	1.84	0.84	1.660	3.573	2.21	1.07	0.981	0.355
Dependence	1.80	0.70	1.315	1.933	2.31	1.03	1.052	1.052
Vulnerability	2.05	0.80	1.267	1.937	2.43	1.04	1.119	1.210
Enmeshment	2.00	0.69	0.993	1.395	2.36	0.96	0.834	0.237
Subjugation	2.13	0.79	1.056	1.421	2.56	0.96	0.684	0.521
Self-sacrifice	3.08	0.87	0.343	-0.207	3.05	0.90	0.310	-0.109
Approval-seeking	2.18	0.84	1.019	1.550	2.51	0.97	0.528	-0.202
Insufficient self-control	2.20	0.73	0.685	0.433	2.59	0.96	0.469	-0.331
Entitlement	2.55	0.78	0.489	0.305	2.69	0.89	0.218	-0.574
Unrelenting standards	2.77	0.82	0.621	0.319	2.76	0.82	0.171	-0.303
Emotional inhibition	2.24	0.91	1.025	1.127	2.64	1.09	0.516	-0.217
Negativism	2.34	0.94	0.786	0.347	2.76	1.05	0.225	-0.629
Self-punitiveness	2.55	0.80	0.483	0.323	2.63	0.85	0.078	-0.427
TDI	28.15	13.17	0.356	-0.214	35.10	16.87	0.164	-0.507
STICSA-trait	35.62	10.18	0.718	-0.044	39.53	12.83	0.839	0.486
STICSA-trait, somatic	17.36	5.20	0.939	0.507	18.32	6.92	1.340	1.540
STICSA-trait, cognitive	18.25	5.99	0.651	-0.211	21.21	7.15	0.313	-0.801
STICSA-state	32.04	10.20	0.692	1.279	35.43	13.14	1.012	2.123
STICSA-state, somatic	15.51	4.81	1.534	2.574	16.33	6.76	1.935	4.399
STICSA-state, cognitive	16.50	5.99	1.047	0.462	18.99	7.30	0.469	-0.048

*Means and Standard Deviations are based on means of EMS.

TABLE 2 | Cronbach alpha and test-retest Reliability of the 18 EMS of the YSQ-L3.

	T0		T1 [†]		T2 [†]		ICC	ICC 95% CI		F	α	
	M	SD	M	DS	M	DS		Lower	Upper		Clinical	Non-clinical
Emotional deprivation	2.13	0.91	2.09	1.01	2.30	1.09	0.925	0.878	0.957	38.148*	0.896	0.895
Abandonment	2.53	0.83	2.35	0.84	2.37	0.85	0.783	0.666	0.869	11.799*	0.894	0.918
Mistrust/Abused	2.37	0.73	2.17	0.80	2.28	0.80	0.786	0.671	0.872	12.042*	0.893	0.911
Social isolation	1.99	0.85	1.96	0.94	2.10	1.06	0.869	0.792	0.924	20.947*	0.883	0.899
Defectiveness	1.73	0.65	1.70	0.71	1.85	0.80	0.889	0.822	0.936	25.087*	0.905	0.911
Failure to achieve	1.90	0.86	1.89	0.80	1.84	0.79	0.703	0.558	0.817	8.103*	0.918	0.901
Dependence	1.86	0.64	1.71	0.65	1.70	0.63	0.768	0.646	0.860	10.930*	0.921	0.941
Vulnerability	2.12	0.80	1.98	0.87	2.05	0.83	0.856	0.770	0.916	18.785*	0.872	0.907
Enmeshment	2.08	0.70	1.87	0.68	1.91	0.72	0.781	0.663	0.868	11.676*	0.804	0.875
Subjugation	2.19	0.77	2.05	0.76	2.15	0.92	0.715	0.574	0.825	8.511*	0.841	0.858
Self-sacrifice	3.19	0.85	2.94	0.91	2.96	0.89	0.854	0.769	0.914	18.596*	0.906	0.895
Approval-seeking	2.20	0.78	2.14	0.81	2.23	0.88	0.766	0.643	0.859	10.811*	0.916	0.914
Insufficient self-control	2.25	0.69	2.11	0.71	2.15	0.63	0.791	0.678	0.875	12.344*	0.862	0.894
Entitlement	2.61	0.73	2.39	0.76	2.37	0.76	0.766	0.634	0.859	10.810*	0.823	0.834
Unrelenting standards	2.82	0.82	2.60	0.82	2.58	0.78	0.802	0.694	0.882	13.185*	0.886	0.868
Emotional inhibition	2.37	0.94	2.17	0.95	2.40	0.95	0.785	0.670	0.871	11.970*	0.840	0.876
Negativism	2.45	0.92	2.24	0.97	2.24	0.90	0.780	0.663	0.868	11.663*	0.903	0.893
Self-punitiveness	2.65	0.78	2.50	0.83	2.54	0.78	0.776	0.657	0.865	11.390*	0.881	0.874

* $p < 0.01$. $N = 40$; ICC, Intraclass Correlation Coefficient; CI, Confidence Interval. [†]Rating at 1-month distance.

with its subscales). As expected, all of the EMS in general showed an average to high correlation with the TDI and the STICSA scales both in the clinical and in the non-clinical groups.

Confirmatory Factor Analyses of the YSQ-L3

Table 4 shows the goodness-of-fit indexes of the five structural models tested both for the non-clinical and the clinical groups. Although the bi-factor model has the best fit, as far as both the non-clinical and the clinical group, it exhibits many flaws at a more detailed level.

The loadings of the Disconnection/Rejection domain are especially not significant for the Abandonment and the Defectiveness/Shame schema in the clinical group; the loadings of the Impaired Autonomy/Performance domain are not significant for all of the four schemas in the clinical group and are not significant for the Failure schema in the non-clinical group; the loadings of the Other-Directedness domain are not significant for the Subjugation and for the Approval-Seeking/Recognition-Seeking schema in the clinical group; the loading of the Impaired Limits domain on the Insufficient Self-Control/Self-Discipline schema is not significant in the clinical group; the loadings of the Overvigilance/Inhibition domain on the Emotional Inhibition, and the Unrelenting Standards/Hypercriticalness schema are not significant in the clinical participants. Not-significant loadings mean that the bi-factor model does not provide adequate measurement properties. **Table 5** shows the loadings of each schema in the five domains

and in the general factor for the bifactor model. Hierarchical (ω_h) and total omegas (ω_t) for each schema are also reported. The ratio ω_t/ω_h expresses the variance component of the general factor in each observed variable in relation to the global variance due to all latent factors (Tommasi et al., 2015).

The distributions of fit indices are affected by sample size and by the distribution of the measured characteristic in population (Yuan, 2005). Therefore, cutoffs of fit indexes cannot be considered as absolutely valid. In addition, the misfit of the models can be due to high covariance residuals instead of model misspecification. Covariance errors and model misspecification do not necessarily correspond (Hayduk et al., 2007). Therefore, not necessarily lower fit indexes indicate a misfit model. Factor loadings represent the quality of measurement of latent variables. Model with poor measurement quality (low factor loadings) can have a better fit than models with excellent measurement quality (high factor loadings). This phenomenon is called *reliability paradox* (Hancock and Mueller, 2011). On the basis of this paradox, McNeish and colleagues (McNeish et al., 2017) recommend to evaluate the validity of factor models not only on goodness of fit indexes, but also on the quality of their measures by reporting also factor loadings, because there is not a perfect correspondence between quality of measurement and fit indexes.

In the second-order model, instead, all loadings of the five domains on schemas are significant both for the non-clinical and for the clinical groups. **Figure 1** shows the path-diagram of the second-order model of the YSQ-L3.

TABLE 3 | correlations between the 18 schemas of the YSQ-L3 with TDI and STICSA for non-clinical and clinical sample.

	TDI		STICSA-trait, somatic		STICSA-trait, cognitive		STICSA-state, somatic		STICSA-state, cognitive				
	Non-clinical	Clinical	Non-clinical	Clinical	Non-clinical	Clinical	Non-clinical	Clinical	Non-clinical	Clinical			
Emotional deprivation	0.347**	0.283	0.365**	0.269	0.256**	0.238	0.399**	0.250	0.362**	0.318	0.227**	0.384**	0.261
Abandonment	0.478**	0.535**	0.571**	0.526**	0.442**	0.464**	0.588**	0.495**	0.511**	0.443**	0.362**	0.533**	0.407**
Mistrust/Abused	0.335**	0.415**	0.436**	0.437**	0.325**	0.400	0.458**	0.397	0.372**	0.375	0.258**	0.404**	0.335
Social isolation	0.449*	0.561**	0.465**	0.466**	0.307**	0.380	0.523**	0.468**	0.392**	0.417**	0.268**	0.440**	0.389
Defectiveness	0.472**	0.546**	0.504**	0.533**	0.376**	0.443**	0.530**	0.528**	0.461**	0.509**	0.365**	0.477**	0.496**
Failure to achieve	0.509**	0.585**	0.482**	0.513**	0.333**	0.415**	0.530**	0.518**	0.438**	0.486**	0.306**	0.482**	0.444**
Dependence	0.499**	0.633**	0.542**	0.577**	0.414**	0.491**	0.563**	0.559**	0.494**	0.531**	0.390**	0.510**	0.508**
Vulnerability	0.411**	0.468**	0.586**	0.617**	0.471**	0.576**	0.586**	0.548**	0.505**	0.473**	0.393**	0.527**	0.384
Enmeshment	0.263**	0.411**	0.384**	0.501**	0.328**	0.448**	0.368**	0.464**	0.350**	0.478**	0.301**	0.338**	0.458**
Subjugation	0.429**	0.513**	0.496**	0.581**	0.355**	0.526**	0.534**	0.532**	0.461**	0.561**	0.338**	0.476**	0.494**
Self-sacrifice	0.191**	0.221	0.354**	0.431**	0.278**	0.411**	0.360**	0.374	0.331**	0.428**	0.241**	0.358**	0.333
Approval-seeking	0.350**	0.467**	0.413**	0.614**	0.278**	0.564**	0.460**	0.555**	0.365**	0.550**	0.240**	0.401**	0.465**
Insufficient self-control	0.484**	0.573**	0.520**	0.527**	0.370**	0.400	0.562**	0.558**	0.462**	0.462**	0.318**	0.490**	0.479**
Entitlement	0.261**	0.444**	0.342**	0.442**	0.241**	0.389	0.372**	0.415**	0.279**	0.370	0.187**	0.294**	0.334
Unrelenting standards	0.125	0.207	0.283**	0.414**	0.212**	0.394	0.297**	0.362	0.231**	0.417**	0.157	0.430**	0.290
Emotional inhibition	0.407**	0.544**	0.537**	0.597**	0.416**	0.512**	0.552**	0.575**	0.454**	0.526**	0.339**	0.467**	0.442**
Negativism	0.525**	0.627**	0.639**	0.683**	0.468**	0.562**	0.680**	0.681**	0.554**	0.566**	0.378**	0.470**	0.511**
Self-punitiveness	0.310**	0.446**	0.421**	0.572**	0.303**	0.462**	0.452**	0.580**	0.384**	0.537**	0.288**	0.464**	0.463**

**p < 0.01. Error α was adjusted with Bonferroni's correction. TDI, Teate Depression Inventory; STICSA, State-Trait Inventory for Cognitive and Somatic Anxiety.

TABLE 4 | Goodness-of-fit indexes of the five models tested in the CFAs both for the non-clinical ($n = 918$) and the clinical sample ($n = 148$).

CFA models	χ^2	df	χ^2/df	CFI	TLI	RMSEA	90% RMSEA	SRMR	AIC	BIC
NON-CLINICAL										
1F model	1794.657	135	13.294	0.785	0.756	0.116	0.111–0.121	0.071	31151.93	31412.33
5F-correlated model	1499.681	125	11.997	0.822	0.782	0.109	0.105–0.114	0.066	30706.75	31015.37
5F-not correlated model	3816.683	136	28.064	0.523	0.463	0.172	0.167–0.176	0.389	33932.87	34188.45
Bi-factor model	935.738	107	8.745	0.893	0.846	0.092	0.086–0.097	0.044	29862.44	30257.86
Second-order model	1527.210	130	11.748	0.819	0.787	0.108	0.103–0.113	0.068	30758.38	31042.89
CLINICAL										
1F model	401.318	135	2.973	0.841	0.820	0.115	0.103–0.129	0.060	5709.360	5871.209
5F-correlated model	296.916	125	2.375	0.897	0.874	0.096	0.082–0.111	0.051	5591.381	5783.202
5F-not correlated model	815.722	136	5.998	0.594	0.543	0.184	0.172–0.196	0.457	6218.797	6377.649
Bi-factor model	244.283	107	2.283	0.918	0.883	0.093	0.078–0.109	0.037	5526.267	5772.039
Second-order model	308.082	130	2.370	0.894	0.875	0.096	0.082–0.110	0.053	5598.831	5775.666

Measurement Invariance of the YSQ-L3 Between Non-clinical and Clinical Groups

Table 6 shows the MG-CFA performed on the second-order model of the YSQ-L3. Because the second-order model has at work order loading, there is a version of the M2 model where the first-order loadings are fixed between groups (M2*) and a version where the first-order and the second-order loadings are fixed (M2**). All ΔCFI are lower than $|0.01|$, therefore the scalar invariance between the non-clinical and the clinical groups of the YSQ-L3 is confirmed. The difference between model M4 and M3 is however significant ($\Delta \chi^2 = 45.824$, $df = 5$, $p < 0.001$). The means of the five domains of the YSQ-L3 are therefore significantly different between the non-clinical and the clinical group. All of the means of the five domains are higher in the clinical than in the non-clinical group.

We therefore calculated the FPR for each schema and for each domain. On these calculations we estimated the percentage of risk in making FPs, multiplying the FPR ratio by 100, for both of the scores attained at the level of YSQ-L3 schemas and on factor scores of the five YSQ-L3 domains. Before estimating the FP risk for each YSQ-L3 schema, we transformed the raw scores of each schema in standardized scores. We estimated different distribution of standardized scores for the non-clinical and the clinical group. The cutoff values for the 0.05 and the 0.025 probability of FPs in the non-clinical group (first-type error) were used to estimate the probability values of TPs in the clinical group. We calculated the factor scores of the five domains to calculate the FPR for each domain. We estimated different distributions of standardized scores for the non-clinical and the clinical groups. The cutoff values for the 0.05 and the 0.025 probability of FPs in the non-clinical group (first-type error) were used to estimate the probability values of TPs in the clinical group. Table 7 shows the FP risk values for each YSQ-L3 schema and for each YSQ-L3 domain. The average FP risk value is 40.6 and 45.0% for the YSQ-L3 schemas, for the 5 and the 2.5% first-type error, respectively, while the average FP risk value for the YSQ-L3 domains is 24.2 and 18.2%, for the 0.05 and the 0.025 first-type error, respectively. FP risk is therefore lower when the factor scores for the five YSQ-L3

domains are used to discriminate between non-clinical and clinical subjects. According to Colquhoun (2014), the usual cutoffs for significance testing (0.05, 0.01 or 0.001) are somewhat misleading, because based on the assumption that if there are no significant differences between clinical and non-clinical subjects (null effect), therefore there is only a 5, 1 or 0.1% probability to judge an individual as a clinical subject while he is perfectly normal. However, this approach does not consider the power of the test or, in other words, the capacity of the psychological test to discriminate between clinical and non-clinical subjects. The test power is the probability to correct recognize the presence of disease in non-clinical subjects (true positives). If test power is not estimated, the correct identification of FPs is underestimated. Therefore, Colquhoun (2014) suggests to use the FPR instead of the usual null hypothesis significance test to determine its capacity to discriminate clinical from non-clinical subjects.

DISCUSSION AND CONCLUSION

The YSQ-L3 (Young and Brown, 1994) is a self-report instrument, developed after a psychometric refinement of the previous version aimed at assessing the 18 EMS according to the ST theoretical framework. Its latent factor structure has not been consistently replicated (for a review, see Oei and Baranoff, 2007). In fact, almost all of the studies on the YSQ psychometric structure scrutinized the previous form (YSQ-L2) or the short form (YSQ-S3) and not the actual long form (YSQ-L3).

Knowledge of its factor structure could be useful both for researchers and for clinicians during assessment and treatment. The current study investigated the factor structure of the Italian YSQ-L3, its reliability, convergent validity with state/trait anxiety and depression measures, and measurement invariance across a large community and clinical groups.

CFAs analyses were conducted separately for the community and for the clinical groups, testing five different models: a single-factor model, a five correlated first-order factor model, a five uncorrelated first-order factor model, a bi-factor model and, finally, a second-order model, with the five first-order factors, according to Young's model, and a general second-order factor.

TABLE 5 | Loadings on the first-order factors (λ_f) and on the general factor (λ_g) and corresponding significance (p -values).

Young-L3 domains	Young-L3 schemas	λ_f	λ_f p -value	λ_g	λ_g p -value	ω_h	ω_t	ω_h/ω_t
NON-CLINICAL (N = 918)								
Disconnection/Rejection	Emotional deprivation	0.49	<0.01	0.38	<0.01	0.15	0.38	0.38
	Abandonment	0.54	<0.01	0.55	<0.01	0.30	0.59	0.51
	Mistrust/Abused	0.75	<0.01	0.36	0.04	0.13	0.69	0.18
	Social isolation	0.46	<0.01	0.63	<0.01	0.40	0.60	0.66
Impaired autonomy/Performance	Defectiveness	0.37	0.03	0.78	<0.01	0.61	0.75	0.81
	Failure to achieve	0.20	0.33	0.80	<0.01	0.63	0.67	0.94
	Dependence	0.38	0.03	0.79	<0.01	0.62	0.76	0.81
	Vulnerability	0.61	<0.01	0.50	<0.01	0.25	0.63	0.40
Impaired autonomy/Performance	Enmeshment	0.55	<0.01	0.36	<0.01	0.13	0.43	0.30
	Subjugation	0.50	<0.01	0.70	<0.01	0.49	0.73	0.66
	Self-sacrifice	0.59	<0.01	0.21	0.13	0.04	0.39	0.12
Impaired limits	Approval-seeking	0.55	<0.01	0.51	<0.01	0.26	0.56	0.45
	Insufficient self-control	0.55	<0.01	0.60	<0.01	0.36	0.66	0.55
Overvigilance/Inhibition	Entitlement	0.90	<0.01	0.20	0.21	0.04	0.85	0.05
	Unrelenting standards	0.77	<0.01	0.03	0.87	<0.01	0.60	<0.01
Overvigilance/Inhibition	Emotional inhibition	0.58	<0.01	0.50	<0.01	0.25	0.59	0.42
	Negativism	0.58	<0.01	0.56	<0.01	0.32	0.65	0.48
	Self-punitiveness	0.67	<0.01	0.29	0.08	0.08	0.53	0.16
CLINICAL (N = 148)								
Disconnection/Rejection	Emotional deprivation	0.47	0.01	0.43	0.04	0.18	0.40	0.45
	Abandonment	0.54	0.24	0.70	0.04	0.48	0.77	0.63
	Mistrust/Abused	0.70	<0.01	0.59	0.08	0.34	0.83	0.41
	Social Isolation	0.46	0.05	0.69	<0.01	0.48	0.69	0.70
	Defectiveness	0.33	0.32	0.81	<0.01	0.65	0.76	0.86
Impaired autonomy/Performance	Failure to achieve	0.12	0.82	0.82	<0.01	0.67	0.68	0.98
	Dependence	0.06	0.94	0.86	<0.01	0.74	0.74	0.99
	Vulnerability	0.42	0.42	0.64	0.03	0.41	0.59	0.70
	Enmeshment	0.50	0.32	0.60	0.06	0.36	0.61	0.59
Impaired autonomy/Performance	Subjugation	0.48	0.21	0.72	<0.01	0.52	0.75	0.69
	Self-sacrifice	0.68	0.01	0.32	0.48	0.10	0.57	0.18
	Approval-seeking	0.47	0.30	0.64	0.04	0.40	0.63	0.65
Impaired Limits	Insufficient self-control	0.54	0.09	0.72	<0.01	0.51	0.80	0.64
	Entitlement	0.68	0.01	0.58	0.06	0.33	0.80	0.42
Overvigilance/Inhibition	Unrelenting standards	0.65	<0.01	0.25	0.49	0.06	0.48	0.13
	Emotional inhibition	0.50	0.11	0.67	<0.01	0.45	0.70	0.65
	Negativism	0.50	0.25	0.72	0.01	0.52	0.76	0.68
	Self-punitiveness	0.67	0.01	0.52	0.13	0.27	0.71	0.37

Not significant loadings ($p > 0.05$) are reported on bold types. Hierarchical omega (ω_h), total omega (ω_t) and ratio between hierarchical and total omega (ω_h/ω_t) are reported.

Although the bi-factor model showed the best fit, both in the clinical group and the community group, some loadings of the five domains did not appear to be significant for their corresponding schemas, as posited by the original factor structure model, thus suggesting an inadequate fit. In the second-order model, instead, all loadings of the five domains on their schemas seemed to be significant both for the community and for the clinical groups. The second-order model was therefore preferred as it showed more adequate measurement properties than the bi-factor model for both of the groups. The original model proposed by Young et al. (2003) was therefore not confirmed in the current study.

Measurement invariance of the YSQ-L3 between community and clinical groups was subsequently tested for the second-order model. Results suggested that all ΔCFI were lower than [0.01], thus supporting the scalar invariance between the community and the clinical groups. Since models M4 and M3 resulted significantly different, the means of the five domains of the YSQ-L3 appeared significantly different across the community and the clinical groups. All of the means of the five domains were higher in the clinical group than in the community group. The YSQ-L3 therefore appeared to be able to discriminate between the community and the clinical groups.

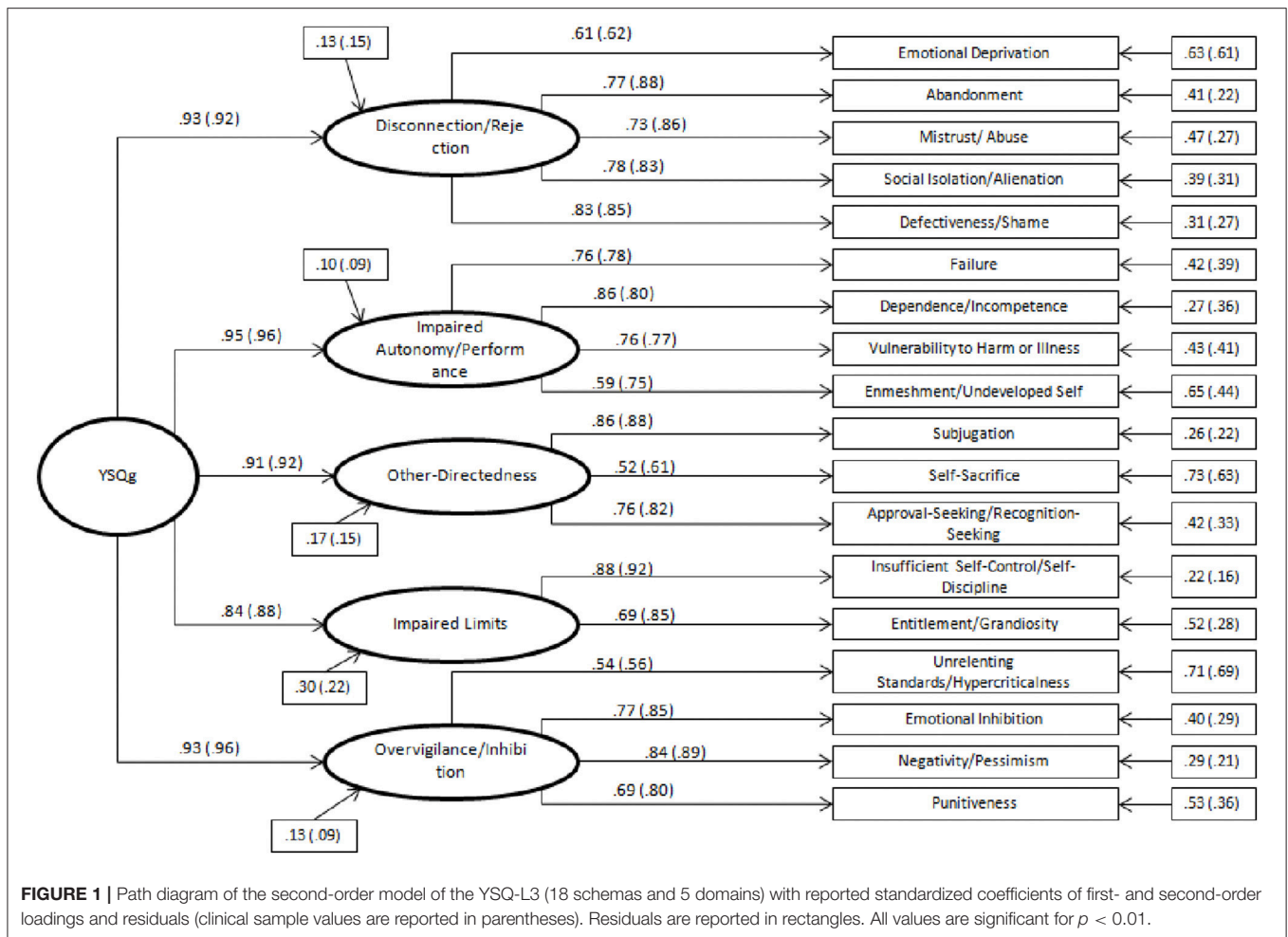


TABLE 6 | MG-CFA for testing measurement invariance of the YSQ-L3 between the non-clinical ($n = 918$) and the clinical ($n = 148$) sample.

Models for measurement invariance	χ^2	df	χ^2/df	RMSEA	90% RMSEA	SRMR	TLI	CFI	ΔCFI
M1	1906.673	26.000	7.333	0.109	0.104–0.114	0.066	0.800	0.830	
M2*	1911.471	273.000	7.002	0.106	0.102–0.111	0.068	0.810	0.831	0.001
M2**	1922.533	277.000	6.941	0.106	0.101–0.110	0.069	0.812	0.830	-0.001
M3	200.130	29.000	6.897	0.105	0.101–0.110	0.071	0.813	0.823	-0.007
M4	2045.954	295.000	6.935	0.106	0.101–0.110	0.089	0.812	0.819	

n.b.: M1, model for configural invariance; M2* model for metric invariance (fixed first-order loadings); M2**, model for metric invariance (fixed first- and second-order loadings); M3 model for scalar invariance; M4, M3 with fixed means of YSQ-L3 domains for each group. ΔCFI s lower than |0.01| are in bold type.

False positive risks indeed appeared lower when the factor scores of the five YSQ-L3 domains were used to discriminate between community and clinical individuals than when all of the 18 EMS were used. This result supported the ST model (Young et al., 2003), which posited that domains constructs are associated with psychopathology.

These results supply proof of the YSQ-L3 discriminant power and, consequently, of its validity. The average to high correlations between both the TDI and the STICSA supply additional proof of the YSQ-L3 capacity to measure psychopathology.

The ICC reliability estimates were in general insufficient or moderate and this could represent a problem for the YSQL-3.

This study bears various strengths. Firstly, it is one of the rare studies available about the YSQ-L3. YSQ-L3 is the most important version of the Young Schema Questionnaire and the most useful one as far as giving psychotherapists indications about patients' schemas. Secondly, at the best of our knowledge, this study is the most comprehensive one available as far as the validity of the Italian version of the YSQ-L3 is concerned. Third, participants were both community and clinical subjects.

TABLE 7 | False Positive Rate (FPR) risk values (in percentage values) for each YSQ-L3 schema and domain.

	First-type errors	
	0.05	0.025
YSQ-L3 SCHEMAS		
Emotional deprivation	33.1	38.2
Abandonment	30.3	31.6
Mistrust/ Abuse	38.1	42.5
Social isolation/Alienation	38.1	42.5
Defectiveness/Shame	33.1	34.6
Failure	36.3	48.1
Dependence/Incompetence	40.2	34.6
Vulnerability to harm or illness	45.1	31.6
Enmeshment/Undeveloped self	36.3	42.5
Subjugation	40.2	42.5
Self-sacrifice	55.2	55.2
Approval-seeking/recognition-seeking	30.3	48.1
Insufficient self-control/self-discipline	33.1	31.6
Entitlement/Grandiosity	51.4	38.2
Unrelenting standards/Hypercriticalness	51.4	78.7
Emotional inhibition	48.0	48.1
Negativity/Pessimism	38.1	42.5
Punitiveness	51.4	78.7
Average FPR	40.6	45.0
YSQ-L3 DOMAINS		
Disconnection/Rejection	21.6	16.3
Impaired autonomy/Performance	28.1	20.9
Other-directedness	29.1	25.2
Impaired limits	24.4	17.0
Overvigilance/Inhibition	17.9	11.7
Average FPR	24.2	18.2

An additional strength is supplied by the specific analyses that it reports for the first time, for example concerning the FPR risk values for each YSQ-L3 schema and domain.

Some limitations of the study should be highlighted. Firstly, the study uses a clinical group with different psychiatric diagnoses. An additional potential bias is that the clinical group included also individuals with comorbid personality disorders and individuals without them. Future research should thus investigate measurement invariance of the YSQ-L3 across different types of psychiatric disorders, such as clinical groups with only personality disorders and groups with only anxiety or depressive disorders. Examining whether the YSQ-L3 can discriminate between individuals with different personality disorders, eating disorders (Innamorati et al., 2015) or clusters of personality disorders could also be interesting.

REFERENCES

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

Another limitation of this study concerns the lack of measures of other constructs related to EMS in the analysis of convergent validity, such as personality traits, attachment styles or functional/dysfunctional personal values (i.e., Balsamo et al., 2013; Picconi et al., 2018). Future studies should also investigate the responsiveness of the questionnaire in participants with psychiatric disorders after CBT or ST.

A further limitation concerns the numerous missing data. We tried to solve this problem in the best possible way. Anyway, particularly for this reason, a replication of the present study is welcomed.

In conclusion, the current study expanded previous knowledge beyond the inconclusive evidence about factor structure of the YSQ-L3, indicating a second-order model for the Italian version, and showing that it can be a valid and reliable instrument of measure than can be used in clinical practice and research.

ETHICS STATEMENT

In accordance with the Declaration of Helsinki, all participants provided written informed consent. Concerning ethics approval, the data collection process does not harm participants neither physically nor mentally.

AUTHOR CONTRIBUTIONS

AS designed the study, assisted with data analyses, wrote part of the paper, and edited the final manuscript. MB assisted with the design of the study and data analyses, and wrote the most part of the paper. LC contributed in the analysis of the data and wrote part of the paper. VC recruited part of the sample. MS collaborated in editing the final manuscript and recruited part of the sample. GdF recruited part of the sample. DD recruited part of the sample and contributed in the analysis of the data. NM recruited part of the sample. IP recruited part of the sample. SP recruited part of the sample. MT assisted with the data analyses and collaborated in writing the manuscript.

ACKNOWLEDGMENTS

Thanks to the following colleagues for their collaboration in the administration of the questionnaires pertaining to this research: Ines D'Ambrosio, Grazia Ferramosca, Grazia Ferrara, Annalisa Gatta, Elisabetta Righini, Daniela Romano. We thank also the following centers that have given their availability to patient recruitment for this study: Centro di Salute Mentale Tocco da Casauria, Centro Diurno L'Airone Pescara, Clinica De Cesaris, Comunità il Castello, Comunità Passaggi, Comunità Soggiorno Proposta.

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

Bach, B., Lee, C., Mortensen, E. L., and Simonsen, E. (2015). How do DSM-5 personality traits align with schema therapy constructs? *J. Pers. Disord.* 30, 502–529. doi: 10.1521/pedi_2015_29_212

- Balsamo, M. (2010). Anger and depression: evidence of a possible mediating role for rumination. *Psychol. Rep.* 106, 3–12. doi: 10.2466/PRO.106.1.3-12
- Balsamo, M. (2013). Personality and depression: evidence of a possible mediating role for anger trait in the relationship between cooperativeness and depression. *Compr. Psychiatry* 54, 46–52. doi: 10.1016/j.comppsy.2012.05.007
- Balsamo, M., Carlucci, L., Murdock, K. K., Sergi, M. R., and Saggino, A. (2015c). The mediating role of early maladaptive schemas in the relation between co-rumination and depression in youths. *PLoS ONE* 10:e0140177. doi: 10.1371/journal.pone.0140177
- Balsamo, M., Carlucci, L., Sergi, M. R., Romanelli, R., D'Ambrosio, I., Fairfield, B., et al. (2016). A new measure for trait and state anxiety: the State Trait Inventory of Cognitive and Somatic Anxiety (STICSA). Standardization in an Italian population. *Psicoterapia Cogn. Comportamentale* 22, 229–232.
- Balsamo, M., Giampaglia, G., and Saggino, A. (2014). Building a new Rasch-based self-report inventory of depression. *Neuropsychiatr. Dis. Treat.* 10, 153–165. doi: 10.2147/NDT.S53425
- Balsamo, M., Imperatori, C., Sergi, M. R., Belvederi Murri, M., Continisio, M., Tamburello, A., et al. (2013b). Cognitive vulnerabilities and depression in young adults: an ROC curves analysis. *Depress. Res. Treat.* 2013:407602. doi: 10.1155/2013/407602
- Balsamo, M., Innamorati, M., Van Dam, N. T., Carlucci, L., and Saggino, A. (2015a). Measuring anxiety in the elderly: psychometric properties of the state trait inventory of cognitive and somatic anxiety (STICSA) in an elderly Italian sample. *Int. Psychogeriatr.* 27, 999–1008. doi: 10.1017/S1041610214002634
- Balsamo, M., Lauriola, M., and Saggino, A. (2013). Work values and college major choice. *Learn. Individ. Diff.* 24, 110–116. doi: 10.1016/j.lindif.2012.12.022
- Balsamo, M., Macchia, A., Carlucci, L., Picconi, L., Tommasi, M., Gilbert, P., et al. (2015b). Measurement of internal shame: an inside view. *J. Pers. Assess.* 97, 81–89. doi: 10.1080/00223891.2014.947650
- Balsamo, M., Romanelli, R., Innamorati, M., Ciccarese, G., Carlucci, L., and Saggino, A. (2013a). The state-trait anxiety inventory: shadows and lights on its construct validity. *J. Psychopathol. Behav. Assess.* 35, 475–486. doi: 10.1007/s10862-013-9354-5
- Balsamo, M., and Saggino, A. (2007). Test per l'assessment della depressione nel contesto italiano: un'analisi critica [Tests for depression assessment in Italian context: a critical review]. *Psicoterapia Cogn. Comp.* 13, 167–199.
- Balsamo, M., and Saggino, A. (2013). *Il Teate Depression Inventory- Manuale [Teate Depression Inventory, Manual]*. Florence: Hoegrefe.
- Balsamo, M., and Saggino, A. (2014). Determining a diagnostic cut-off on the teate depression inventory. *Neuropsychiatr. Dis. Treat.* 10, 987–995. doi: 10.2147/NDT.S55706
- Bamelis, L. L., Evers, S. M., Spinhoven, P., and Arntz, A. (2014). Results of a multicenter randomized controlled trial of the clinical effectiveness of schema therapy for personality disorders. *Am. J. Psychiatry* 171, 305–322. doi: 10.1176/appi.ajp.2013.12040518
- Caci, H., Baylé, F. J., Dossios, C., Robert, P., and Boyer, P. (2003). The Spielberger trait anxiety inventory measures more than anxiety. *Eur. Psychiatry* 18, 394–400. doi: 10.1016/j.eurpsy.2003.05.003
- Chen, S. F., Wang, S., and Chen, C. Y. (2012). A simulation study using EFA and CFA programs based the impact of missing data on test dimensionality. *Expert Syst. Appl.* 39, 4026–4031. doi: 10.1016/j.eswa.2011.09.085
- Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Eq. Model.* 9, 233–255. doi: 10.1207/S15328007SEM0902_5
- Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *R. Soc. Open Sci.* 1:140216. doi: 10.1098/rsos.140216
- Farrell, J. M., Shaw, I. A., and Webber, M. A. (2009). A schema-focused approach to group psychotherapy for outpatients with borderline personality disorder: a randomized controlled trial. *J. Behav. Ther. Exp. Psychiatry* 40, 317–328. doi: 10.1016/j.jbtep.2009.01.002
- Giesen-Bloo, J., Van Dyck, R., Spinhoven, P., Van Tilburg, W., Dirksen, C., Van Asselt, T., et al. (2006). Outpatient psychotherapy for borderline personality disorder: randomized trial of schema-focused therapy vs. transference-focused psychotherapy. *Arch. Gen. Psychiatry* 63, 649–658. doi: 10.1001/archpsyc.63.6.649
- Grös, D. F., Antony, M. M., Simms, L. J., and McCabe, R. E. (2007). Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): comparison to the State-Trait Anxiety Inventory (STAI). *Psychol. Assess.* 19, 369–381. doi: 10.1037/1040-3590.19.4.369
- Gude, T., and Hoffart, A. (2008). Change in interpersonal problems after cognitive agoraphobia and schema-focused therapy versus psychodynamic treatment as usual of inpatients with agoraphobia and Cluster C personality disorders. *Scand. J. Psychol.* 49, 195–199. doi: 10.1111/j.1467-9450.2008.00629.x
- Halvorsen, M., Wang, C. E., Richter, J., Myrland, I., Pedersen, S. K., Eisemann, M., et al. (2009). Early maladaptive schemas, temperament and character traits in clinically depressed and previously depressed subjects. *Clin. Psychol. Psychother.* 16, 394–407. doi: 10.1002/cpp.618
- Hancock, G. R., and Mueller, R. O. (2011). The reliability paradox in assessing structural relations within covariance structure models. *Educ. Psychol. Meas.* 71, 306–324. doi: 10.1177/0013164410384856
- Hawke, L. D., Provencher, M. D., and Arntz, A. (2011). Early maladaptive schemas in the risk for bipolar spectrum disorders. *J. Affect. Disord.* 133, 428–436. doi: 10.1016/j.jad.2011.04.040
- Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., and Boulianne, S. (2007). Testing! testing! one, two, three—testing the theory in structural equation models!. *Pers. Individ. Dif.* 42, 841–850. doi: 10.1016/j.paid.2006.10.001
- Hoffart, A., Sexton, H., Hedley, L. M., Wang, C. E., Holthe, H., Haugum, J. A., et al. (2005). The structure of maladaptive schemas: a confirmatory factor analysis and a psychometric evaluation of factor-derived scales. *Cogn. Ther. Res.* 29, 627–644. doi: 10.1007/s10608-005-9630-0
- Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Med.* 30, 1–15. doi: 10.2165/00007256-200030010-00001
- Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equation Model.* 6, 1–55. doi: 10.1080/10705519909540118
- Innamorati, M., Imperatori, C., Meule, A., Lamis, D. A., Contardi, A., Balsamo, M., et al. (2015). Psychometric properties of the Italian Food Cravings Questionnaire-Trait-reduced (FCQ-T-r). *Eating and Weight Dis.* 20, 129–135. doi: 10.1007/s40519-014-0143-2
- Innamorati, M., Tamburello, S., Contardi, A., Imperatori, C., Tamburello, A., Saggino, A., et al. (2013). Psychometric properties of the attitudes toward self-revised in Italian young adults. *Depress. Res. Treatment* 2013:209216. doi: 10.1155/2013/209216
- Kriston, L., Schäfer, J., von Wolff, A., Härter, M., and Hölzel, L. P. (2012). The latent factor structure of Young's early maladaptive schemas: are schemas organized into domains? *J. Clin. Psychol.* 68, 684–698. doi: 10.1002/jclp.21846
- Lee, C. W., Taylor, G., and Dunn, J. (1999). Factor structure of the schema questionnaire in a large clinical sample. *Cogn. Ther. Res.* 23, 441–451.
- Malogiannis, I. A., Arntz, A., Spyropoulou, A., Tsartsara, E., Aggeli, A., Karveli, S., et al. (2014). Schema therapy for patients with chronic depression: a single case series study. *J. Behav. Ther. Exp. Psychiatry* 45, 319–329. doi: 10.1016/j.jbtep.2014.02.003
- Mash, E. J., and Dozois, D. J. A. (2003). "Child psychopathology: a developmental-systems perspective," in *Child Psychopathology, 2nd Edn.* eds E. J. Mash and R. A. Barkley (New York, NY: The Guilford Press), 3–74.
- McGraw, K. O., and Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* 1, 30–46.
- McNeish, D., An, J., and Hancock, G. R. (2017). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *J. Pers. Assess.* 1–10. doi: 10.1080/00223891.2017.1281286
- Meins, E. (2017). The predictive power of attachment. *Psychologist* 30, 21–24.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58, 525–543. doi: 10.1007/BF02294825
- Muris, P. (2006). Maladaptive schemas in nonclinical adolescents: relations to perceived parental rearing behaviours, big five personality factors and psychopathological symptoms. *Clin. Psychol. Psychother.* 13, 405–413. doi: 10.1002/cpp.506
- Muthén, L. K., and Muthén, B. O. (2012). *Mplus Statistical Modeling Software: Release 7.0*. Los Angeles, CA: Muthén & Muthén.
- Nadort, M., Arntz, A., Smit, J. H., Giesen-Bloo, J., Eikelenboom, M., Spinhoven, P., et al. (2009). Implementation of outpatient schema therapy for borderline personality disorder with versus without crisis support of the therapist outside office hours: a randomized trial. *Behav. Res. Ther.* 47, 961–973. doi: 10.1016/j.brat.2009.07.013

- Oei, T. P. S., and Baranoff, J. (2007). Young schema questionnaire: review of psychometric and measurement issues. *Aust. J. Psychol.* 59, 78–86. doi: 10.1080/00049530601148397
- Picconi, L., Jackson, C. J., Balsamo, M., Tommasi, M., and Saggino, A. (2018). Factor structure and measurement invariance across groups of the Italian Eysenck Personality Questionnaire-Short form (EPP-S). *Pers. Individ. Dif.* 123, 76–80. doi: 10.1016/j.paid.2017.11.013
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Achievement Tests*. Copenhagen: Danish Institute for Educational Research.
- Ree, M. J., French, D., MacLeod, C., and Locke, V. (2008). Distinguishing cognitive and somatic dimensions of state and trait anxiety: development and validation of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA). *Behav. Cogn. Psychother.* 36, 313–332. doi: 10.1017/S135246580004232
- Renner, F., Arntz, A., Leeuw, I., and Huibers, M. (2013). Treatment for chronic depression using schema therapy. *Clin. Psychol. Sci. Pract.* 20, 166–180. doi: 10.1111/cpsp.12032
- Renner, F., Lobbestael, J., Peeters, F., Arntz, A., and Huibers, M. (2012). Early maladaptive schemas in depressed patients: stability and relation with depressive symptoms over the course of treatment. *J. Affect. Disord.* 136, 581–590. doi: 10.1016/j.jad.2011.10.027
- Riso, L. P., du Toit, P. L., Stein, D. J., and Young, J. E. (2017). *Cognitive Schemas and Core Beliefs in Psychological Problems: A Scientist-Practitioner Guide*. Washington, DC: American Psychological Association.
- Roberts, K. E., Hart, T. A., and Eastwood, J. D. (2016). Factor structure and validity of the State-Trait Inventory for Cognitive and Somatic Anxiety. *Psychol. Assess.* 28, 134–146. doi: 10.1037/pas0000155
- Saggino, A., Balsamo, M., Carlucci, L., Sergi, M. R., Da Fermo, G., Dèttore, D., et al. (2017). Analysis of the factor structure of the Italian version of the Young Schema Questionnaire L-3 in an Italian clinical and nonclinical sample: preliminary results of a multicenter study. *G. Ital. Psicol.* 44, 445–466. doi: 10.1421/87349
- Saggino, A., Carlucci, L., Sergi, M. R., D'Ambrosio, I., Fairfield, B., Cera, N., et al. (2017). A validation study of the psychometric properties of the other as shamer scale—2. *SAGE* 7, 1–10. doi: 10.1177/2158244017704241
- Schermerle-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. Online* 8, 23–74.
- Schmidt, N. B., Joiner, T. E., Young, J. E., and Telch, M. J. (1995). The Schema questionnaire: investigation of psychometric properties and the hierarchical structure of maladaptive schemas. *Cognit. Ther. Res.* 19, 295–321. doi: 10.1007/BF02230402
- Sempertegui, G. A., Karreman, A., Arntz, A., and Bekker, M. H. (2013). Schema therapy for borderline personality disorder: a comprehensive review of its empirical foundations, effectiveness and implementation possibilities. *Clin. Psychol. Rev.* 33, 426–447. doi: 10.1016/j.cpr.2012.11.006
- Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428.
- Stopa, L., Thorne, P., Waters, A., and Preston, J. (2001). Are the short and long forms of the Young Schema Questionnaire comparable and how well does each version predict psychopathology scores? *J. Cogn. Psychother.* 15, 253–272.
- Tommasi, M., Pezzuti, L., Colom, R., Abad, F. J., Saggino, A., and Orsini, A. (2015). Increased educational level is related with higher IQ scores but lower g-variance: evidence from the standardization of the WAIS-R for Italy. *Intelligence* 50, 68–74. doi: 10.1016/j.intell.2015.02.005
- Van Dam, N. T., Gros, D. F., Earleywine, M., and Antony, M. M. (2013). Establishing a trait anxiety threshold that signals likelihood of anxiety disorders. *Anxiety Stress Coping* 26, 70–86. doi: 10.1080/10615806.2011.631525
- Van Vlierbergh, L., Braet, C., Bosmans, G., Rosseel, Y., and Bögels, S. (2010). Maladaptive schemas and psychopathology in adolescence: on the utility of Young's schema theory in youth. *Cogn. Ther. Res.* 34, 316–332. doi: 10.1007/s10608-009-9283-5
- Vincent, W. J. (1999). *Statistics in Kinesiology*. Champaign, IL: Human Kinetics.
- Waller, G., Kennerley, H., and Ohanian, V. (2007). "Schema-focused cognitive-behavioral therapy for eating disorders," in *Cognitive Schemas and Core Beliefs in Psychological Problems. A Scientist-Practitioner Guide*, eds L. P. Riso, P. L. du Toit, D. J. Stein, and J. E. Young (Washington, DC: American Psychological Association), 139–175.
- West, S. G., Finch, J. F., and Curran, P. J. (1995). *Structural Equation Models with Non normal Variables: Problems and Remedies. Structural Equation Modeling: Concepts, Issues, and Applications*. Thousand Oaks, CA: Sage Publications.
- Wright, M. O. (2007). The long-term impact of emotional abuse in childhood: Identifying mediating and moderating processes. *J. Emotional Abuse* 7, 1–8. doi: 10.1300/J135v07n02_01
- Young, J. E. (1994). *Cognitive Therapy for Personality Disorders: A Schema-Focused Approach*. Rev. Sarasota, FL: Professional Resource Press/Professional Resource Exchange.
- Young, J. E., and Brown, G. (1994). "Young Schema-Questionnaire," in *Cognitive Therapy for Personality Disorders: A Schema-Focused Approach, 2nd Edn.*, ed J. E. Young (Sarasota, FL: Professional Resource Press), 63–76.
- Young, J. E., Klosko, J. S., and Weishaar, M. E. (2003). *Schema Therapy. A Practitioner's Guide*. New York, NY: The Guilford Press.
- Yuan, K. H. (2005). Fit indices versus test statistics. *Multivariate Behav. Res.* 40, 115–148. doi: 10.1207/s15327906mbr4001_5

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

Copyright © 2018 Saggino, Balsamo, Carlucci, Cavalletti, Sergi, da Fermo, Dèttore, Marsigli, Petruccelli, Pizzo and Tommasi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Measuring Intimate Partner Violence and Traumatic Affect: Development of VITA, an Italian Scale

Gina Troisi*

Department of Humanities, University of Naples Federico II, Naples, Italy

OPEN ACCESS

Edited by:

Marco Innamorati,
Università Europea di Roma, Italy

Reviewed by:

Nicoletta Cera,
Universidade do Porto, Portugal
Elisa Pedroli,
Istituto Auxologico Italiano (IRCCS), Italy
Rosa Scardigno,
Università degli Studi di Bari Aldo Moro, Italy

*Correspondence:

Gina Troisi
gina.troisi2@unina.it

Specialty section:

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

Received: 17 April 2018

Accepted: 04 July 2018

Published: 26 July 2018

Citation:

Troisi G (2018) Measuring Intimate Partner Violence and Traumatic Affect: Development of VITA, an Italian Scale. *Front. Psychol.* 9:1282. doi: 10.3389/fpsyg.2018.01282

In a global context where the percentage of women who are victim of violence is still high (World Health Organization, 2013), intimate partner violence (IPV) can be considered the most widespread form of violence against women: in such cases violent attacks are perpetuated or threatened by a partner or ex-partner within an intimate relationship, which makes its recognition more difficult. IPV requires specific tools and, although the literature has highlighted the specific role played by some emotions (such as shame, guilt, and fear) that keep women experiencing this violence in a state of passivity and confusion, to date too little attention has been given to the construction of sound instruments able to detect post-traumatic affectivity. Such instruments could facilitate women who have suffered from IPV in recognizing it and in making the responses of women's health services more sensitive and structured. This study illustrates a sequential item development process to elaborate a new self-report instrument (VITA Scale: Intimate Violence and Traumatic Affects Scale) for assessing the intensity of post-traumatic affect derived from IPV. Within a psychodynamic perspective, the scale is characterized by four affects: fear, as a state of alarm elicited by the avoidance of the danger; terror, as a paralyzing state that hinders an active process of reaction; shame as a strong exposure to the other that disarms the individual and the guilt as a defensive dimension aiming at the restoring of the link with the abusive partner. Through specific methodological steps, a 28-item set was selected and administered to a sample of 302 Italian women who declared themselves as having suffered from IPV. Explorative and confirmatory factor analysis, as well as correlations with well-established concurrent tools were computed in order to investigate its psychometric property. A factorial structure composed of four factors, consistent with theoretical scales and a good internal consistency (Cronbach's alphas from 0.80 to 0.90) emerged. The VITA Scale could be a useful tool for clinicians and researchers to investigate the intensity of the affective state of the woman suffered from IPV. It could be useful to better address the clinical practice and therapeutic intervention planning.

Keywords: intimate partner violence, shame, guilt, terror, fear, psychodynamic perspective, women's health

INTRODUCTION

Intimate Partner Violence

With an estimated global prevalence of 30% (World Health Organization, 2013), intimate partner violence (IPV), can be considered the most widespread form of violence against women.

According to the definition of the American Psychological Association and Presidential Task Force on Violence and the Family (1996), IPV, is the physical, sexual, psychological, economical or stalking abuse, both concrete and menaced, perpetuated by current or ex-partners. In the European Union Member States the 22% of women have suffered from physical and/or sexual violence by partners since the age of 15, with a prevalence across countries ranging from 13 to 32% (European Union Agency for Fundamental Rights, 2014). In Italy, according to a national survey by the National Institute of Statistics (ISTAT, 2015), two million eight hundred thousand women between 16 and 70 years have experienced at least one episode of sexual or physical violence by partner or ex-partner. Indeed current or ex-partners commit the most serious violence and are involved in 62.7% of rapes. IPV can include sexual assault: according of World Health Organization, Department of Injuries and Violence Prevention (2002) one in four women experiences sexual violence by her intimate partner. On the other hand, sexual harassment, such as sex-related verbal or physical behavior that is annoying or disrespectful to the person who suffers it Rubinstein (1987) and Piotrkowski and Brannen (2002) is perpetrated more frequently in the work environment by colleagues or employers. IPV and sexual harassment have many similarities: they are both mainly crimes against women by known perpetrators, and occur in places perceived as safe by victims, like the home or the workplace (Lawson, 2012).

Although it cannot be viewed as a unidirectional phenomenon, IPV concerns a higher percentage of violence of the man against the woman (World Health Organization, 2013). Furthermore, according to World Health Organization (2013) even if IPV can occur against men, men injured by their partners had high rates of IPV perpetration themselves and the violence carried out by women may often present itself as self-defense.

Initially, this phenomenon was investigated within Feminist Movements. In this perspective, IPV was linked to male dominance, rooted in the patriarchal traditions of heterosexual relationships, expressed through control and power dynamics (Dobash and Dobash, 1979; Pence and Paymar's, 1993; Ferraro, 1997; Campbell et al., 1998). According to a recent overview (Bell and Naugle, 2008) the "Feminist Theory," and the "Power Theory" constitute the *Sociocultural theories*, that derive the roots of violence not only from culture but also from the family structure (Straus, 1976).

On the other hand *Individual Theories*, include the "Social learning theory," the "Background/situational model" (Riggs and O'Leary, 1996) and the "Personality/typology theories" that bring the origins of violent conduct back to behaviors learned during childhood (Mihalic and Elliott, 1997; Shook et al., 2000), or to situational factors or elements linked to individual background or, again, to personal characteristics of victims and perpetrators (Koss et al., 1994).

These two classifications were mentioned to explain the complexity of the phenomenon of IPV, whose origins can be traced both at a sociocultural level and in the relationship dynamics of the specific couple.

Other studies focused on the descriptive factors of different types of IPV. Johnson (1995) has distinguished two forms of male violence against female partners: *intimate terrorism* and *situational couple violence*. This distinction may be important in planning prevention and intervention programs and to understand the specific consequences that these two forms of violence can have at the psychic level. In *intimate terrorism* the perpetrator imposes strict control on the partner, through emotional abuse, using children, isolation, threats, intimidation, economic abuse, and blaming. On the other hand, *situational couple violence* concerns a certain altercation that turns into an unstoppable series of escalating violence but with no evidence of the perpetrator exerting control over the partner (Kelly and Johnson, 2008). It is most likely to be described within the conceptual framework of family conflict theory (Straus and Gelles, 1990; Bradbury et al., 2001). Intimate terrorism is probably best conceptualized through the patriarchal pattern of male dominance (Frieze and Browne, 1989). This violence is rarely an isolated incident, as it often turns into more severe episodes of violence in an escalation (Walker, 1977; Coleman, 1997), which may have dangerous consequences for the partner's physical, psychological, and social well-being.

Several studies suggest that depression, panic attacks, inability to cope, suicide attempts, non-suicidal self-injury, post-traumatic stress disorder (PTSD), and alcohol or drugs abuse may be some possible consequences of IPV on the health of the victims (Campbell, 2002; Ellsberg et al., 2008; Pico-Alfonso et al., 2008; Gargiulo et al., 2014). However, few studies in this field have underscored the role of the subjective affective experience of victimization. The different forms in which the IPV can manifest itself within the couple can even result in different affective reactions (Jaquier and Sullivan, 2014).

Psychological violence is always present where there is any other form of violence within a romantic relationship and it is identified as their main source of distress by women who have suffered from IPV (Murphy and O'Leary, 1989; Ronfeldt et al., 1998; Hamby and Sugarman, 1999). IPV, in its form of *Intimate terrorism* can be conceptualized as a sort of private dictatorship that is developed through progressive and disguised attacks. The implicit aim of the abuser is to deprive the victim of his/her individuality, destroy his/her subjectivity, by imposing strict control and exerting physical and psychological violence in order to make the victim a powerless object at the mercy of the dominant partner. The affect of terror seems to play a major role. When the violence appears to be isolated and not restricted to a relationship that assumes the characteristics of private dictatorship, the affect of fear is more likely to be present with the behavioral reaction that follows. The victim would be forced to escape in anguish or, alternately, to react with anger and attack (Nunziante Cesàro and Troisi, 2016).

The subjective affective experience of the victim of violence plays a relevant role in the maintaining the violent relationship (Herman-Lewis, 1992; Hirigoyen, 2005).

Many studies testified how IPV is more difficult to be recognized both from the victims' perspective (Herman-Lewis, 1992; Hirigoyen, 2005; Reale, 2011) and from the society perspective (Romito, 2005; Arcidiacono and Di Napoli, 2012).

Few studies focused on the reasons for exiting or remaining in the violent relationship (Bell and Naugle, 2005).

Several authors showed that the silence of victims of IPV and their ability of carry out help seeking strategies can be influenced by a combination of different factors (Tjaden and Thoennes, 1998; Rennison and Welchans, 2000) Together with cognitive, social and psychosocial factors, emotional factors, such as emotional dependence, fear, guilt and shame, play a main role in maintaining of abusive relationship (Tjaden and Thoennes, 2000; Margherita and Troisi, 2014).

This study focuses on emotional factors maintaining the violent relationship in particular guilt, shame, fear, and terror.

Affect and Trauma in Intimate Partner Violence

Trauma is the main consequence of IPV (Resnick et al., 1993; O'Keefe, 1998; Ehrensaft, 2009). If the traumatic events have occurred repeatedly or chronically, complex PTSD is diagnosed (Herman-Lewis, 1992). This involves specific alterations in affect regulation.

Several studies remark on the importance of emotion dysregulation in PTSD (Van der Kolk, 1996; Cloitre, 1998) since it leads to a lack of awareness of the emotional states the trauma may induce (Litz et al., 2000; Bouton et al., 2001; Hunt and Evans, 2004; Orsillo et al., 2004; Veazey et al., 2004). Through the negative effect on interpersonal relationships and on an individual's overall functional capacity, emotion dysregulation may have an impact on the maintenance of PTS symptoms (Cloitre et al., 2002). Particularly, negative emotions were important for understanding the PTSD (Dagleish and Power, 2004; Resick and Miller, 2009). Shame and guilt contribute to the development and maintenance of PTSD (Lee et al., 2001; Wilson et al., 2006).

Few empirical studies have addressed the specific relation between emotions and PTSD in IPV.

In traumatic experiences such as sexual/physical abuse perpetrated by a known and/or trusted perpetrator, heightened levels of shame compared to fear that would probably accompany a trauma characterized by physical threat, in particular among women (Andrews et al., 2000).

In this study guilt, shame, fear, and terror are considered "affects" within a metapsychological and psychoanalytic framework.

"Affects" were defined as a range of emotions, feelings and passions, which could be represented by a metaphorical image (Green, 1973; Imbasciati, 1991).

Psychoanalytic theories on the trauma suggest how it leads to the collapse of the construction of the meanings processes (Bohleber, 2007; Levine, 2014) and how it disrupts the capacity for representation of mentalization (Levine, 2014).

In literature little attention has been devoted to the affect of shame in interpersonal violence.

In victims of violence, the sense of passivity and helplessness and the feeling of being treated as an object could be traced back to the affect of shame, understood as something that makes the victim feel exposed, naked at the mercy of the other, who, as in the primary impotence at the origin of the life, has the power of life and death over the subject (Margherita and Troisi, 2014). A masking of shame through guilt can more easily permit forgiveness, through a reparative gesture, assuring the maintenance of the link with the partner and restoring an active position in the relationship by taking responsibility for other's behavior. This could explain why self-blaming and silence are such widespread phenomena in IPV (Margherita et al., 2014).

More recently the psychodynamic of affects is moving toward a more precise differentiation between guilt and shame (see, for example Tisseron, 1992; Morrison, 1999; Tangney and Dearing, 2002; Ciccone and Ferrant, 2015). Shame was conceptualized as an archaic and destructive affect that draws the individual's primary impotence and puts a narcissistic failure at stake. Shame accompanies the perception of a failure and the Self is placed in a passive state, where hiding prevails (Morrison, 1999).

In contrast, guilt may be associated with transgression and the Self remains active, absorbed in the action, even during the repair (Tisseron, 1992).

A previous qualitative study allowed an in-depth analysis of the role played by the affects of fear, shame, and guilt in women victims of IPV (Nunziante Cesàro and Troisi, 2016). Authors underscored the difference between fear, associated with the escape from danger and therefore understood as an active defense, and terror associated with paralysis and freezing, in line with psychoanalytical (Diel, 1956; Clit, 2002) and neurophysiological studies (Hagenaars et al., 2014). Considering the three possible reactions that an individual can develop in the face of danger, the attack is associated with anger, the escape is associated with fear and abandonment is associated with terror. Fear, therefore, seems to be a protection that puts the subject in a state of activity and makes them alert, activating sensorial and perceptive systems linked to the awareness of an event that is perceived as traumatic (Nunziante Cesàro and Troisi, 2016).

It would be inappropriate to involve the affect of fear, instead, in situations of private dictatorship because it presupposes an actual danger and foresees a peculiar reactivity on both the behavioral and psychic level.

The situations of extreme violence crystallize the experiences of archaic terror, re-actualize the proven perceptions and the defenses used by the subject (Nunziante Cesàro and Troisi, 2016).

Measuring the Traumatic Affect in Intimate Partner Violence

Affectivity involved in IPV requires valid and specific tools to be measured for quality and quantity. Among existing screening instruments used for the identification of women victims of IPV as the Index of Spouse Abuse (ISA; Hudson and McIntosh, 1981), the Abuse Risk Inventory (ARI; Yegidis, 1989), the Composite Abuse Scale (CAS; Hegarty et al., 1999), or the Conflict Tactics

Scale (CTS; Straus, 2017) should be mentioned. However, these screening tools seem to take into account all possible forms of violence. More importantly, to our knowledge (Rabin et al., 2009) a comprehensive examination of their psychometric properties is lacking. There are also several validated IPV risk assessment instruments, for example Danger Assessment (DA; Campbell, 2004) to assess risk factors for Intimate Partner Femicide, the Ontario Domestic Risk Assessment (ODARA; Hilton et al., 2004) and Spousal Assault Risk Assessment (SARA; Kropp and Hart, 2000; Baldry, 2006).

However, with the aim of measuring the consequences of IPV, several, non-IPV specific tools are used, and the distress is often just evaluated asking the women to assess their general mood.

Examples of tools used for symptom detection, or multidimensional self-report symptom inventories include: Symptom Checklist-90-R (SCL - 90; Derogatis and Cleary, 1977), or scales that investigate specific dimensions such as depression, using by Beck Depression Inventory (BDI; Beck, 1961), or PTSD tools like The Post-traumatic Stress Disorder Checklist (PCL; Weathers et al., 1991) or the Peritraumatic Dissociative Experiences Questionnaire (PDEQ; Marmar et al., 1997) or the Peritraumatic Distress Inventory (PDI; Brunet et al., 2001) or scales that measure psychological well-being, such as the Psychological General Well-Being Index (PGWBI; Dupuy, 1984), or the quality of life, such as Quality of life (QOL; Flanagan, 1978) or on resilience, such as Resilience Scale for Adult (RSA; Friberg et al., 2003).

However, no tools were developed to measure the traumatic impact that IPV can generate on the affective world of women, hence the necessary importance to the level of emotional abuse that accompanies victimization is not considered (Jaquier and Sullivan, 2014).

The development of a valid and reliable scale could aim at measuring the post-traumatic affectivity in situations of IPV, facilitating the identification and the therapeutic process of women victims of IPV, as well as making the responses of health services more sensitive and structured. This study illustrates methodological steps aimed at the development of a self-report instrument for identifying the post-traumatic affectivity in women who have suffered from IPV.

MATERIALS AND METHODS

Participants

The sample comprised 302 Italian women (M : 30.63; SD : 18.5 years) recruited online, through mailing lists and social networks. The whole sample was split randomly into two congruous subsamples (subsamples A and B) for the analysis of its factor structure (Bollen, 1989).

The two sub-samples did not differ significantly in age ($t_{290} = 1.39$, $p = 0.164$) marital status ($t_{300} = 0.124$, $p = 0.901$), nor awareness of violence ($t_{217} = 1.94$, $p = 0.06$) and period of violence ($t_{217} = 1.58$, $p = 0.116$).

Subjects included in the whole sample were mostly unmarried (81.4%) while 12.6% were married, 1.9% divorced, 3.5% separated and 0.6% through a divorce.

Regarding sexual orientation, 87.1% stated that they were heterosexual, 6% bisexual, 2.2% homosexual. The study participants mostly had a high level of education: 31.3% had a master's degree and 20.2% a 3-year degree, 16.7% a post-graduate degree and 30% a high-school diploma. As regards the employment, 34.4% of the entire sample were students, 25.5% were self-employed workers, 15.5% were employed, 5.4% was made up of women without any employment and the 1.9% of the sample were managers.

At the time of the compilation of the questions, 71.3% of the subjects were involved in a current romantic relationship, while 28.4% were single. Moreover, 84.5% of women had no children and the remaining had from one to three offspring.

Measures

Intimate Violence and Traumatic Affect Scale (VITA Scale) (Troisi, 2017)

The original version of the VITA is an Italian 28-item self-report measure used to assess the intensity to affects in women that suffered from IPV. Of all the items, 5 were related to the affect of *fear*, 7 belonged to the affect of *terror*, 10 were related to *shame* and 6 to *guilt*. Items are rated on a Likert type scale (from 1 = *never* to 5 = *often*). In the present study, the Cronbach's alpha was 0.79 for Fear, 0.9 for Terror, 0.93 for Shame, 0.87 for Guilt and 0.93 for the total scale. The process of the development of VITA Scale is described below.

Intimate Partner Violence Check List

The checklist was obtained from the National Association DiRe "Networking of Women against violence," the first Italian association of non-institutional anti-violence centers. The descriptions are set up in the form of questions rated on five-point Likert type scale (from 1 = *never* to 5 = *always*). Five forms of violence are included: *psychological violence* (18 items) containing every form of abuse that damages the identity of the victim; *sexual violence* (3 items) including the imposition of sexual practices or sexual relationships that cause physical harm obtained through threats of various kinds; *physical violence* (7 items) including the use of any act guided by the intention to do evil or to terrorize the woman who has suffered violence; *stalking* (8 items) including controlling behavior performed by the persecutor, *economic violence* (6 items) as a form of direct control, which limits the victim's economic independence. This checklist was used for descriptive purposes to identify what kind of violence the study participants suffered.

Questions About Violence

Further questions on the awareness of violence (*Have you ever suffered any form of violence?*), on the period of her life in which the violence had been experienced, on the author of the violence, on the intensity of suffered violence (if isolated or repeated).

Questions About Help Seeking

These questions aimed at identifying the presence or absence of reporting and a help seeking process, and if present the type of help requested (informal and formal), if absent the reasons behind the failure to request help.

Difficulties in Emotion Regulation Strategies (DERS) (Gratz and Roemer, 2004)

It is a 36-item multidimensional self-report measure of difficulties in emotion regulation. Items are assessed on a five-point Likert scale ordered from 1 = *almost never* to 5 = *almost always*. The DERS assesses difficulties in six clinically relevant dimensions of emotion regulation: (a) non-acceptance of emotional (*Non-acceptance*), (b) difficulty engaging in goal-directed behavior in distress situations (*Goals*), (c) inability to control behavior when distressed (*Impulse*), (d) lack of awareness of emotions (*Awareness*), (e) limited access to strategies that are perceived as effective for emotion regulation, and (f) lack of emotional clarity. The DERS showed adequate construct and predictive validity, as well as good test-retest reliability ($\rho I = 0.88$; Gratz and Roemer, 2004). Also, the Italian adaptation it showed good psychometric properties (Giromini et al., 2012). In the present study, internal consistencies for the total and subscale scores were good, ranging from 0.81 (for *Awareness*) to 0.89 (for *Non-acceptance*).

Impact of Event Scale (IES) (Horowitz et al., 1979)

It is a self-report measure composed of 15 items on a four-point Likert scale (ranging from 1 = *not at all* to 4 = *often*). The IES tapped two-specific answers to traumatic events: (a) *intrusion*, intended as emerging of undesired ideas, images, feelings, or dreams that remind to the event and (b) *avoidance*, intended as elusion of certain ideas, feelings, or situations linked to the stressful situation. In its Italian version Cronbach's alphas were 0.84 for the intrusion subscale and 0.71 for the avoidance subscale (Pietrantonio et al., 2003). In the present study, the IES total score yielded Cronbach's alpha of 0.93, 0.92 for the *intrusion* subscale and 0.89 for the *avoidance* subscale.

Other as Shamer Scale (OAS) (Goss et al., 1994)

It includes 18 items to measure the external shame, as a global judgment about how the self is evaluated by others. Items are rated on a five-point Likert-type scale (ranging from 0 = *never* to 4 = *almost always*). It was made up of three subscales: (a) *inferiority*, related to being seen as inferior; (b) *emptiness*, related to being seen as empty; and (c) *mistake*, related to how others are vigilant to mistakes one makes (Goss et al., 1994). In the Italian version of OAS, the Cronbach's alphas were 0.87 (Balsamo et al., 2015c; Saggino et al., 2017).

In the present study, the OAS yielded Cronbach's alpha of 0.94 for the total score and 0.92 for the *Inferiority*, 0.83 for *Emptiness*, and 0.86 for *Mistake* subscale.

Coping Orientation to Problems Experienced 25 (COPE-NVI-25) (Foà et al., 2015)

This 25-item scale is a reduced form of the Coping Orientation to Problem Experienced (COPE-NVI) developed by Carver et al.

(1989). Items ask to assess how often the subject implements a certain coping process in difficult or stressful situations. The selected subscales measure: *Avoidance Strategies* (5 items), which concern the negation and natural detachment; *Transcendent Orientation* (4 items); *Positive Attitude* (6 items); *Social Support* (5 items) related to the search for understanding, information and of emotional outpouring; *Problem Orientation* (5 items) related to the use of active planning strategies and suppression of alternative activities. The Cronbach's alpha was 0.70 for all dimensions, excepting for *avoidance strategies*, which nevertheless presents values considered satisfactory (Sica et al., 2008). In our sample, for the COPE-NVI 25 total score Cronbach α is equal to 0.85. For the different subscales, it was 0.76 for *avoidance strategies* 0.96 for *transcendent orientation* = 0.80 for *positive attitude*, 0.94 for *social support* and 0.83 for *orientation to the problem*.

Teate Depression Inventory (TDI) (Balsamo and Saggino, 2014; Balsamo et al., 2014)

It is a 21-item self-report tool that aims to measure depressive symptoms as described by the latest editions of the Diagnostic and Statistical Manual of Mental Disorders (DSM; American Psychiatric Association, 2013) on a five-point Likert-type scale (ranging from 0 = *always* to 4 = *never*). It was developed via Rasch logistic analysis of responses (Rasch, 1960), within the framework of Item Response Theory, in order to overcome psychometric weaknesses of existing measures of depression (Balsamo and Saggino, 2007). Recent literature suggested that TDI demonstrate good psychometric properties (Balsamo et al., 2013a,c, 2015a,b; Innamorati et al., 2013; Carlucci et al., 2018; Contardi et al., 2018; Saggino et al., 2018). In the present sample, Cronbach's α was 0.95.

Procedure

Questionnaire Development

Generation of the preliminary item list

In a qualitative study, a group of 10 women (age *M*: 42.25; *SD*: 4.9 years) who had suffered from IPV and who had sought help services, were interviewed. Affects of guilt, shame, fear, and terror were identified and explored as associated with the situation of violence suffered by the women (Nunziante Cesàro and Troisi, 2016). A pilot study was carried out, through an online *ad hoc* questionnaire developed, in order to test this method of administration. The on-line administration appeared to be more appropriate for recruiting participants who did not want to access the help services, because guaranteed them protection and respect for their own privacy. Furthermore, the pilot study allowed investigation of the means of expression, the sequencing rule for the questions and the types of IPV suffered and to evaluate the response format (Troisi, 2017).

A qualitative selection of the pool of items was carried out on the basis of the words used by the women in the qualitative study, the results from the pilot study and the theoretical assumption. Items were expressed in the metaphorical form. Typically, the language of affects can be more readily evoked by the use of metaphor, often linked to a shared collective

symbolization (Imbasciati, 1991; Tisseron, 1992). Therefore, items have been organized through their insertion into different areas related to the following affects: fear, terror, shame, and guilt.

Within the semantic area related to the fear, fear was considered as a state of alarm elicited by the avoidance of the danger (Diel, 1956; Hagedaars et al., 2014). Within the semantic area of the terror, this affect was framed as a paralyzing state that hinders an active process of reaction to danger (Clit, 2002; Nunziante Cesàro and Troisi, 2016). The semantic area of the shame defines it as a strong exposure to the other that disarmed the individual and makes him animated by a sense of failure and passivity (Tisseron, 1992; Lewis, 1995; Pandolfi, 2002; Ciccone and Ferrant, 2015). The semantic area of the guilt focused on its defensive dimension aimed at the restoration of the relationship with the partner, assuming responsibility for the violence suffered and taking an active position in the relationship (Tisseron, 1992; Pandolfi, 2002; Ciccone and Ferrant, 2015).

As a strategy for developing items useful for capturing the meaning of the psychological constructs of affects here defined, three experts, of whom one psychotherapist/researcher and two clinical psychologists, were asked independently to assess items on a Likert type scale. The item pool generated by these procedures comprised 30 items, including 6 for the semantic area of fear, 7 for the semantic area of terror, 10 for the semantic area of shame and 7 for the semantic area of guilt. A five-point Likert-type scale raking from 1 = *never* to 5 = *often*, was chosen as appropriate response format for these items.

Refinement of the initial item pool

The resulting 30-item pool was examined by a second group of three experts, composed of one psychologist, one psychoanalyst and one social methodologist who were asked to independently rate the relevance to the construct of each item on a 1 to 5 Likert scale (1 = *strongly disagree*, 5 = *strongly agree*). The psychoanalyst and the psychologist expert of the health services set up for violence against women evaluated the relevance of the emerged item pool with theoretical principles and with studied phenomenology in order to guarantee the content validity of the instrument. The methodologist, instead, rated the degree of adherence to the response format and the formulation of the items according to the criteria of brevity, simplicity, exclusion of possible linguistic ambiguities. Based on the collected rating, the 30-item pool underwent syntactic changes and reformulations, which led to a reduction in the number of items. Two items were deleted; one, related to the area of the affect of fear, was evaluated redundant, and another, belonging to the area of the affect of guilt, resulted ambiguous. Furthermore, linguistic ambiguities, double statements, multiple negations and redundant frequency adverbs in the response format were deleted and some changes were made to the instructions and to the terminology.

At the end of this selection, 28 items were retained and grouped as follows: 5 related to fear, 7 related to terror, 10 related to shame, and 6 related to guilt.

The 28-item pool was submitted to a further screening aimed at examining its comprehensibility. It was administered

to the same group of 10 women who had participated in the qualitative study described above, since it was considered as a representative sample of the population under examination. This preliminary administration confirmed the comprehensibility of the item and, therefore, did not result in any changes.

All aspects of the study involved the informed consent of each participant, according to the ethical guidelines of Helsinki Declaration. Moreover, participants were informed about the confidentiality of their responses and their anonymous treatment. Participants read a web page with the informed consent document before starting the online survey. The online consent form containing all the required elements, consisting of purpose of the research, nature of participation, description of research procedures, description of risks, voluntariness of participation, right to withdraw at any time without penalty, handling of data (anonymity and confidentiality), contact information for researcher, and contact information for concerns about the project were read and submitted by clicking a button below the text to consent to participate to the survey. There was no honorarium for completing the assessment.

The protocol was approved by the ethics committee of Section of Psychology and Education Sciences, University of Naples Federico II, Italy.

Statistical Analysis

A split-sample cross-validation procedure (Bollen, 1989) was performed on our sample. Data from subsample A and subsample B were respectively subjected to an exploratory factor analysis (EFA- study 1), and to a confirmatory factorial analysis (CFA-study 2), based on the factor structure derived from the exploratory analysis. Model fit was measured by means of the following fit indexes that are suggested as most important (Hu and Bentler, 1998, 1999; MacCallum and Austin, 2000): (a) the chi-square (χ^2) statistic and its degrees of freedom; (b) the Tucker-Lewis Index (TLI); (c) the comparative fit index (CFI); (d) the root-mean-square error of approximation (RMSEA) and its 90% confidence interval (CI); and (e) the standardized root mean square residual (SRMR). According to Schermelleh-Engel et al. (2003), the model fit the data when: $\chi^2/df < 2$, CFI and TLI > 0.97 , SRMR < 0.05 , and RMSEA < 0.05 (90% CI: the lower boundary of the CI should contain zero for exact fit and be < 0.05 for close fit), in any case also values between 0.05 and 0.08 were considered by some authors as indicative of a good suitability of the model (Browne and Cudeck, 1993; Hu and Bentler, 1999).

Cronbach's alpha and correlations have been used to assess descriptive statistics and internal consistency.

RESULTS

Study 1

Subsample A

The sample included 151 participants (age M : 30.23; SD : 8.87 years). As regards their marital status, 83.4% of the

women was unmarried, 10.6% married, 2.6% divorced, 2% separated, and 1.3% through a divorce. As regards the sexual orientation 83.4% of the women were heterosexual, 5.3% bisexual and 1.3% homosexual. Regarding level of education: 39.1% had a master's degree, 21.2% a 3- year degree and 21.3% a high-school diploma and 16.7% a post- graduate degree. As regards the employment 29.8% of subjects were students, 25.2% were self- employed workers, 14.6% were employed and 4.6% were unemployed. Moreover, 68.9% of study participants were involved in a current romantic relationship, 88.1% had no children and the remaining had from one to three offspring.

Exploratory factor analysis (EFA)

The structure of the VITA Scale has been evaluated through a series of EFA using the principal axis factoring (PFA) method in subsample A. PFA has been chosen because of its capacity to recover weak factors and be fairer than principal component analysis (Widaman, 1993) especially when working on small samples (Briggs and MacCallum, 2003). Firstly, it has been tested a one-factor model where all the VITA Scale items were free to load on a single latent component. The one-factor solution explained the 37.54% of the total variance, with eigenvalue equal to 11.05. Absolute factor loadings for each item were greater than 0.30, except for the items (#5) "Ho reagito alla paura chiedendo aiuto" ("I reacted to the fear by asking for help") and (#3) "L'agitazione mi ha spinto a reagire" ("Agitation pushed me to react"). However, a careful inspection of the scree test (Cattell, 1966) and the inclusion of the factor with eigenvalues > 1 (Kaiser, 1960) suggested the extraction of four or five factor latent components. Based on the previous results, a second PFA was performed extracting five factors with Direct Oblimin rotation. Despite the solution accounting for more than 60% of the total variance, several double factor loadings (>0.30) were observed in the pattern structure. Again, the fifth factor resulted to be composed of a single item "Mi sento/sentivo sporca" (#20) ("I fee/felt dirty"). Therefore, the five-factor model tested could not retained as a reliable solution, both from the statistical and theoretical examination. The last model tested the presence of four factor latent components. Following the authors construct theory of the VITA Scale, a four-factor solution was extracted using PFA with Direct Oblimin rotation. The Kaiser-Meyer-Olkin index was 0.900 suggesting an appropriate measure of sampling adequacy (Tabachnick and Fidell, 2007). The significant Bartlett's Test of Sphericity (2767,990; gdl = 153; $p = 0.001$) suggested the adequacy of sample to the EFA. According to the scree test (Cattell, 1966), four factors could be extracted, accounting for the 63.96% of the total variance. All the VITA Scale items showed absolute loadings for each item greater than 0.30 (see **Table 1**). Only six items showed secondary loadings (#2, #10, #13, #15, #18, and #25). Based on the content analysis, nine items (from #13 to #22, without #18) loaded on the first factor called "Shame"; five items (from #1 to #5) loaded on the second factor called "Fear"; six items (#from #23 to #28) loaded on "Guilt" factor; and, eight items (from #6 to #12 and item #18) loaded on factor defined "Terror."

Study 2

Subsample B

The sample included 151 participants ($M: 30.53$; $SD: 23.7$ years). 79.5% of the women was unmarried, 14.3% married, 0.7% divorced, 5.3% separated. 90.7% of the women are heterosexual, 6% bisexual, 3.3% homosexual. Regarding level of education: 40.4% had a high-school diploma 22.5% had a master's degree, 18.5% a 3- year degree and 17.2% a post- graduate degree. 41.7% of the women were female students, 27.2% was self-employed workers, 6.6% were without any employment, 13.2% were employed.

74.2% were in romantic relationship, 84.1% had no children and the remaining had from 1 to 3 offspring.

Confirmatory factor analysis (CFA)

A CFA (Bollen, 1989) was carry out using the MPLUS 7 statistical package (Muthén and Muthén, 2012) on the subsample B. Descriptive statistics for the Subsample B revealed no missing values and several departures from the normality of the data. Specifically, item #10 showed skewness and kurtosis values that exceeded the cut-off criteria of ± 3 (Curran et al., 1996).

Due to the asymmetrical distribution of data, the responses to the VITA Scale items should be better evaluated at the categorical rather than the metric level. However, the robust unweighted least squares (ULSMW) method using a diagonal weight matrix and robust standard errors and a mean- and variance adjusted χ^2 test statistic (Muthén and Muthén, 1998, 2004) was used to estimate parameters. Like WLSMV, the ULSMV estimator were more likely to catch small structural links with precision when data was slightly or moderately asymmetric, and when small sample sizes were used.

The one-factor and the four-factor refined models (without items with double loadings) emerged from the previous EFA, versus the four-factor model, that followed authors' theoretical assumptions (Troisi, 2017), were tested. The unstable five-factor model was excluded from the comparison, based on the results from the previous EFA.

As seen in **Table 2**, the four-factor model fit the data slightly better than the refined four-factor model and the one-factor model. Specifically, the one-factor model fitted the data worst. All the chi-squared values were significant ($p < 0.001$), and the ratio χ^2/df indicated a slightly better fit of the four-factor model ($\chi^2 = 540.789$; $\chi^2/df = 1.57$) than the refined model ($\chi^2 = 355.389$; $\chi^2/df = 1.75$). Likewise, CFI (four-factor, 0.970; refined model, 0.960) and RMSEA (four-factor, 0.062; refined model, 0.071) indices confirmed the better fit of the four-factor model. The TLI index of the two models showed no difference. TLI and CFI were all above 0.97 and SRMR was close to 0.05 (Schermelleh-Engel et al., 2003), indicating a close fit of both models to the empirical data. These results showed that the exclusion of the items with EFA double loadings (#2, #10, #13, #15, #18, and #25) did not contribute to improving the model.

In **Figure 1** the standardized factor loadings of the four-factor CFA model was shown, as well as the path coefficients, among

TABLE 1 | Exploratory factor analysis (EFA) loadings performed on subsample A (N = 151).

#Item	Original dimension	Item	Factor			
			1	2	3	4
19	Shame	VITA ITEM 19	0,960			
16	Shame	VITA ITEM 16	0,911			
21	Shame	VITA ITEM 21	0,619			
14	Shame	VITA ITEM 14	0,619			
20	Shame	VITA ITEM 20	0,591			
17	Shame	VITA ITEM 17	0,536			
22	Shame	VITA ITEM 22	0,532			
15	Shame	VITA ITEM 15	<i>0,534</i>		0,351	
13	Shame	VITA ITEM 13	<i>0,360</i>		0,326	
3	Fear	VITA ITEM 3		0,727		
1	Fear	VITA ITEM 1		0,689		
5	Fear	VITA ITEM 5		0,627		
4	Fear	VITA ITEM 4		0,479		
2	Fear	VITA ITEM 2		0,551		-0,409
24	Guilt	VITA ITEM 24			0,910	
26	Guilt	VITA ITEM 26			0,737	
27	Guilt	VITA ITEM 27			0,717	
28	Guilt	VITA ITEM 28			0,706	
23	Guilt	VITA ITEM 23			0,602	
25	Guilt	VITA ITEM 25			0,396	-0,338
8	Terror	VITA ITEM 8				-0,821
9	Terror	VITA ITEM 9				-0,811
12	Terror	VITA ITEM 12				-0,802
7	Terror	VITA ITEM 7				-0,768
6	Terror	VITA ITEM 6				-0,600
11	Terror	VITA ITEM 11				-0,591
18	Shame	VITA ITEM 18	0,386			-0,447
10	Terror	VITA ITEM 10		0,317		-0,317
Eigenvalues			10,678	3,267	1,393	0,968
% explained variance			38,134	11,67	4,975	3,458

In *Italic* double item loadings.

TABLE 2 | Confirmatory factorial analysis (CFA) indices performed on the subsample B.

Models	χ^2	df	χ^2/df	TLI	CFI	RMSEA	90% CI
One factor	704.995	350	2,01	0.939	0.943	0.082	0.073 0.091
Four factors	540.789	344	1,57	0.966	0.970	0.062	0.051 0.071
Four factors refined*	355.389	203	1,75	0.965	0.960	0.071	0.058 0.083

df, degrees of freedom; TLI, Tucker-Lewis Index; CFI, Comparative Fit Index; RMSEA, root-mean-square error of approximation; 90% CI, 90% confidence interval of RMSEA; SRMR, standardized root mean square residual.

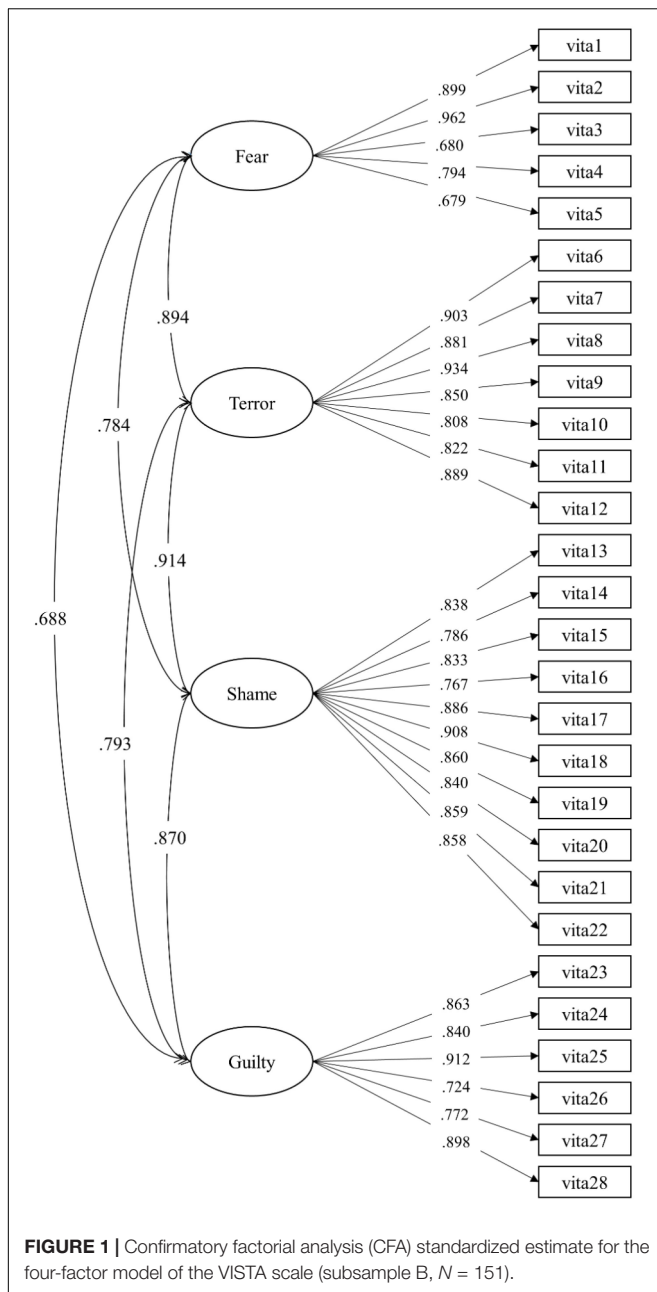
*Four factor refined model without EFA double loadings items #2, #10, #13, #15, #18, and #25.

the four latent factors. All the items loaded considerably (>0.75) on the respective factors, and all the four latent factors highly correlated (from 0.68 to 0.89).

Construct validation

Pearson correlational analyses were used to explore the associations between the VITA subscales and other related measures (Table 3).

The subscales of the VITA Scale *Shame* and *Guilt* were positively correlated with the DERS subscales *Non-acceptance*, *Goals* and *Impulse*, with both subscales of the IES (*Intrusion* and *Avoidance*) and with all subscales of the OAS (*inferiority*, *emptiness*, and *mistake*). Furthermore, the subscale of *Shame* and *Guilt* were positively correlated with the subscale of COPE-NVI 25 that were *Avoidance Strategies* and *Positive Attitude* and with the total score of the TDI. The VITA subscale of



Shame was positively correlated with the DERS subscale of *Strategies*.

The VITA subscale of *Terror* was positively correlated with the DERS subscales of *Non-acceptance*, *Goals* and *Impulse* and *Strategies*, with both the subscales of IES (*Intrusion* and *Avoidance*), with all subscales of OAS and with the subscale of COPE-NVI 25 related to *Avoidance Strategies*.

VITA-Scale *Fear* was positively correlated with both the IES subscales (*Intrusion* and *Avoidance*) and positively correlated with the COPE-NVI 25 subscale of *Problem Orientation*.

Significant correlations were found between the total score of the VITA and the total scores the used concurrent measures, that were OAS, IES, DERS ($p < 0.01$) and TDI ($p < 0.05$).

DISCUSSION

Intimate partner violence is the most common form of violence against women (Campbell, 2002). Unfortunately, it is also the most difficult form of violence to recognize (Herman-Lewis, 1992; Hirigoyen, 2005).

The results of the current study widely confirmed the main role played by four affects of terror, fear, guilt and shame, in situation of IPV in line with the authors' theoretical assumptions (Margherita and Troisi, 2014; Margherita et al., 2014; Nunziante Cesàro and Troisi, 2016).

Particularly, this research confirmed the main role, within the women's subjective affective experience of victimization, of the affect of shame (Follingstad et al., 1991; Sippel and Marshall, 2011; Shorey et al., 2011), of guilt (Beck et al., 2011), and of fear (Kilpatrick et al., 1989; Riggs et al., 1992; Weaver and Clum, 1995; Scheffer Lindgren and Renck, 2008).

Summing up, the current study aimed at measuring the variety and the complexity of the post-traumatic affectivity of the women suffered from IPV experienced through a specific and newly developed instrument, named VITA Scale (Intimate Violence and Traumatic Affects Scale).

This scale showed a clear factor structure and strong psychometric properties in a sample composed of women suffered from IPV.

Reliability analysis indicated that the VITA Scale, as well its subscales related to the different affects, showed an excellent Cronbach alpha value. The EFA and CFA showed a fully satisfactory fit. The dimensions emerging from these analyses were in line with theoretical expectations (Margherita and Troisi, 2014; Margherita et al., 2014; Nunziante Cesàro and Troisi, 2016).

Results of correlational analysis was in line with the theoretical expectations: the intensity of post-traumatic affectivity was positively and significantly correlated with external shame, the intensity of depression, the impact of trauma, the affective dysregulation and the lack of strategies and it is not related to the implementation of coping skills. The VITA subscale of *terror* was positively correlated with the subscale *avoidance strategy* of the COPE-INV25 and the subscale of *fear* was positively correlated with the *Problem Orientation* subscale of the COPE-NVI. These last two correlations showed that fear was more associated with the possibility of facing the problem in an "activity" dimension, while terror was more associated with a "passive" avoidance response.

Particular attention should be devoted to the distinction of the affects of terror and fear. While fear assumes a more protective function, terror is denoted as a psychic state, more intense than fear: it emerges when facing with a threat of extreme danger, that could be real or fictitious and would lead to a state of passivity (Clit, 2002; Nunziante Cesàro and Troisi, 2016; Troisi, 2017).

Previous studies, not making a specific distinction between fear and terror, considered fear in the IPV resulting from both the perceived risk of violence and the uncontrollability of this risk (Smith et al., 1995). The distinction between fear and terror, as proposed in the current study, is also supported by neurophysiological studies, which underlined that neuronal circuits of the amygdala, the hypothalamus and the

TABLE 3 | Pearson correlations between VITA and external measures and descriptive statistics.

	VITA subscale					Descriptive statistics		
	Shame	Guilt	Terror	Fear	Total scale	a	M	SD
DERS	0.242**	0.315**	0.303**	0.043	0.252**	0.86	80.95	22.25
Non-accept	0.338**	0.313**	0.305**	0.062	0.346**	0.89	14.91	6.20
Goals	0.401**	0.285**	0.386**	0.117	0.346**	0.85	5.54	2.44
Impulse	0.262**	0.187*	0.252**	0.097	0.202*	0.85	14.72	5.56
Awareness	-0.132	-0.059	-0.041	-0.097	-0.096	0.81	11.20	4.2
Strategies	0.239**	0.158	0.194*	-0.025	-0.066	0.82	15.63	4.74
Clarity	-0.004	-0.75	0.004	-0.021	0.147	0.84	19.11	6.36
IES	0.550**	0.328**	0.534**	0.327**	0.575**	0.93	34.24	12.6
Intrusion	0.510**	0.349**	0.508**	0.294**	0.544**	0.92	15.76	6.61
Avoidance	0.499**	0.258**	0.486**	0.324**	0.520**	0.89	18.27	7.11
OAS	0.349**	0.461**	0.331**	0.120	0.382**	0.94	41.44	16.19
Inferiority	0.317**	0.435**	0.284**	0.077	0.332**	0.92	17.83	7.93
Emptiness	0.292**	0.334**	0.315**	0.143	0.337**	0.83	8.67	4.13
Mistake	0.327**	0.450**	0.314**	0.128	0.367**	0.86	14.95	5.80
COPE-NVI 25	0.144	0.232**	0.122	0.122	0.144	0.85	86.38	17.19
Avoidance strategies	0.228**	0.267**	0.232**	0.059	0.231**	0.76	12.84	5.056
Transcendent orientation	-0.005	0.041	-0.030	-0.076	-0.037	0.96	8.55	6.17
Positive attitude	0.178*	0.231**	0.102	0.096	0.206*	0.80	24.40	6.193
Social support	0.054	0.062	-0.020	0.070	-0.034	0.94	20.46	6.47
Problem orientation	0.035	0.027	0.091	0.204*	0.046	0.83	20.08	5.41
TDI	0.348**	0.365**	0.202	0.062	0.280*	0.95	40.72	16.13

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

periaqueductal gray substance had sub-zones distinguished for active defenses, such as attack-escape, and for passive ones, such as freezing (Hagenaars et al., 2014).

In the face of danger, the emergence of the affects of terror and shame shows itself as the first defense to maintain the psychic equilibrium, a first form of protection against the disorganization induced by the trauma. If, in dealing with the malaise, the possible actions are no longer adaptive, these affects can come into play with the loss of the feeling of Self.

There are, for example, two different levels of shame: a *toxic shame*, associated with a sense of constriction, anger and withdrawal, and with an intolerable punitive isolation, and on the other level a *humanizing shame* that sharpens empathy. This is a shame that has been given recognition either by others or by themselves or by both (Kilborne, 2002).

Feeling shame is the first sign of subjectivation, of being still a subject among subjects, this affect besides being first human feeling, it is a social feeling at the limit between intra and inter-subjective (Barazer, 2000).

Several studies have shown that the affect of shame is associated with the maintenance of PTSD symptoms over time (Andrews et al., 2000).

In the guilt proneness of women who have suffered from violence, not specifically psychoanalytic approaches identify a tendency to feel regrets or remorse for past behavior judged as wrong, while shame refers more to a lowering of self-esteem and emptiness feelings (Gilbert, 2000) and sense of inferiority to the other (Tangney, 1996; Tangney and Dearing, 2002), and is considered a less adaptive affects than the guilt.

Some other studies have questioned whether shame could be a predisposing factor, rather than a consequence of IPV (Harper and Arias, 2004).

Working with such affects, in situations of violence, can be useful in processing the trauma, in order to return to the victim that functional role for the psychic life of the individual that allows a subject to become aware of his/her own internal world and to inscribe the traumatic experience in temporality (Herman-Lewis, 1992; Bohleber, 2007; Levine, 2014).

It is necessary that these affects can emerge to be recognized and elaborated, to limit the disruptive effects of the trauma and reconstruct the event by placing it in a space and in a time and to increase the woman's ability to think and elaborate those affective experiences that escape any attempt of nomination.

Also, in line with this assumption, the scale of post-traumatic affects here built (VITA Scale) has foreseen the use of metaphor, which stands as a mediator between unspeakable affections and representations (Tisseron, 1992).

It is important to underline that the traumatic experience of the women who suffered from IPV differs from traumas after another stressful event of life. Trauma after IPV is an interpersonal trauma, caused by another person. This type of trauma, whose nature is relational and lasting, is often a "complex trauma" (Herman-Lewis, 1992), whereas the traumatic experience is not a single event but it is repeated and prolonged. This specific situation can make the nervous system reactive, as if in a constant state of alarm and has a higher PTSD risk than other types of trauma, such as trauma associated with natural disasters (Kessler et al., 2017).

The VITA Scale could be a useful tool for the clinician to investigate the affective state of the woman at the time of access to services, to assess the awareness of the woman in her internal world after the trauma, in order to better address the clinical practice and therapeutic intervention planning. Moreover, the use of the instrument could facilitate the recognition of the affects that emerged in the woman following the traumatic experience. This tool could also be useful to broaden the scientific knowledge on the subjective affective experience of victimization, for which several studies have emphasized the need (Harper and Arias, 2004) and for recognition of change in the therapeutic process (Halfon et al., 2016). Furthermore the VITA Scale may be helpful in investigating the role played by affects in different situations of violence.

The treatment of the traumatized woman requires specific clinical work aimed at developing the ability to process traumatic affects and only an adequate tuning with the precise affective states can support the therapeutic alliance and reconstruct the sense of security threatened by the traumatic event (Caretta et al., 2013; De Luca Picione et al., 2017, 2018).

REFERENCES

- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Washington, DC: American Psychiatric Pub. doi: 10.1176/appi.books.9780890425596
- American Psychological Association and Presidential Task Force on Violence and the Family (1996). *Violence and the Family: Report of the American Psychological Association Presidential Task Force on Violence and the Family*. Washington, DC: American Psychological Association.
- Andrews, B., Brewin, C. R., Rose, S., and Kirk, M. (2000). Predicting PTSD symptoms in victims of violent crime: the role of shame, anger, and childhood abuse. *J. Abnorm. Psychol.* 109, 69–73. doi: 10.1037/0021-843X.109.1.69
- Arcidiacono, C., and Di Napoli, I. (eds). (2012). *Sono Caduta Dalle(scale): I Luoghi e Gli Attori Della Violenza di Genere*, Vol. 40. Milano: Franco Angeli.
- Baldry, A. C. (2006). *Dai Maltrattamenti All'omicidio. La Valutazione del Rischio di Recidiva e Dell'uxoricidio*. Milano: Franco Angeli.
- Balsamo, M., Carlucci, L., Murdock, K. K., Sergi, M. R., and Saggino, A. (2015a). The mediating role of early maladaptive schemas in the relation between co-rumination and depression in youths. *PLoS One* 10:e0140177. doi: 10.1371/journal.pone.0140177
- Balsamo, M., Giampaglia, G., and Saggino, A. (2014). Building a new Rasch-based self-report inventory of depression. *Neuropsychiatr. Dis. Treat.* 10, 153–165. doi: 10.2147/NDT.S53425
- Balsamo, M., Innamorati, M., Van Dam, N. T., Carlucci, L., and Saggino, A. (2015b). Measuring anxiety in the elderly: psychometric properties of the state trait inventory of cognitive and somatic anxiety (STICSA) in an elderly Italian sample. *Int. Psychogeriatr.* 27, 999–1008. doi: 10.1017/S1041610214002634
- Balsamo, M., Imperatori, C., Sergi, M. R., Belvederi Murri, M., Continisio, M., Tamburello, A., et al. (2013a). Cognitive vulnerabilities and depression in young adults: an ROC curves analysis. *Depress. Res. Treat.* 2013:407602. doi: 10.1155/2013/407602
- Balsamo, M., Lauriola, M., and Saggino, A. (2013b). Work values and college major choice. *Learn. Individ. Dif.* 24, 110–116. doi: 10.1016/j.lindif.2012.12.022
- Balsamo, M., Macchia, A., Carlucci, L., Picconi, L., Tommasi, M., Gilbert, P., et al. (2015c). Measurement of external shame: an inside view. *J. Pers. Assess.* 97, 81–89. doi: 10.1080/00223891.2014.947650
- Balsamo, M., Romanelli, R., Innamorati, M., Ciccarese, G., Carlucci, L., and Saggino, A. (2013c). The state-trait anxiety inventory: shadows and lights on its construct validity. *J. Psychopathol. Behav. Assess.* 35, 475–486. doi: 10.1007/s10862-013-9354-5
- Balsamo, M., and Saggino, A. (2007). Test per l'assessment della depressione nel contesto italiano: un'analisi critica. *Psicoter. Cogn. Comport.* 13, 167–199.
- Balsamo, M., and Saggino, A. (2014). Determining a diagnostic cut-off on the teate depression inventory. *Neuropsychiatr. Dis. Treat.* 10, 987–995. doi: 10.2147/NDT.S55706
- Barazer, C. (2000). *Hontes sans issue. Documents et débats, Bulletin de l'Association psychanalytique de France*. Paris: PUF.
- Beck, A. T. (1961). A systematic investigation of depression. *Compr. Psychiatry* 2, 163–170. doi: 10.1016/S0010-440X(61)80020-5
- Beck, J. G., McNiff, J., Clapp, J. D., Olsen, S. A., Avery, M. L., and Hagewood, J. H. (2011). Exploring negative emotion in women experiencing intimate partner violence: shame, guilt, and PTSD. *Behav. Ther.* 42, 740–750. doi: 10.1016/j.beth.2011.04.001
- Bell, K. M., and Naugle, A. E. (2005). Understanding stay/leave decisions in violent relationships: a behavior analytic approach. *Behav. Soc. Issues* 14, 21–45. doi: 10.5210/bsi.v14i1.119
- Bell, K. M., and Naugle, A. E. (2008). Intimate partner violence theoretical considerations: moving towards a contextual framework. *Clin. Psychol. Rev.* 28, 1096–1107. doi: 10.1016/j.cpr.2008.03.003
- Bohleber, W. (2007). Remembrance, trauma and collective memory: the battle for memory in psychoanalysis. *Int. J. Psychoanal.* 88, 329–352. doi: 10.1516/V5H5-8351-7636-7878
- Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociol. Methods Res.* 17, 303–316. doi: 10.1177/0049124189017003004
- Bouton, M. E., Mineka, S., and Barlow, D. H. (2001). A modern learning theory perspective on the etiology of panic disorder. *Psychol. Rev.* 108, 4–32. doi: 10.1037/0033-295X.108.1.4
- Bradbury, T. N., Rogge, R., and Lawrence, E. (2001). "Reconsidering the role of conflict in marriage," in *Couples in Conflict*, eds A. Bootj, A. C. Crouter, and M. Clements (Mahwah, NJ: Erlbaum Associates), 59–81.
- Briggs, N. E., and MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behav. Res.* 38, 25–56. doi: 10.1207/S15327906MBR3801_2
- Browne, M. W., and Cudeck, R. (1993). "Alternative ways of assessing model fit," in *Testing Structural Equation Models*, Vol. 154, eds K. A. Bollen, and J. S. Long (Newbury Park, CA: Sage), 136–136.
- Brunet, A., Weiss, D. S., Metzler, T. J., Best, S. R., Neylan, T. C., Rogers, C., et al. (2001). The peritraumatic distress inventory: a proposed measure of PTSD criterion A2. *Am. J. Psychiatry* 158, 1480–1485. doi: 10.1176/appi.ajp.158.9.1480
- Campbell, J., Rose, L., Kub, J., and Nedd, D. (1998). Voices of strength and resistance: a contextual and longitudinal analysis of women's responses to battering. *J. Interpers. Violence* 13, 743–762. doi: 10.1177/088626098013006005
- Campbell, J. C. (2002). Health consequences of intimate partner violence. *Lancet* 359, 1331–1336. doi: 10.1016/S0140-6736(02)08336-8

AUTHOR CONTRIBUTIONS

The author contributes at the whole article, in each of its part.

- Campbell, J. C. (2004). *Danger Assessment*. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.
- Caretti, V., Craparo, G., and Schimmenti, A. (eds). (2013). "Emozioni, corpo, mentalizzazione: il trattamento delle sindromi traumatiche," in *Memorie Traumatiche e Mentalizzazione*, (Rome: Casa Editrice Astrolabio Ubaldini Editore), 169–182.
- Carlucci, L., D'Ambrosio, I., Innamorati, M., Saggino, A., and Balsamo, M. (2018). Co-rumination, anxiety, and maladaptive cognitive schemas: when friendship can hurt. *Psychol. Res. Behav. Manage.* 11, 133–144. doi: 10.2147/PRBM.S144907
- Carver, C. S., Scheier, M. F., and Weintraub, J. K. (1989). Assessing coping strategies: a theoretically based approach. *J. Pers. Soc. Psychol.* 56, 267–283. doi: 10.1037/0022-3514.56.2.267
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behav. Res.* 1, 245–276. doi: 10.1207/s15327906mbr0102_10
- Ciccone, A., and Ferrant, A. (2015). *Honte, Culpabilité et Traumatisme*. Malakoff: Dunod. doi: 10.3917/dunod.cicco.2015.01
- Clit, R. (2002). La terreur comme «passivité». *Topique* 81, 141–154. doi: 10.3917/top.081.0141
- Cloitre, M. (1998). "Intentional forgetting and clinical disorders," in *Intentional Forgetting: Interdisciplinary Approaches*, eds J. M. Golding and C. M. MacLeod (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 395–412.
- Cloitre, M., Koenen, K. C., Cohen, L. R., and Han, H. (2002). Skills training in affective and interpersonal regulation followed by exposure: a phase-based treatment for PTSD related to childhood abuse. *J. Consult. Clin. Psychol.* 70, 1067–1074. doi: 10.1037/0022-006X.70.5.1067
- Coleman, F. L. (1997). Stalking behavior and the cycle of domestic violence. *J. Interpers. Violence* 12, 420–432. doi: 10.1177/088626097012003007
- Contardi, A., Imperatori, C., Amati, I., Balsamo, M., and Innamorati, M. (2018). Assessment of affect lability: psychometric properties of the ALS-18. *Front. Psychol.* 9:427. doi: 10.3389/fpsyg.2018.00427
- Curran, P. J., West, S. G., and Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychol. Methods* 1, 16–29. doi: 10.1037/1082-989X.1.1.16
- Dalgleish, T., and Power, M. J. (2004). Emotion-specific and emotion-non-specific components of posttraumatic stress disorder (PTSD): implications for a taxonomy of related psychopathology. *Behav. Res. Ther.* 42, 1069–1088. doi: 10.1016/j.brat.2004.05.001
- De Luca Picione, R., Martino, M. L., and Freda, M. F. (2017). Understanding cancer patients' narratives: meaning-making process, temporality and modalities. *J. Constr. Psychol.* 30, 339–359. doi: 10.1080/10720537.2016.1227738
- De Luca Picione, R., Martino, M. L., and Freda, M. F. (2018). Modal articulation: the psychological and semiotic functions of modalities in the sensemaking process. *Theory Psychol.* 28, 84–103. doi: 10.1177/0959354317743580
- Derogatis, L. R., and Cleary, P. A. (1977). Confirmation of the dimensional structure of the SCL-90: a study in construct validation. *J. Clin. Psychol.* 33, 981–989. doi: 10.1186/1471-2288-11-114
- Diel, P. (1956). *La Peur et Langoisse: Phénomène Central de la Vie et de Son Évolution*, Vol. 116. Paris: Payot.
- Dobash, R. E., and Dobash, R. (1979). *Violence Against Wives: A Case Against the Patriarchy*. New York, NY: Free Press, 179–206.
- Dupuy, H. J. (1984). "The psychological general well-being (PGWB) index," in *Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies*, eds N. K. Wenger, M. E. Mattson, C. D. Furberg, and J. Elison (New York, NY: Le Jacq Publishing), 184–188.
- Ehrensaft, M. K. (2009). "Family and relationship predictors of psychological and physical aggression," in *Psychological and Physical Aggression in Couples: Causes and Interventions*, eds K. D. O'Leary and E. M. Woodin (Washington, DC, US: American Psychological Association), 99–118.
- Ellsberg, M., Jansen, H. A., Heise, L., Watts, C. H., and Garcia-Moreno, C. (2008). Intimate partner violence and women's physical and mental health in the WHO multi-country study on women's health and domestic violence: an observational study. *Lancet* 371, 1165–1172. doi: 10.1016/S0140-6736(08)60522-X
- European Union Agency for Fundamental Rights (2014). *Violence Against Women: An EU-Wide Survey*. Technical Report IUB-CS-TR61. Vienna: FRA.
- Ferraro, K. J. (1997). "Battered women: strategies for survival," in *Violence between Intimate Partners: Patterns, Causes, and Effects*, ed. A. Carderelli (New York, NY: Macmillan), 124–143.
- Flanagan, J. C. (1978). A research approach to improving our quality of life. *Am. Psychol.* 33, 138–147. doi: 10.1037/0003-066X.33.2.138
- Foà, C., Tonarelli, A., Caricati, L., and Fruggeri, L. (2015). COPE-NVI-25: validazione italiana della versione ridotta della coping orientation to the problems experienced (COPE-NVI). *Psicol. Salute* 2, 121–138.
- Follingstad, D. R., Wright, S., Lloyd, S., and Sebastian, J. A. (1991). Sex differences in motivations and effects in dating violence. *Fam. Relat.* 40, 51–57. doi: 10.2307/585658
- Friborg, O., Hjemdal, O., Rosenvinge, J. H., and Martinussen, M. (2003). A new rating scale for adult resilience: what are the central protective resources behind healthy adjustment? *Int. J. Methods Psychiatr. Res.* 12, 65–76. doi: 10.1002/mpr.143
- Frieze, I. H., and Browne, A. (1989). Violence in marriage. *Crime Justice* 11, 163–218. doi: 10.1086/449154
- Gargiulo, A., Plener, P. L., Baus, N., Margherita, G., Brunner, R., Kaess, M., et al. (2014). Autolesività non suicidaria (NSSI) e Disturbo da comportamento suicidario (SBD) nella recente pubblicazione del DSM-5. *Minerva Psichiatr.* 56, 83–90.
- Gilbert, P. (2000). The relationship of shame, social anxiety and depression: the role of the evaluation of social rank. *Clin. Psychol. Psychother.* 7, 174–189. doi: 10.1002/1099-0879(200007)7:3<174::AID-CPP236>3.0.CO;2-U
- Giromini, L., Velotti, P., de Campora, G., Bonalume, L., and Cesare Zavattini, G. (2012). Cultural adaptation of the difficulties in emotion regulation scale: reliability and validity of an Italian version. *J. Clin. Psychol.* 68, 989–1007. doi: 10.1002/jclp.21876
- Goss, K., Gilbert, P., and Allan, S. (1994). An exploration of shame measures—I: The other as Shamer scale. *Pers. Individ. Dif.* 17, 713–717. doi: 10.1016/0191-8869(94)90149-X
- Gratz, K. L., and Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. *J. Psychopathol. Behav. Assess.* 26, 41–54. doi: 10.1023/B:JOBA.0000007455.08539.94
- Green, A. (1973). *Le Discours Vivant: La Conception Psychanalytique de L'affect*, Vol. 2. Paris: Presses universitaires de France.
- Hagenaars, M. A., Oitzl, M., and Roelofs, K. (2014). Updating freeze: aligning animal and human research. *Neurosci. Biobehav. Rev.* 47, 165–176. doi: 10.1016/j.neubiorev.2014.07.021
- Halfon, S., Çaydar, A., Orsucci, F., Schiepek, G. K., Andreassi, S., Giuliani, A., et al. (2016). The non-linear trajectory of change in play profiles of three children in psychodynamic play therapy. *Front. Psychol.* 7:1494. doi: 10.3389/fpsyg.2016.01494
- Hamby, S. L., and Sugarman, D. B. (1999). Acts of psychological aggression against a partner and their relation to physical assault and gender. *J. Marriage Fam.* 959–970. doi: 10.2307/354016
- Harper, F. W., and Arias, I. (2004). The role of shame in predicting adult anger and depressive symptoms among victims of child psychological maltreatment. *J. Fam. Violence* 19, 359–367. doi: 10.1007/s10896-004-0681-x
- 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. Violence* 14, 399–415. doi: 10.1023/A:1022834215681
- Herman-Lewis, J. (1992). *Trauma and Recovery: The Aftermath of Violence-From Domestic Abuse to Political Terror*. New York, NY: Basic Books.
- Hilton, N. Z., Harris, G. T., Rice, M. E., Lang, C., Cormier, C. A., and Lines, K. J. (2004). A brief actuarial assessment for the prediction of wife assault recidivism: the Ontario domestic assault risk assessment. *Psychol. Assess.* 16, 267–275. doi: 10.1037/1040-3590.16.3.267
- Hirigoyen, M. F. (2005). *Femmes Sous Emprise: Les Ressorts de la Violence Dans le Couple*. OH! Éditions. London: Pocket.
- Horowitz, M., Wilner, N., and Alvarez, W. (1979). Impact of event scale: a measure of subjective stress. *Psychosom. Med.* 41, 209–218. doi: 10.1097/00006842-197905000-00004
- Hu, L. T., and Bentler, P. M. (1998). Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychol. Methods* 3, 424–453. doi: 10.1037/1082-989X.3.4.424

- 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: 10.1080/1070519909540118
- Hudson, W. W., and McIntosh, S. R. (1981). The assessment of spouse abuse: two quantifiable dimensions. *J. Marriage Fam.* 43, 873–888. doi: 10.2307/351344
- Hunt, N., and Evans, D. (2004). Predicting traumatic stress using emotional intelligence. *Behav. Res. Ther.* 42, 791–798. doi: 10.1016/j.brat.2003.07.009
- Imbasciati, A. (1991). *Affetto e Rappresentazione. Per una Psicoanalisi dei Processi Cognitive*. Milano: Franco Angeli.
- Innamorati, M., Tamburello, S., Contardi, A., Imperatori, C., Tamburello, A., Saggino, A., et al. (2013). Psychometric properties of the attitudes toward self-revised in Italian young adults. *Depress. Res. Treat.* 2013:209216. doi: 10.1155/2013/209216
- ISTAT (2015). *Dipartimento Per Le Pari Opportunità: La Violenza Contro Le Donne*. Rome: ISTAT.
- Jaquier, V., and Sullivan, T. P. (2014). Fear of past abusive partner (s) impacts current posttraumatic stress among women experiencing partner violence. *Violence Against Women* 20, 208–227. doi: 10.1177/1077801214525802
- Johnson, M. P. (1995). Patriarchal terrorism and common couple violence: two forms of violence against women. *J. Marriage Fam.* 57, 283–294. doi: 10.2307/353683
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* 20, 141–151. doi: 10.1177/001316446002000116
- 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
- Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Benjet, C., Bromet, E. J., Cardoso, G., et al. (2017). Trauma and PTSD in the WHO world mental health surveys. *Eur. J. Psychotraumatol.* 8:1353383. doi: 10.1080/20008198.2017.1353383
- Kilborne, B. (2002). *Disappearing Persons: Shame and Appearance*. Albany, NY: SUNY Press.
- Kilpatrick, D. G., Saunders, B. E., Amick-McMullan, A., Best, C. L., Veronen, L. J., and Resnick, H. S. (1989). Victim and crime factors associated with the development of crime-related post-traumatic stress disorder. *Behav. Ther.* 20, 199–214. doi: 10.1016/S0005-7894(89)80069-3
- Koss, M. P., Goodman, L. A., Browne, A., Fitzgerald, L. F., Keita, G. P., and Russo, N. F. (1994). *No Safe Haven: Male Violence against Women at Home, at Work, and in the Community*. Worcester, MA: American Psychological Association. doi: 10.1037/10156-000
- Kropp, P. R., and Hart, S. D. (2000). The spousal assault risk assessment (SARA) guide: reliability and validity in adult male offenders. *Law Hum. Behav.* 24, 101–118. doi: 10.1023/A:1005430904495
- Lawson, J. (2012). Sociological theories of intimate partner violence. *J. Hum. Behav. Soc. Environ.* 22, 572–590. doi: 10.1080/10911359.2011.598748
- Lee, D. A., Scragg, P., and Turner, S. (2001). The role of shame and guilt in traumatic events: a clinical model of shame-based and guilt-based PTSD. *Psychol. Psychother. Theory Res. Pract.* 74, 451–466. doi: 10.1348/000711201161109
- Levine, H. B. (2014). Psychoanalysis and trauma. *Psychoanal. Inq.* 34, 214–224. doi: 10.1080/07351690.2014.889475
- Lewis, M. (1995). *Shame: The Exposed Self*. New York, NY: Simon and Schuster.
- Litz, B. T., Orsillo, S. M., Kaloupek, D., and Weathers, F. (2000). Emotional processing in posttraumatic stress disorder. *J. Abnorm. Psychol.* 109, 26–39. doi: 10.1037/0021-843X.109.1.26
- MacCallum, R. C., and Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annu. Rev. Psychol.* 51, 201–226. doi: 10.1146/annurev.psych.51.1.201
- Margherita, G., and Troisi, G. (2014). Gender violence and shame. The visible and the invisible, from the clinical to the social systems. *La Camera Blu* 10, 166–185.
- Margherita, G., Troisi, G., and Nunziante Cesaro, A. (2014). “La vergogna e il femminile: del rendersi invisibile,” in *Chiaroscuri dell’identità. Sesso e Genere*, ed. A. Nunziante Cesaro (Roma: Franco Angeli).
- Marmar, C. R., Weiss, D. S., and Metzler, T. J. (1997). The peritraumatic dissociative experiences questionnaire. *Assess. Psychol. Trauma PTSD* 2, 144–168.
- Mihalic, S. W., and Elliott, D. (1997). A social learning theory model of marital violence. *J. Fam. Violence* 12, 21–47. doi: 10.1023/A:1021941816102
- Morrison, A. P. (1999). Shame, on either side of defense. *Contemp. Psychoanal.* 35, 91–105. doi: 10.1016/j.lpm.2014.06.037
- Murphy, C. M., and O’Leary, K. D. (1989). Psychological aggression predicts physical aggression in early marriage. *J. Consult. Clin. Psychol.* 57, 579–582. doi: 10.1037/0022-006X.57.5.579
- Muthén, L. K., and Muthén, B. O. (1998). *Mplus: The Comprehensive Modeling Program for Applied Researchers; User’s Guide; [Version 1.0]*. Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., and Muthén, B. O. (2004). *Mplus User’s Guide: Statistical Analysis with Latent Variables: User’s Guide*. Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., and Muthén, B. O. (2012). *Mplus Statistical Modeling Software: Release 7.0*. Los Angeles, CA: Muthén & Muthén.
- Nunziante Cesaro, A., and Troisi, G. (2016). Le ferite della violenza tra paura e terrore. *La Camera Blu. Rivista di Studi di Genere* 14, 90–112.
- O’Keefe, M. (1998). Factors mediating the link between witnessing interparental violence and dating violence. *J. Fam. Violence* 13, 39–57. doi: 10.1023/A:1022860700118
- Orsillo, S. M., Batten, S. V., Plumb, J. C., Luterek, J. A., and Roessner, B. M. (2004). An experimental study of emotional responding in women with posttraumatic stress disorder related to interpersonal violence. *J. Trauma Stress* 17, 241–248. doi: 10.1023/B:JOTS.0000029267.61240.94
- Pandolfi, A. M. (2002). *La Vergogna. Un Affetto Psicico Che Sta Scomparendo*. Milano: Franco Angeli.
- Pence, E., and Paymar, M. (1993). *Education Groups for Men Who Batter: The Duluth Model*. New York, NY: Springer Publishing Company. doi: 10.1891/9780826179913
- Pico-Alfonso, M. A., Echeburúa, E., and Martinez, M. (2008). Personality disorder symptoms in women as a result of chronic intimate male partner violence. *J. Fam. Violence* 23, 577–588. doi: 10.1007/s10896-008-9180-9
- Pietrantonio, F., De Gennaro, L., Di Paolo, M. C., and Solano, L. (2003). The impact of event scale: validation of an Italian version. *J. Psychosom. Res.* 55, 389–393. doi: 10.1016/S0022-3999(02)00638-4
- Piotrkowski, C. S., and Brannen, S. J. (2002). Exposure, threat appraisal, and lost confidence as predictors of PTSD symptoms following September 11, 2001. *Am. J. Orthopsychiatry* 72, 476–485. doi: 10.1037/0002-9432.72.4.476
- Rabin, R. F., Jennings, J. M., Campbell, J. C., and Bair-Merritt, M. H. (2009). Intimate partner violence screening tools: a systematic review. *Am. J. Prev. Med.* 36, 439–445. doi: 10.1016/j.amepre.2009.01.024
- Rasch, G. (1960). *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests*. Oxford: Nielsen & Lydiche.
- Reale, E. (2011). *Maltrattamento e Violenza Sulle Donne*. Milano: Franco Angeli.
- Rennison, C. M., and Welchans, S. (2000). *Intimate Partner Violence*. Washington, DC: U.S. Department of Justice. doi: 10.1037/e378132004-001
- Resick, P. A., and Miller, M. W. (2009). Posttraumatic stress disorder: anxiety or traumatic stress disorder? *J. Trauma Stress* 22, 384–390. doi: 10.1002/jts.20437
- Resnick, H. S., Kilpatrick, D. G., Dansky, B. S., Saunders, B. E., and Best, C. L. (1993). Prevalence of civilian trauma and posttraumatic stress disorder in a representative national sample of women. *J. Consult. Clin. Psychol.* 61, 984–991. doi: 10.1037/0022-006X.61.6.984
- Riggs, D. S., Kilpatrick, D. G., and Resnick, H. S. (1992). Long-term psychological distress associated with marital rape and aggravated assault: a comparison to other crime victims. *J. Fam. Violence* 7, 283–296. doi: 10.1007/BF00994619
- Riggs, D. S., and O’Leary, K. D. (1996). Aggression between heterosexual dating partners: an examination of a causal model of courtship aggression. *J. Interpers. Violence* 11, 519–540. doi: 10.1177/088626096011004005
- Romito, P. (2005). *Un Silenzio Assordante: la Violenza Occultata su Donne e Minori*, Vol. 36. Milano: Franco Angeli.
- Ronfeldt, H. M., Kimerling, R., and Arias, I. (1998). Satisfaction with relationship power and the perpetration of dating violence. *J. Marriage Fam.* 70–78. doi: 10.2307/353442
- Rubinstein, M. (1987). *The Dignity of Women at Work: Report on the Problem of Sexual Harassment in the Member States of the European Community*. Brussels: Publication Office of the European Community.
- Saggino, A., Balsamo, M., Carlucci, L., Cavalletti, V., Sergi, M. R., da Fermo, G., et al. (2018). Psychometric properties of the Italian version of the young schema questionnaire I-3: Preliminary results. *Front. Psychol.* 9:312. doi: 10.3389/fpsyg.2018.00312

- Saggino, A., Carlucci, L., Sergi, M. R., D'Ambrosio, I., Fairfield, B., and Balsamo, M. (2017). A validation study of the psychometric properties of the other as Shamer scale-2. *SAGE Open* 7, 1–10. doi: 10.1177/0706743716673320
- Scheffer Lindgren, M., and Renck, B. (2008). 'It is still so deep-seated, the fear': psychological stress reactions as consequences of intimate partner violence. *J. Psychiatr. Ment. Health Nurs.* 15, 219–228. doi: 10.1111/j.1365-2850.2007.01215.x
- 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.
- Shook, N. J., Gerrity, D. A., Jurich, J., and Segrist, A. E. (2000). Courtship violence among college students: a comparison of verbally and physically abusive couples. *J. Fam. Violence* 15, 1–22. doi: 10.1023/A:1007532718917
- Shorey, R. C., Sherman, A. E., Kivisto, A. J., Elkins, S. R., Rhatigan, D. L., and Moore, T. M. (2011). Gender differences in depression and anxiety among victims of intimate partner violence: the moderating effect of shame proneness. *J. Interpers. Violence* 26, 1834–1850. doi: 10.1177/0886260510372949
- Sica, C., Magni, C., Ghisi, M., Altoè, G., Sighinolfi, C., Chiri, L. R., et al. (2008). Coping orientation to problems experienced-nuova versione Italiana (COPE-NVI): uno strumento per la misura degli stili di coping. *Psicoter. Cogn. Comport.* 14, 27–53.
- Sippel, L. M., and Marshall, A. D. (2011). Posttraumatic stress disorder symptoms, intimate partner violence perpetration, and the mediating role of shame processing bias. *J. Anxiety Disord.* 25, 903–910. doi: 10.1016/j.janxdis.2011.05.002
- Smith, P. H., Tessaro, I., and Earp, J. A. L. (1995). Women's experiences with battering: a conceptualization from qualitative research. *Womens Health Issues* 5, 173–182. doi: 10.1016/1049-3867(95)00615-X
- Straus, M. A. (1976). Sexual inequality, cultural norms, and wife-beating. *Victimology* 1, 54–70.
- Straus, M. A. (2017). "Measuring intrafamily conflict and violence: the conflict tactics (CT) scales," in *Physical Violence in American Families*, eds M. A. Straus, and R. J. Gelles (Abingdon: Routledge), 29–48.
- Straus, M. A., and Gelles, R. J. (1990). How violent are American families? Estimates from the national family violence resurvey and other studies. *Phys. Violence Am. Fam.* 8, 95–112.
- Tabachnick, B. G., and Fidell, L. S. (2007). *Using Multivariate Statistics*. Boston, MA: Allyn & Bacon.
- Tangney, J. P. (1996). Conceptual and methodological issues in the assessment of shame and guilt. *Behav. Res. Ther.* 34, 741–754. doi: 10.1016/0005-7967(96)00034-4
- Tangney, J. P., and Dearing, R. L. (2002). *Emotions and Social Behavior. Shame and Guilt*. New York, NY: Guilford.
- Tisseron, S. (1992). *La Honte. Psychanalyse D'un Lien Social*. Paris: Dunod.
- Tjaden, P., and Thoennes, N. (1998). *Prevalence, Incidence, and Consequences of Violence against Women: Findings from the National Violence against Women Survey*. Washington, DC: National Institute of Justice and the Centers for Disease Control and Prevention.
- Tjaden, P. G., and Thoennes, N. (2000). *Extent, Nature, and Consequences of Intimate Partner Violence*. Washington, DC: National Institute of Justice. doi: 10.1037/e300342003-001
- Troisi, G. (2017). *Gli Affetti e la Violenza: Uno Studio Sull'intimate Partner Violence*. Ph.D. thesis, University of Naples Federico II, Naples.
- Van der Kolk, B. A. (1996). "The complexity of adaptation to trauma: self-regulation, stimulus discrimination, and characterological development," in *Traumatic Stress: The Effects of Overwhelming Experience on Mind, Body, and Society*, eds B. A. van der Kolk, A. C. McFarlane, and L. Weisaeth (New York, NY: Guilford Press), 182–213.
- Veazey, C. H., Blanchard, E. B., Hickling, E. J., and Buckley, T. C. (2004). Physiological responsiveness of motor vehicle accident survivors with chronic posttraumatic stress disorder. *Appl. Psychophysiol. Biofeedback* 29, 51–62. doi: 10.1023/B:APBI.0000017863.35714.a1
- Walker, L. E. (1977). Battered women and learned helplessness. *Victimology* 2, 525–534.
- Weathers, F., Huska, J., and Keane, T. (1991). *The PTSD Checklist Military Version (PCL-M)*. Boston, MA: National Center for PTSD, 42.
- Weaver, T. L., and Clum, G. A. (1995). Psychological distress associated with interpersonal violence: a meta-analysis. *Clin. Psychol. Rev.* 15, 115–140. doi: 10.1016/0272-7358(95)00004-9
- Widaman, K. F. (1993). Common factor analysis versus principal component analysis: differential bias in representing model parameters? *Multivariate Behav. Res.* 28, 263–311. doi: 10.1207/s15327906mbr2803_1
- Wilson, J. P., Drozdek, B., and Turkovic, S. (2006). Posttraumatic shame and guilt. *Trauma Violence Abuse* 7, 122–141. doi: 10.1177/1524838005285914
- World Health Organization, Department of Injuries and Violence Prevention (2002). *The Injury Chart Book: A Graphical Overview of the Global Burden of Injuries*. Geneva: World Health Organization.
- World Health Organization [WHO] (2013). *Responding to Intimate Partner Violence and Sexual Violence against Women: WHO Clinical and Policy Guidelines*. Geneva: World Health Organization.
- Yegidis, B. L. (1989). *Abuse Risk Inventory for Women: Manual*. Palo Alto, CA: Consulting Psychologists Press.

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

Copyright © 2018 Troisi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Reliability of the DEM Test in the Clinical Environment

Alessio Facchin^{1,2,3,4*} and Silvio Maffioletti^{4,5}

¹ Department of Psychology, University of Milano-Bicocca, Milan, Italy, ² COMIB – Research Center in Optics and Optometry, University of Milano-Bicocca, Milan, Italy, ³ NeuroMi - Milan Center for Neuroscience, Milan, Italy, ⁴ IRSOO – Institute for Research and Studies in Optics and Optometry, Vinci, Italy, ⁵ Optics and Optometry, University of Turin, Turin, Italy

OPEN ACCESS

Edited by:

Michela Balsamo,
Università degli Studi G. d'Annunzio
Chieti e Pescara, Italy

Reviewed by:

Cesar Merino-Soto,
Universidad de San Martín de Porres,
Peru
Andrea Spoto,
Università degli Studi di Padova, Italy

*Correspondence:

Alessio Facchin
alessiopietro.facchin@gmail.com

Specialty section:

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

Received: 23 February 2018

Accepted: 04 July 2018

Published: 25 July 2018

Citation:

Facchin A and Maffioletti S (2018) The
Reliability of the DEM Test
in the Clinical Environment.
Front. Psychol. 9:1279.
doi: 10.3389/fpsyg.2018.01279

The developmental eye movement (DEM) test is a practical and simple method for assessing and quantifying ocular motor skills in children. Different studies have previously assessed the reliability of the DEM test and they have generally found high values for vertical and horizontal time, whereas those for Ratio and Errors were medium and low, respectively. In the second application of test were found an improvement in performance in all subtests. Our aim was to evaluate the reliability of the DEM test using seconds and percentile scoring and looking in depth at the improvement in performance when the test is repeated. We tested the reliability of the DEM test on a group of 115 children from the 2nd to the 5th grade using different statistical methods: correlations, ANOVA, limits of agreement for results expressed in seconds and as percentile scoring and pass-fail diagnostic classification. We found high reliability with excellent values for vertical and adjusted horizontal time, medium-to-high for ratio and medium for errors. We have re-confirmed the presence of a significant improvement of performance on the second session for vertical time, horizontal time and ratio. The stability of binary classification of Pass-Fail criteria appears to be medium. We found high reliability for the DEM test when compared with the published results of other research but the improvement of performance, the learning effect was still present, but at a lower level than previously found. With the awareness of these limitations the DEM test can be used in clinical practice in evaluating performance over time.

Keywords: DEM test, reliability, test-retest, learning effect, psychometrics, clinical assessment

INTRODUCTION

The developmental eye movement (DEM) test is a practical and simple method for assessing and quantifying ocular motor skills in children. The DEM test allows clinicians interested in vision to obtain an easy quantitative measurement of ocular-movement skills by means of a psychometric test. The task consists of naming numbers in a simulated reading-like condition (Garzia et al., 1990).

The DEM test comprises three different individual plates. Two plates contain regularly spaced numbers, each displaced in two different columns for vertical reading (Card A and Card B). These determine the automaticity of number-naming ability. The third plate contains unevenly spaced numbers, displaced in sixteen different lines for horizontal reading (Card C). This evaluates number naming in a reading-like task. The ratio score is calculated by dividing the adjusted horizontal time, corrected for errors, by the vertical time. The vertical time, adjusted horizontal time, ratio, and error scores are compared with the published normative dataset and used to

identify dysfunctions related to either number naming, ocular motor skills, or a combination of the two.

The choice of a psychometric test such as DEM is determined by considering the three factors that characterize its properties: validity, reliability, and normative values (Anastasi and Urbina, 1997; Facchin et al., 2011). The validity of the DEM test in assessing ocular movement has been the subject of some discussion (Medland et al., 2010; Webber et al., 2011). Some studies concluded that DEM did not measure ocular movements (Ayton et al., 2009). Conversely, others studies have evaluated the validity of the DEM test (Garzia et al., 1990; Facchin et al., 2011) and, although it did not seem to correlate directly with pure eye movement parameters, it was related with different aspects of reading performance and it is useful in clinical practice (Powers et al., 2008; Ayton et al., 2009; Palomo-Álvarez and Puell, 2009). Even where there are influences from many cognitive processes, such as sustained attention (Coulter and Shallo-Hoffmann, 2000), number recognition and retrieval, visual verbal integration time, speaking time and visuo-spatial attention (Facchin et al., 2011), the DEM test provided the potential to measure visual skills related to ocular movements in a reading-like condition. In actual fact, normative values are available for the English (Richman and Garzia, 1987), Spanish (Fernandez-Velazquez and Fernandez-Fidalgo, 1995; Jimenez et al., 2003), Cantonese (Pang et al., 2010), Japanese (Okumura and Wakamiya, 2010), Portuguese (Baptista et al., 2011), Italian (Facchin et al., 2012), Mandarin (Xie et al., 2016), and Latvian (Serdjukova et al., 2016) languages.

Test-retest reliability means that a test should produce the same score for each subject when it is performed twice without apparent changes in the variable measured (Urbina, 2004; Kline, 2014). As applied to the DEM test, reliability was tested several times over periods of years. The test manuals (Richman and Garzia, 1987; Richman, 2009) reports that reliability was tested on forty subjects from grades one through seven and gives the following correlation coefficients (Pearson r): for vertical time, $r = 0.89$, $p < 0.001$; for adjusted horizontal time, $r = 0.86$, $p < 0.01$; for ratio, $r = 0.57$, $p < 0.05$; for errors, $r = 0.07$ n.s. Taken together, these data show that the DEM test has good reliability (test-retest correlation) for vertical and horizontal time, but medium for ratio, and low for errors. Santiago and Perez (1992) have replicated these results, finding only a higher value for errors.

Rouse et al. (2004) tested a group of 30 3rd grade children, and retested them 2 weeks later. They found that vertical and adjusted horizontal time both have fair to good repeatability, whereas that for the ratio score was found to be poor. It is necessary to take into account that a single classroom was used in the study and not a stratified sample. Interestingly, in this study, it was introduced the concept of limits of agreement with a corresponding graphical representation (Altman and Bland, 1983; Bland and Altman, 1986).

Tassinari and DeLand (2005) tested two groups, in office and in school environments. The correlation coefficients were higher than those previously found and, remarkably, good agreement was reported between test and retest in terms of pass-fail classification only for the office group.

Orlansky et al. (2011) performed a more extensive evaluation of reliability in a multi-center study. More than 180 subjects

were tested in two sessions, in each of which they were each evaluated three times. The most important results are the fair to good correlation coefficients between-session for both the vertical and horizontal scores and the poor results for the ratio and error scores. Regarding pass-fail classification, the proportion of subjects who stayed in the same classification was in the range from 71 to 100% for both vertical and horizontal scores. For ratio and error scores, the proportion of subjects that remained classified as pass or fail was between 47 and 100%. However, they found that children in this age range could show improvements in all four test scores without any intervention. Finally, it was concluded that clinicians should be careful about using the DEM test for diagnosis or to monitor the effectiveness of treatment. The pass/fail analyses were performed based on two cut-offs at the 16th and 30th percentiles. The researchers administered three parallel versions of the DEM test (the same 80 numbers in different sequences) in order to eliminate implicit or explicit memorization of the numbers. In a clinical setting it is impossible to use parallel versions because the original test was not designed to have such forms. Indeed, from a theoretical point of view, parallel forms seem plausible and the normative data appear to be equally valid.

In the last case, the parallel form of test reliability was in fact evaluated, but it did not represent the true test-retest reliability of a single version of the clinical test. Moreover, unlike manual instructions, the vertical time for errors was also corrected; when the original manual (and the large part of norms) did not require this correction to be performed (the scoresheet in the 1987 manual reported this calculation incorrectly). Again, the multiple repetition of test within each session could affect the true between session test-retest reliability.

In the studies mentioned previously, the general term reliability has incorporated concepts and scores derived from the agreement term. The border between the concepts of reliability and agreement may not always be clear (Costa Santos et al., 2011a,b), and for this reason we discuss reliability and agreement separately.

Broadly speaking, from a pure psychometric point of view, the reliability is the correlation coefficient between test and retest (Anastasi and Urbina, 1997; Urbina, 2004). On the other hand, it provides information regarding the ability of the score to distinguish between subjects (Kottner and Streiner, 2011). The DEM test shows a high reliability, with the exception of ratio which shows a medium to high relationship. Correlation refers to the linear relationship with two sessions of administration, but it can provide nothing regarding the changes with respect to the absolute score. In fact, this concept was better explained by the agreement term, which represents the similarity of scores, and judgment or diagnoses with respect to the degree in which they differ (Kottner and Streiner, 2011). Rouse et al. (2004) and Orlansky et al. (2011) have shown that the true problem with the DEM test appears to be the improvement between sessions, which can be defined as a form of lack of agreement. This improvement was also defined as learning effect (Orlansky et al., 2011) and reported in terms of mean change and its respective limits of agreement (Altman and Bland, 1983).

Based on the aforementioned considerations, when compared with the study by Orlansky et al. (2011), using a single test, we predict an equal or higher reliability, but a low agreement expressed with a high learning effect (high bias and wider limits of agreements). Different comparisons were performed with all other reliability studies in order to assess and compare reliability and agreement.

Consequently, in performing the present study we have three aims. Firstly, we wanted to test the reliability, quantify the learning effect and assess the agreement between sessions using only one established classification criterion and only one version of the test as used in clinical practice. Secondly, from a clinical and rehabilitation point of view, because DEM scores have previously been observed to improve between sessions in absence of intervention, we wanted to calculate the minimum amount of change that needs to be observed to consider the change a real change using percentile score. Thirdly, considering the recent needs of replication studies (Open Science Collaboration, 2015), we wanted to replicate the results of previous studies on DEM reliability involving a different population and norms.

MATERIALS AND METHODS

Subjects

Children were taken from a school screening program performed in the “V.Muzio” public school in Bergamo, north of Italy. Only children with written informed consent from their parents to take part in the study were enrolled (Facchin et al., 2011, 2012). All participants were selected on the basis of the following criteria: they were required to use their glasses or contact lenses (if required) during testing; to have a monocular visual acuity at distance of at least 0.63 decimal (20/32 with Goodlite n. 735000 table), to have a near binocular visual acuity of at least 0.8 decimal (20/25 with Goodlite n. 250800 table); and not to present binocular anomalies (strabismus) at cover test and distance and near phoria in a normal range (± 4 at distance and ± 6 at near) measured with a Thorington technique (Rainey et al., 1998; Scheiman and Wick, 2013). Testing was performed in two sessions. Subjects who performed in only one session were excluded. 135 children from two primary schools in the north of Italy were screened, but only 115 met the required inclusion criteria (three participants were excluded for strabismus, eight for lower monocular distance visual acuity, nine for the absence of second session test; see **Table 1** for details of the final participants). The study was carried out in accordance with the guidelines given in the Declaration of Helsinki and the school council of the “V.Muzio” school approved the procedure.

Tests and Procedures

A short description of tests and procedures is given below.

Four cards comprise the DEM test: the pretest card, two vertical cards (A and B) and one horizontal card (C). The test was administered using the methodology given in the DEM manual. The vertical time represents the sum of that spent on naming the number printed on the two cards, A and B. The vertical time returns the time required to read 80 numbers organized vertically.

The adjusted horizontal time represents the time required for card C corrected for omission or addition errors. The adjusted horizontal time reflects the time required to read the 80 numbers organized in a horizontal pattern, together with that needed to perform saccadic movements. Dividing the adjusted horizontal by the vertical time, the ratio score was calculated. This is used to assess ocular motility dysfunction. The total number of errors returns the accuracy of reading of card C. Italian normative tables (Facchin et al., 2012) were used to determine the percentile score for vertical time, adjusted horizontal time, ratio and error.

The DEM test was administered as reported in the manual on an inclined reading desk set at 40 cm, with constant illumination and without noise. The tests were administered in two different sessions, separated by between 14 and 20 days, in the same room, for every subject who completed the test in the first session.

Statistical Methods

We have analyzed all aspects of test–retest reliability and agreement between the two measurements as a function of time. Wherever possible, our data were compared with the results obtained in other published research. In order to look at the results from a meaningful clinical viewpoint, additional analyses were applied using percentile scoring.

Firstly, because previous studies used three different correlation indexes (Richman and Garzia, 1987; Rouse et al., 2004; Tassinari and DeLand, 2005; Richman, 2009; Facchin et al., 2011; Orlansky et al., 2011) in order to perform inter-study comparison, the test–retest reliability for DEM was analyzed using: Pearson *r* correlation, partial correlation (adjusted for age) and intra class correlation (ICC) using the average score and One-Way model (McGraw and Wong, 1996). Confidence intervals for correlations were calculated following a specific procedure (Zou, 2007; Diedenhofen and Musch, 2015), and ICC and Cohen’s *K* difference were also calculated and reported using a specific methodology (Dormer and Zou, 2002; Ramasundarahettige et al., 2009).

Because Orlansky et al. (2011) have performed the test–retest evaluation with three repetitions in each session (30–90'' distance) in two sessions (1–4 weeks apart), from this study, only the first administration of each session was taken into account for comparison of correlation coefficients. According to the study

TABLE 1 | Sample description subdivided by grade and gender (M, male; F, female).

Grade	Mean age	Gender	<i>n</i>
2nd	7 years and 6 months	M	16
		F	9
3rd	8 years and 5 months	M	14
		F	15
4th	9 years and 10 months	M	22
		F	10
5th	10 years and 7 months	M	12
		F	17
Total			115

of Fleiss and Cohen (1973) and the study of Viera and Garrett (2005), interpretation of correlation coefficients, Kappa and AC₁ was based on five steps each of 0.2 points between 0 and 1 with the respective subdivision: low, low to moderate, moderate, moderate to high and high.

Secondly, in order to test the agreement, we calculated and plotted the Bland – Altman 95% limits of agreement (LoA; 1.96 * SD) that gives the value and the range of differences between the test and re-test scores (Bland and Altman, 1986). If the test is truly reliable, differences outside of LoA limits have only 5% of occurrence. These limits have an error margin and consequently their respective confidence intervals (95% CI) were calculated. With these data expressed in seconds and in percentiles we can estimate the minimum change necessary in the second session to have a statistical confirmation of amelioration over two sessions of administration is due to a treatment and not to lack of agreement. In order to evaluate the mean bias between sessions, a repeated measure ANOVA was applied to each specific subtest.

To quantify the magnitude of the improvement over time, we proposed a simple index of learning effect, adapted to reliability. This index was calculated for each DEM subtest and can be summarized as:

$$Learning\ Effect\ (\%) = 100 * \frac{ReTest\ Mean - Test\ Mean}{Test\ Mean}$$

where,

ReTest Mean = the mean value of all subjects in the second session,

Test Mean = the mean value for all subjects in the first.

The learning effect can give us an absolute mean percentage of improvement (in seconds). For clinical use, it is better to know the same effect scored in percentile in order to determine whether there is a significant amelioration over time. Finally, a standard error of measurement expressed as the standard deviation of errors of measurement that are associated with test reliability was

calculated using the formula (Rouse et al., 2004):

$$Se_m = SD\sqrt{1 - r_{xx}}$$

where,

Se_m = standard error of measurement,

SD = standard deviation,

R_{xx} = reliability of the test.

Thirdly, in order to evaluate and compare the agreement between sessions of the DEM test classification using pass–fail cut-off criteria, the Cohen’s Kappa (Fleiss et al., 1969) and the AC₁ index (Gwet, 2008) were applied. Kappa was selected for the comparison of studies and AC₁ was applied in order to avoid the paradoxical results found using Kappa index (Gwet, 2008). Before calculating Kappa and AC₁, for each subject, a percentile scoring through DEM test specific Italian norms were calculated. In previous studies and in the manual, two criteria were used. The first refers to the first edition of manual (version 1/1987, 30th percentile criterion), whilst the second refers to the new edition (version 2; 2009, 16th percentile criterion). In order to be aligned with other Italian national psychoeducational criteria used in the cognitive evaluation of children, we applied the cutoff at the 16th percentile (Associazione Italiana Dislessia, 2007). If vertical time, adjusted horizontal time, ratio and errors presented a score that was equal or below the 16th percentile, it was marked as “fail.” If the score was above the 16th percentile, it was marked as “pass.” Data were analyzed using R statistical environment and specific packages (R Core Team, 2017).

RESULTS

Reliability

The different correlation coefficients for test–retest reliability were determined and these are listed in **Table 2**. The results

TABLE 2 | Correlations, partial correlations and intra-class correlation coefficients (ICC) with relative 95% confidence intervals for the four DEM subtests.

	Pearson correlations		Partial correlations		Intraclass correlations	
	<i>r</i> (95% CI)	<i>p</i>	<i>r_p</i> (95% CI)	<i>p</i>	ICC (95% CI)	<i>p</i>
VT	0.933 (0.905 to 0.953)	<0.0001	0.902 (0.861 to 0.930)	<0.0001	0.932 (0.904 to 0.953)	<0.0001
AHT	0.901 (0.860 to 0.931)	<0.0001	0.816 (0.742 to 0.873)	<0.0001	0.892 (0.848 to 0.924)	<0.0001
Ratio	0.668 (0.552 to 0.758)	<0.0001	0.597 (0.459 to 0.701)	<0.0001	0.649 (0.529 to 0.743)	<0.0001
Errors	0.692 (0.582 to 0.776)	<0.0001	0.643 (0.524 to 0.742)	<0.0001	0.692 (0.583 to 0.776)	<0.0001

VT, vertical time; AHT, adjusted horizontal time.

TABLE 3 | Pearson correlation coefficients comparison between this study and those of Rouse et al. (2004) and Richman (2009).

	This work	DEM manual	Difference (95% CI)	<i>p</i>	Rouse et al. (2004)	Difference (95% CI)	<i>p</i>
Subjects (<i>n</i>)	115	40			30		
VT	0.93	0.89	0.04 (–0.019 to 0.132)	n.s.	0.60	0.33 (0.138 to 0.625)	<0.001
AHT	0.90	0.86	0.04 (0.036 to 0.155)	n.s.	0.55	0.35 (0.136 to 0.665)	<0.0001
Ratio	0.66	0.57	0.09 (–0.124 to 0.362)	n.s.	0.27	0.39 (0.064 to 0.772)	<0.05
Errors	0.69	0.07	0.62 (0.297 to 0.948)	<0.01			

VT, vertical time; AHT, adjusted horizontal time.

TABLE 4 | ICC correlation coefficients comparison between this study and those of Tassinari and DeLand (2005) and Orlansky et al. (2011).

	This work	Tassinari "Office"	Difference (95% CI)	p	Tassinari "School"	Difference (95% CI)	p	Orlansky	Difference (95% CI)	p
n	115	53			13			181		
VT	0.932	0.96	0.028 (-0.013 to 0.059)	n.s.	0.85	-0.082 (-0.356 to 0.023)	n.s.	0.83	-0.102 (-0.236 to -0.019)	<0.0001
AHT	0.892	0.92	0.028 (-0.032 to 0.084)	n.s.	0.89	-0.002 (-0.216 to 0.085)	n.s.	0.78	-0.112 (-0.291 to -0.005)	<0.005
Ratio	0.649	0.76	0.111 (-0.062 to 0.263)	n.s.	0.59	-0.059 (-0.582 to 0.229)	n.s.	0.49	-0.159 (-0.462 to 0.085)	0.05
Errors	0.692	0.79	0.098 (-0.058 to 0.234)	n.s.	0.44	-0.252 (-0.7 to 0.112)	n.s.	0.39	-0.306 (-0.585 to -0.042)	<0.0005

VT, vertical time; AHT, adjusted horizontal time.

TABLE 5 | Limits of Agreement for the DEM subtest stratified by grades.

Grade	Subtest	Lower limit (95% CI)	Mean difference (95% CI)	Upper limit (95% CI)
2	VT	-9.28 (-11.92 to -6.64)	-2.04 (-3.57 to -0.52)	5.2 (2.56 to 7.84)
2	AHT	-30.05 (-38.24 to -21.86)	-7.59 (-12.32 to -2.86)	14.87 (6.68 to 23.07)
2	Ratio	-0.52 (-0.67 to -0.36)	-0.1 (-0.19 to -0.01)	0.32 (0.17 to 0.48)
2	Errors	-14.96 (-21.36 to -8.55)	2.6 (-1.1 to 6.3)	20.16 (13.75 to 26.56)
3	VT	-9.06 (-11.42 to -6.7)	-2.04 (-3.4 to -0.68)	4.98 (2.62 to 7.34)
3	AHT	-18.75 (-22.98 to -14.52)	-6.17 (-8.61 to -3.73)	6.42 (2.19 to 10.65)
3	Ratio	-0.34 (-0.43 to -0.25)	-0.08 (-0.13 to -0.03)	0.19 (0.1 to 0.27)
3	Errors	-14.55 (-18.83 to -10.27)	-1.83 (-4.3 to 0.64)	10.9 (6.62 to 15.17)
4	VT	-8.2 (-10.69 to -5.72)	-0.39 (-1.83 to 1.04)	7.42 (4.93 to 9.91)
4	AHT	-14.61 (-18.25 to -10.96)	-3.17 (-5.27 to -1.07)	8.27 (4.62 to 11.91)
4	Ratio	-0.34 (-0.42 to -0.25)	-0.06 (-0.11 to -0.01)	0.21 (0.12 to 0.3)
4	Errors	-7.84 (-10.37 to -5.3)	0.12 (-1.34 to 1.59)	8.09 (5.55 to 10.62)
5	VT	-10.19 (-12.92 to -7.46)	-2.06 (-3.64 to -0.48)	6.07 (3.34 to 8.81)
5	AHT	-12.48 (-15.07 to -9.9)	-4.79 (-6.29 to -3.3)	2.89 (0.31 to 5.48)
5	Ratio	-0.25 (-0.3 to -0.19)	-0.07 (-0.11 to -0.04)	0.1 (0.04 to 0.16)
5	Errors	-10.59 (-13.48 to -7.7)	-2 (-3.67 to -0.33)	6.59 (3.7 to 9.48)

In each column the lower, mean and upper limits of agreement are reported, together with their specific ±95% CI intervals. (The units of these raw data are seconds for vertical time and adjusted horizontal time and number for errors). VT, vertical time; AHT, adjusted horizontal time.

show high values for vertical time and adjusted horizontal time, and moderate to high for ratio and errors. This pattern was confirmed by partial correlation when the component due to age was removed. The ICC correlations also confirmed the good repeatability of all variables. Moreover, the confidence intervals are very small and the values vary from medium-high to high.

The different studies on the repeatability of the DEM test used different correlation coefficients. To enable comparison, in the case of the studies of Richman and Garzia (1987) and Rouse et al. (2004), the evaluation was performed with the Pearson correlation coefficient, and for the study of Tassinari and DeLand (2005) and Orlansky et al. (2011) using the ICC (Tables 3, 4)

Independent of the correlation used the results of the present study show significantly higher repeatability compared with other studies. Only with the Tassinari "school" group are there no significant differences, and the higher number of subjects involved in the present study confirms the previous result.

Agreement

An efficient way to verify the agreement is to use the Bland and Altman limits of agreement graphical analysis and its associated statistics (Bland and Altman, 1986). In Table 5, we have listed the limits of agreement with the 95% upper and lower limits, with the 95% confidence limits.

Because the limits of agreement calculation could also be performed with transformed data (Giavarina, 2015), we carried out these analyses with percentiles. The results are listed in Table 6 and shown in Figure 1.

Another way to view the bias between sessions is to observe the mean and SD for vertical time, adjusted horizontal time, and ratio score for each age group are listed in Table 7 and presented in Figure 2.

Apart from the errors in grades 2 and 4, there is an evident improvement in performance on the second administration of the test. In order to verify this improvement, a series of ANOVA for each DEM subtest was performed. ANOVA was performed with one factor within (Time, with two levels), and one factor between (Grade, with four levels). The results for vertical

TABLE 6 | Limits of Agreement for the DEM subtest expressed in percentile score.

	Lower limit (95% CI)	Mean difference (95% CI)	Upper limit (95% CI)
VT	-28.45 (-33.9 to -23)	4.93 (1.78 to 8.08)	38.31 (32.86 to 43.76)
AHT	-21.11 (-26.85 to -15.37)	14.05 (10.74 to 17.37)	49.22 (43.48 to 54.96)
Ratio	-35.44 (-43.64 to -27.23)	14.81 (10.07 to 19.54)	65.05 (56.85 to 73.26)
Errors	-67.12 (-78.53 to -55.72)	2.75 (-3.84 to 9.33)	72.62 (61.21 to 84.02)

In each column the lower, mean and upper limits of agreement are reported, together with their specific $\pm 95\%$ CI intervals. (The values are listed as percentiles). VT, vertical time; AHT, adjusted horizontal time.

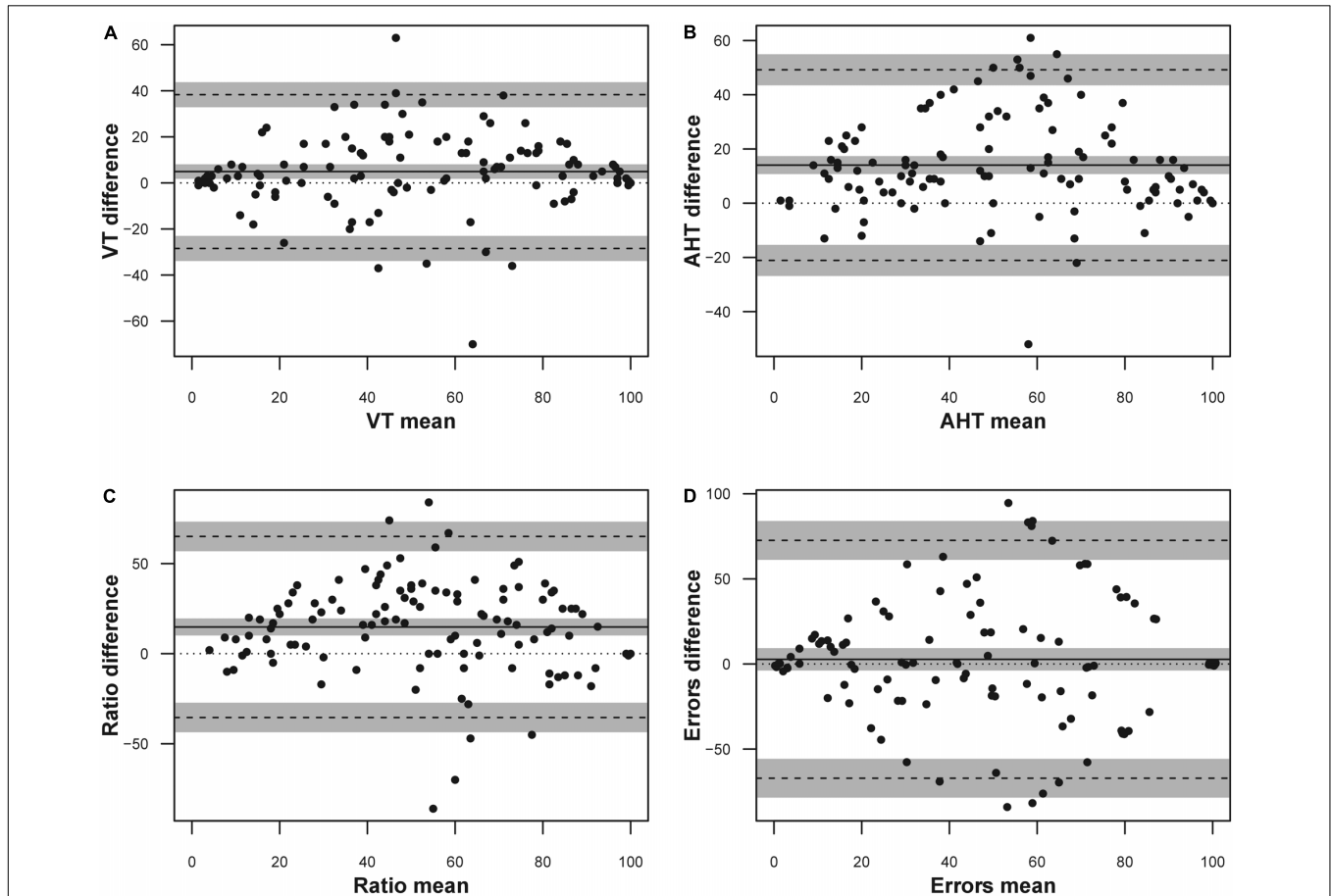


FIGURE 1 | Bland Altman plot for the DEM subtests expressed as percentile scores. The solid line represents the mean difference, the dashed lines represent the upper and lower boundaries of the 95% limits of agreement, the gray areas represent the 95% confidence intervals of the limits of agreement (LoA). (A) Vertical time; (B) adjusted horizontal time; (C) ratio; (D) errors. Only the errors data were jittered ($x \pm 1$; $y \pm 1$) to visualize point density.

time show significance for the factor Grade [$F_{(3,111)} = 19.24$, $p < 0.0001$, $\eta_p^2 = 0.34$] and for the factor Time [$F_{(1,111)} = 19.33$, $p < 0.0001$, $\eta_p^2 = 0.16$]. The adjusted horizontal time results show significance for the factor Grade [$F_{(3,111)} = 35.04$, $p < 0.0001$, $\eta_p^2 = 0.48$] and for the factor Time [$F_{(1,111)} = 61.8$, $p < 0.0001$, $\eta_p^2 = 0.37$]. The results for ratio show significance for the factor Grade [$F_{(3,111)} = 12.08$, $p < 0.0001$, $\eta_p^2 = 0.25$] and for the factor Time [$F_{(1,111)} = 31.75$, $p < 0.0001$, $\eta_p^2 = 0.22$]. The results for errors show significance for the factor Grade [$F_{(3,111)} = 11.76$, $p < 0.0001$, $\eta_p^2 = 0.24$] and for the interaction Time \times Grade [$F_{(3,111)} = 3.62$, $p < 0.05$, $\eta_p^2 = 0.08$]. Across all grades, the subjects

showed improvements with retest for vertical time, adjusted horizontal time and ratio scores (except for 2nd and 4th grade error; see Table 7 for details).

In order to show the mean improvement of performance on retest in a different way, it is possible to view these results in terms of learning effect according to raw data and percentile improvement. The learning effect during sessions performed for each DEM subtest and grade (from 2nd to 5th) shows an improvement, respectively, of: 3.7, 4.32, 0.95, and 5.34% for vertical time, 10.16; 9.6; 6.10; and 10.8% for adjusted horizontal time, 7.3; 5.88; 5.47; and 6.03% for ratio and -26.21; 20.65; 6.31; and 47.51% for Errors. In percentile terms, the same results

TABLE 7 | Descriptive statistics for test and re-test data on the four subtests of DEM.

Subtest	Grade	Test mean	SD	Retest mean	SD
VT	2	55.12	9.19	53.08	8.5
	3	47.19	8.32	45.15	8.21
	4	41.05	8.9	40.66	9.4
	5	38.56	9.44	36.5	7.93
AHT	2	74.71	16.23	67.12	10.48
	3	64.29	14.35	58.12	13.92
	4	51.96	10.12	48.79	9.95
	5	44.35	9.05	39.56	7.95
Ratio	2	1.37	0.28	1.27	0.14
	3	1.36	0.17	1.28	0.16
	4	1.28	0.13	1.21	0.12
	5	1.16	0.12	1.09	0.11
Errors	2	9.92	9.09	12.52	11.15
	3	8.86	10.28	7.03	8.19
	4	2.06	3.21	2.19	3.74
	5	4.21	5.31	2.21	2.24

VT, vertical time; AHT, adjusted horizontal time.

(unstratified) correspond to 4.93% for vertical time, 14.05% for adjusted horizontal time, 14.81% for ratio and 2.75% for errors. Lastly we report the standard error of measurement for all tests: 40.95 s for vertical time, 50.01 s for adjusted horizontal time, 0.301 for ratio and 14.44 for errors.

The pass-fail criteria for both administrations were only applied using the specific Italian norms for the 16th percentile criterion (Facchin et al., 2012).

The results listed in **Table 8** show a high or medium to high level of agreement for binary classification for vertical time, adjusted horizontal time, ratio and error. The same data of agreement reported in percentage show a range between 88 and 97% for vertical time, between 84 and 93% for adjusted horizontal time, between 75 and 97% for ratio and between 72 and 79% for errors. This level of agreement of binary classification appears to be equal or higher when compared with other studies, probably because it uses the last criterion of the 16th percentile (Tassinari and DeLand, 2005; Orlansky et al., 2011). Based on these data, we performed the Cohen *K* and AC_1 as a measure of agreement. The results of Cohen *K* are listed in **Table 9**. These results on Cohen *K* are moderate to high for vertical time and low to moderate for adjusted horizontal time, ratio and errors. These values are lower than others that have been previously reported (Tassinari and DeLand, 2005), but the different criterion used (16th vs. 30th percentile) may explain the differences. The AC_1 coefficients of agreement ($\pm 95\%$ CI) were 0.89 (0.81 – 0.96) for vertical time, 0.84 (0.75 – 0.92) for adjusted horizontal time, 0.79 (0.69 – 0.90) for ratio and 0.59 (0.44 – 0.74) for error.

DISCUSSION

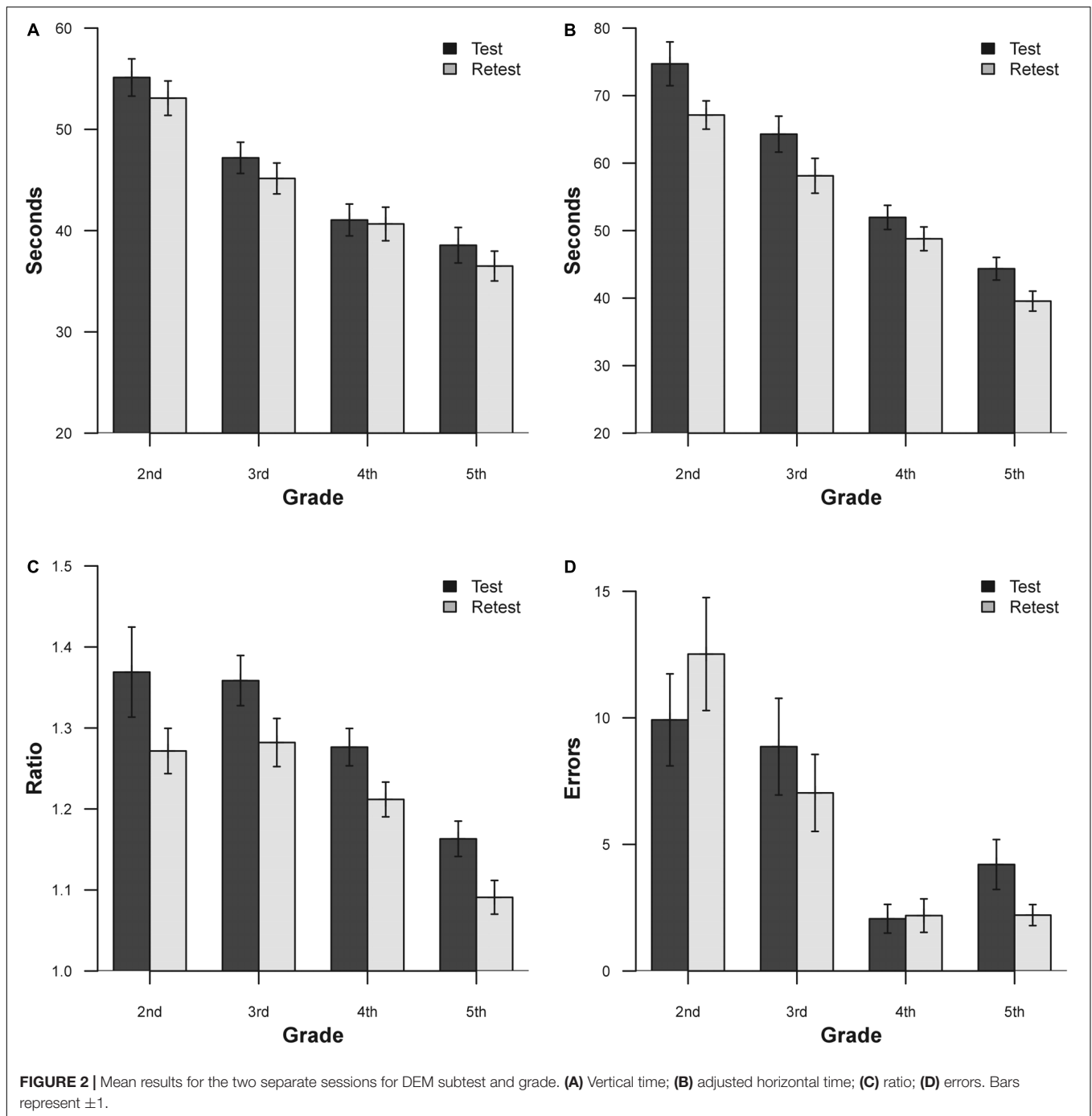
The purpose of this study was to re-evaluate the reliability of the DEM test with a test–retest method applying the

original test (as used in practice) twice, scored in seconds and percentile and evaluating in depth the improving of performance between sessions. It is worth noting that the replication of experiments and confirmation of the results play an important role in science (Open Science Collaboration, 2015; Gelman and Geurts, 2017). One of the purposes of the present study was to perform a replication study in the context of another population and language and also using different norms.

Taking into account the strict definition of reliability as the correlation between test and retest, we have obtained results that are consistent with some studies that have reported high values (Rouse et al., 2004; Tassinari and DeLand, 2005), and our results are significantly higher than others (Orlansky et al., 2011), probably for the use the same test cards and which are not different parallel versions. In fact, we have reconfirmed the conclusions of previous studies for the good to excellent reliability for vertical and adjusted horizontal time but a medium to high reliability for ratio and error scores. On the other hand, it seems that the parallel and test–retest reliabilities are slightly different, with higher results for the latter which, in practice, is the most important because the original parallel forms are not practical available for this test.

The results of agreement analyses show that there is a significant and distinct trend in the amelioration of performance in the second repetition. This lack of agreement and the presence of a learning effect is the main problem with reliability of the DEM test.

Based on the previous well-known phenomenon of the learning effect, the main focus of our study was to calculate these results as percentile scores, besides confirming the phenomena using a different population and language. In fact, for monitoring the performance of a child over time or the use of the DEM test to assess the effectiveness of a therapy, there is a requirement



to take into account the reliability of the test and its learning effect. The changes found in a second repetition of the test need to be greater than the repeatability itself. Our results, as expressed in seconds, show that, in order to be sure that the changes in the second administration can be attributed to therapy rather than test–retest variability, the results need to be higher than: about 9 s for the 2nd and 3rd grade, about 8 s for 4th and 10 s for 5th grade for vertical time; 30 s for 2nd grade, about 19 s for 3rd, about 15 s for 4th grade and 12 s

for 5th grade for horizontal time; 0.5 for 2nd grade, 0.3 for 3rd and 4th grade and 0.25 for 5th grade for ratio; 15, 15, 8, and 11 errors, respectively, for 2nd, 3rd, 4th, and 5th grade for errors. These results are objectively weak but lower than the previously found which suggested 20 s for vertical time; 60 s for adjusted horizontal time, 0.6 for ratio, and 23 for errors, respectively (Orlansky et al., 2011). Moreover, we calculated not only the 95% limits of agreement, but also the 95% confidence interval to have a statistical confidence for this measure. Also

TABLE 8 | Agreement between sessions separated for subtest (top) and grades (left), based on classification of the DEM findings (16th percentile).

Test	Grade		Retest							
			VT		AHT		Ratio		Error	
			P	F	P	F	P	F	P	F
	2nd	P	20	0	21	0	20	0	13	5
		F	2	3	4	0	5	0	2	5
	3rd	P	25	0	22	0	22	0	17	2
		F	1	3	4	3	4	3	4	6
	4th	P	24	3	26	2	22	2	23	2
		F	1	4	3	1	6	2	6	1
	5th	P	22	2	25	0	26	0	18	1
		F	0	5	2	2	1	2	7	3

P, pass; F, fail. VT, vertical time; AHT, adjusted horizontal time.

TABLE 9 | Cohen's K comparison between this study and those of Tassinari and DeLand (2005).

	This work	Tassinari "Office"	Difference (95% CI)
n	115	53	
VT	0.72 (0.55 to 0.89)	0.56 (0.34 to 0.79)	-0.16 (-0.438 to 0.126)
AHT	0.38 (0.14 to 0.62)	0.77 (0.56 to 0.98)	0.39 (0.071 to 0.709)
Ratio	0.37 (0.15 to 0.58)	0.73 (0.48 to 0.98)	0.36 (0.034 to 0.693)
Errors	0.36 (0.16 to 0.53)	0.77 (0.28 to 0.74)	0.41 (-0.109 to 0.612)

VT, vertical time; AHT, adjusted horizontal time.

with considering the confidence intervals, the difference between the results obtained by Orlansky et al. (2011) did not change (see Table 6).

Using percentile scoring, a score useful in practice, a change lower than 39 percentile points for vertical time, 49 for adjusted horizontal time, 65 for ratio and 72 for error as indicated could be interpreted, with care, to confirm amelioration. These values reflect the previous scores (limits of agreement) translated as percentiles and are useful for direct and easy clinical application. Confidence intervals on limits of agreement are calculated also for percentile scoring and reported in Table 7.

The lack of agreement and a remarkable learning effect was reflected in the generally moderate agreement of binary classification between sessions, with some changes in classification. The Kappa indexes of agreement were moderate to low and smaller than previously found. The AC₁ index gave better results and part of the low scores in kappa could arise from the limitation of this index when data are highly asymmetrical. Nevertheless all these values have to be taken into account for clinical use. The improvement over sessions is the main problem with DEM test reliability, but knowing and quantifying it could permit the correct decisions to be taken when different sessions need to be compared.

A possible source of the aforementioned learning effect could be the lack of a true pre-test on DEM, especially in the first session (Facchin et al., 2014). Indeed, the manual reports that, in cases of doubt, the test needs to be performed twice, although

the normative data were only collected for the first application and the improvement of time was not considered in the norms (Richman, 2009).

CONCLUSION

Developmental eye movement test reliability has some limitations due to the lack of agreement between sessions, but our results show that this problem is lower than previously found. We have confirmed that the results should be evaluated carefully when the DEM test is used in monitoring the effectiveness of treatment with new values in seconds and percentiles. With awareness of this limitation, the DEM test can be used in clinics in performing ocular movement assessment over time from the professionals interested in vision assessment.

ETHICS STATEMENT

We obtain the authorization from the "Istituto scolastico comprensivo "V.Muzio", Via S.Pietro ai Campi 1, 24126 Bergamo, Italy" School Council to perform the screening and the research. They act as a control council over the all activities performed in the school. The authorization has a reference number 23/2010 and was obtained on April 5, 2010.

Regarding the informed consent, we asked at parents (or tutor) of the child to compile and sign the written informed consent.

Only children with written informed consent from their parents participated in the study.

Children without written informed consent were not admitted.

AUTHOR CONTRIBUTIONS

The authors contributed differently for the several aspects of this study. AF and SM conceptualization, methodology, writing – original draft, and writing – review and editing.

AF data curation and formal analysis. SM investigation and supervision. All authors approved the final version of the work.

ACKNOWLEDGMENTS

We would like to thank Sabrina Prudenzano and Stefano Bellino for collecting data.

REFERENCES

- Altman, D. G., and Bland, J. M. (1983). Measurement in medicine: the analysis of method comparison studies. *Stat. Methods Med. Res.* 32, 307–317. doi: 10.2307/2987937
- Anastasi, A., and Urbina, S. (1997). *Psychological Testing*, 7th Edn. Prentice Hall, NJ: Upper Saddle River.
- Associazione Italiana Dislessia (2007). *Learning Disabilities. Recommendations for Clinical Practice Defined with Consensus Conference*. Bologna: Associazione Italiana Dislessia.
- Ayton, L. N., Abel, L. A., Fricke, T. R., and McBrien, N. A. (2009). Developmental eye movement test: what is it really measuring? *Optom. Vis. Sci.* 86, 722–730. doi: 10.1097/OPX.0b013e3181a6a4b3
- Baptista, A. M., de Sousa, R. A., de Moraes Guerra Casal, C. C., Marques, R. J., and da Silva, C. M. (2011). Norms for the developmental eye movement test for portuguese children. *Optom. Vis. Sci.* 88, 864–871. doi: 10.1097/OPX.0b013e3182195dae
- Bland, J. M., and Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 327, 307–310. doi: 10.1016/S0140-6736(86)90837-8
- Costa Santos, C., Bernardes, J., and Ayres-de-Campos, D. (2011a). Observer reliability and agreement: differences, difficulties, and controversies. *J. Clin. Epidemiol.* 64, 701–702. doi: 10.1016/j.jclinepi.2010.12.002
- Costa Santos, C., Bernardes, J., Ayres-de-Campos, D., Costa, A., and Costa, C. (2011b). The limits of agreement and the intraclass correlation coefficient may be inconsistent in the interpretation of agreement. *J. Clin. Epidemiol.* 64, 264–269. doi: 10.1016/j.jclinepi.2009.11.010
- Coulter, R. A., and Shallo-Hoffmann, J. (2000). The presumed influence of attention on accuracy in the developmental eye movement (DEM) test. *Optom. Vis. Sci.* 77, 428–432. doi: 10.1097/00006324-200008000-00010
- Diedenhofen, B., and Musch, J. (2015). cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS One* 10:e0121945. doi: 10.1371/journal.pone.0121945
- Dormer, A., and Zou, G. (2002). Interval estimation for a difference between intraclass kappa statistics. *Biometrics* 58, 209–215. doi: 10.1111/j.0006-341X.2002.00209.x
- Facchin, A., Maffioletti, S., and Carnevali, T. (2011). Validity reassessment of developmental eye movement (DEM) test in the Italian population. *Optom. Vis. Dev.* 42, 155–167.
- Facchin, A., Maffioletti, S., and Carnevali, T. (2012). The developmental eye movement (DEM) test: normative data for Italian population. *Optom. Vis. Dev.* 43, 162–179.
- Facchin, A., Ruffino, M., Facchetti, A., and Maffioletti, S. (2014). Modified direction of DEM test suggests differences in naming and eye movements. *Optom. Vis. Perform.* 2, 103–111.
- Fernandez-Velazquez, F. J., and Fernandez-Fidalgo, M. J. (1995). Do DEM test scores change with respect to the language? Norms for Spanish-speaking population. *Optom. Vis. Sci.* 72, 902–906. doi: 10.1097/00006324-199512000-00009
- Fleiss, J. L., and Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ. Psychol. Meas.* 33, 613–619. doi: 10.1177/001316447303300309
- Fleiss, J. L., Cohen, J., and Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychol. Bull.* 72, 323–327. doi: 10.1037/h0028106
- Garzia, R. P., Richman, J. E., Nicholson, S. B., and Gaines, C. S. (1990). A new visual-verbal saccade test: the development eye movement test (DEM). *J. Am. Optom. Assoc.* 61, 124–135.
- Gelman, A., and Geurts, H. M. (2017). The statistical crisis in science: how is it relevant to clinical neuropsychology? *Clin. Neuropsychol.* 31, 1000–1014. doi: 10.1080/13854046.2016.1277557
- Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochem. Med.* 25, 141–151. doi: 10.11613/BM.2015.015
- Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *Br. J. Math. Stat. Psychol.* 61, 29–48. doi: 10.1348/000711006X126600
- Jimenez, R., Gonzalez, M. D., Perez, M. A., and Garcia, J. A. (2003). Evolution of accommodative function and development of ocular movements in children. *Ophthalmic Physiol. Opt.* 23, 97–107. doi: 10.1046/j.1475-1313.2003.00093.x
- Kline, P. (2014). *The New Psychometrics: Science, Psychology and Measurement*. Milton Park: Taylor & Francis.
- Kottner, J., and Streiner, D. L. (2011). The difference between reliability and agreement. *J. Clin. Epidemiol.* 64, 701–702. doi: 10.1016/j.jclinepi.2010.12.001
- McGraw, K. O., and Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* 1, 30–46. doi: 10.1037/1082-989X.1.1.30
- Medland, C., Walter, H., and Woodhouse, M. J. (2010). Eye movements and poor reading: does the Developmental Eye Movement test measure cause or effect? *Ophthalmic Physiol. Opt.* 30, 740–747. doi: 10.1111/j.1475-1313.2010.00779.x
- Okumura, T., and Wakamiya, E. (2010). *Visual Skills in Children with Learning Difficulties*. Tokyo: Meijitoshu Shuppan Corporation.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716
- Orlansky, G., Hopkins, K. B., Mitchell, G. L., Huang, K., Frazier, M., Heyman, C., et al. (2011). Reliability of the developmental eye movement test. *Optom. Vis. Sci.* 88, 1507–1519. doi: 10.1097/OPX.0b013e318230f03a
- Palomo-Álvarez, C., and Puell, M. C. (2009). Relationship between oculomotor scanning determined by the DEM test and a contextual reading test in schoolchildren with reading difficulties. *Graefes Arch. Clin. Exp. Ophthalmol.* 247, 1243–1249. doi: 10.1007/s00417-009-1076-8
- Pang, P. C., Lam, C. S., and Woo, G. C. (2010). The Developmental Eye Movement (DEM) test and Cantonese-speaking children in Hong Kong SAR, China. *Clin. Exp. Optom.* 93, 213–223. doi: 10.1111/j.1444-0938.2010.00470.x
- Powers, M., Grisham, D., and Riles, P. (2008). Saccadic tracking skills of poor readers in high school. *Optometry* 79, 228–234. doi: 10.1016/j.optm.2007.07.014
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rainey, B. B., Schroeder, T. L., Goss, D. A., and Grosvenor, T. P. (1998). Inter-examiner repeatability of heterophoria tests. *Optom. Vis. Sci.* 75, 719–726. doi: 10.1097/00006324-199810000-00016
- Ramasundarahettige, C. F., Donner, A., and Zou, G. Y. (2009). Confidence interval construction for a difference between two dependent intraclass correlation coefficients. *Stat. Med.* 28, 1041–1053. doi: 10.1002/sim.3523
- Richman, J. E. (2009). *Developmental Eye Movement Test, Examiner's Manual, version 2.0*. South Bend, IN: Bernell Corp.
- Richman, J. E., and Garzia, R. P. (1987). *Developmental Eye Movement Test, Examiners Booklet, version 1*. South Bend, IN: Bernell Corp.

- Rouse, M. W., Nestor, E. M., Parot, C. J., and Deland, P. N. (2004). A reevaluation of the Developmental Eye Movement (DEM) test's repeatability. *Optom. Vis. Sci.* 81, 934–938.
- Santiago, H. C., and Perez, M. A. (1992). "Test-retest reliability of the developmental eye movement test," in *Proceedings of the AAO Annual Meeting*, Chicago, IL.
- Scheiman, M., and Wick, B. (2013). *Clinical Management of Binocular Vision: Heterophoric, Accommodative, and Eye Movement Disorders*. Philadelphia, PA: Lippincott Williams & Wilkins.
- Serdjukova, J., Ekimane, L., Valeinis, J., Skilters, J., and Krumina, G. (2016). How strong and weak readers perform on the Developmental Eye Movement test (DEM): norms for Latvian school-aged children. *Read. Writ.* 30, 233–252. doi: 10.1007/s11145-016-9671-7
- Tassinari, J. T., and DeLand, P. (2005). Developmental Eye Movement Test: reliability and symptomatology. *Optometry* 76, 387–399. doi: 10.1016/j.optm.2005.05.006
- Urbina, S. (2004). *Essentials in Psychological Testing*. Hoboken, NJ: John Wiley and Sons.
- Viera, A. J., and Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam. Med.* 37, 360–363.
- Webber, A., Wood, J., Gole, G., and Brown, B. (2011). DEM test, visagraph eye movement recordings, and reading ability in children. *Optom. Vis. Sci.* 88, 295–302. doi: 10.1097/OPX.0b013e31820846c0
- Xie, Y., Shi, C., Tong, M., Zhang, M., Li, T., Xu, Y., et al. (2016). Developmental Eye Movement (DEM) test norms for Mandarin Chinese-speaking Chinese children. *PLoS One* 11:e0148481. doi: 10.1371/journal.pone.0148481
- Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychol. Methods* 12, 399–413. doi: 10.1037/1082-989X.12.4.399

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

Copyright © 2018 Facchin and Maffioletti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Could Time Detect a Faking-Good Attitude? A Study With the MMPI-2-RF

Paolo Roma^{1*}, Maria C. Verrocchio², Cristina Mazza¹, Daniela Marchetti², Franco Burla¹, Maria E. Cinti¹ and Stefano Ferracuti¹

¹ Department of Human Neuroscience, Sapienza University of Rome, Rome, Italy, ² Department of Psychological, Health, and Territorial Sciences, University "G. d'Annunzio", Chieti-Pescara, Chieti, Italy

OPEN ACCESS

Edited by:

Marco Innamorati,
Università Europea di Roma, Italy

Reviewed by:

Elisa Pedrolì,
Istituto Auxologico Italiano (IRCCS),
Italy
Luigi Janiri,
Università Cattolica del Sacro Cuore,
Italy

*Correspondence:

Paolo Roma
paolo.roma@uniroma1.it

Specialty section:

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

Received: 12 March 2018

Accepted: 06 June 2018

Published: 25 July 2018

Citation:

Roma P, Verrocchio MC, Mazza C,
Marchetti D, Burla F, Cinti ME and
Ferracuti S (2018) Could Time Detect
a Faking-Good Attitude? A Study
With the MMPI-2-RF.
Front. Psychol. 9:1064.
doi: 10.3389/fpsyg.2018.01064

Background and Purpose: Research on the relationship between response latency (RL) and faking in self-administered testing scenarios have generated contradictory findings. We explored this relationship further, aiming to add further insight into the reliability of self-report measures. We compared RLs and T-scores on the MMPI-2-RF (validity and restructured clinical [RC] scales) in four experimental groups. Our hypotheses were that: the Fake-Good Speeded group would obtain a different completion time; show higher RLs than the Honesty Speeded Group in the validity scales; show higher T-Scores in the L-r and K-r scales and lower T-scores in the F-r and RC scales; and show higher levels of tension and fatigue. Finally, the impact of the speeded condition in malingering was assessed.

Materials and Methods: The sample was comprised of 135 subjects ($M = 26.64$; $SD = 1.88$ years old), all of whom were graduates (having completed at least 17 years of instruction), male, and Caucasian. Subjects were randomly assigned to four groups: Honesty Speeded, Fake-Good Speeded, Honesty Un-Speeded, and Fake-Good Un-Speeded. A software version of the MMPI-2-RF and Visual Analog Scale (VAS) were administered. To test the hypotheses, MANOVAs and binomial logistic regressions were run.

Results: Significant differences were found between the four groups, and particularly between the Honest and Fake-Good groups in terms of test completion time and the L-r and K-r scales. The speeded condition increased T-scores in the L-r and K-r scales but decreased T-scores in some of the RC scales. The Fake groups also scored higher on the VAS Tension subscale. Completion times for the first and second parts of the MMPI-2-RF and T-scores for the K-r scale seemed to predict malingering.

Conclusion: The speeded condition seemed to bring out the malingerers. Limitations include the sample size and gender bias.

Keywords: MMPI-2-RF, faking-good, speed, response latency, self-report, malingering

INTRODUCTION

A common concern for those using self-report inventories of personality and psychopathology is the susceptibility of such inventories to malingering or faking (Anastasi, 1988; Holden et al., 1992). Ziegler et al. (2012) defined faking as an intentional and deliberate behavior that helps an individual achieve personal goals. Specifically, *fake-good behavior* involves presenting the self in a more positive manner, relative to honest self-evaluation (Maricuțoiu and Sârbescu, 2016). In any assessment setting, a subject completing a personality inventory can answer truthfully or not, according to his or her goal. For this reason, detection of malingering represents an area of considerable interest for researchers of individual differences (Holden et al., 2001). Over the past years, psychologists have searched for methods to identify the occurrence of this phenomenon (Fluckinger et al., 2008).

In the 1970s, Dunn et al. (1972) suggested that *response latency* (RL; i.e., the amount of time elapsed between an item's presentation and a subject's response) could be used to detect dissimulation tendencies. Beginning in the 1990s, RL was proposed more insistently as an additional method of testing the validity of the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway and McKinley, 1951), together with the MMPI's own validity scales (L-r, F-r, and K-r).

Nevertheless, over the decades, divergent perspectives regarding RL and faking have emerged in the literature, and empirical research has produced mixed findings. According to the *semantic evaluation perspective* (Hsu et al., 1989), shorter RL is associated with higher scores on social desirability scales, because it is easier to evaluate the meaning of an item than to evaluate that item according to autobiographic information, which involves recalling episodes to direct the answer. More specifically, Dunn et al. (1972) found, in administering the MMPI, that participants in faking conditions had shorter RLs relative to participants in honesty conditions. Hsu et al. (1989), referring to the theories of response process proposed by Nowakowska (1970), Rogers (1971, 1977), and Kuncel (1973), studied the RL in the subtle-obvious scales of the MMPI on a sample of 100 undergraduate students who were instructed to fake-bad or fake-good, with or without an incentive. The results indicated that RL was shorter in the fake condition and that RL had incremental validity in detecting both faking-good and faking-bad. This finding is supported by the theory that responding to MMPI items with the intent to dissemble involves accessing a less elaborate information schema or network (Brunetti et al., 1998).

Several researchers have proposed theories and shown empirical results that diverge from the idea that faking speeds the processing of personality test items. Authors who support the *self-schema model* (McDaniel and Timm, 1990; Holden and Kroner, 1992; Holden et al., 1992; Walczyk et al., 2003, 2005; Foerster et al., 2013) argue that faking is a complex process that, relative to honest answering, requires extra cognitive processing and editing. Maricuțoiu and Sârbescu (2016) assumed that "honest respondents answer consistently with their self-schemas, while dishonest respondents decide not to provide self-schematic

information, after an evaluation of schematic information" (p. 2). Vasilopoulos et al. (2000) stated that fakers must reflect and, in turn, keep real information in memory, and they must inhibit and replace this real information with fake information taken from the target's ideal schema. This schema is hypothesized and, for this reason, more complex and not immediately available for recall; thus, it takes longer for faking respondents to provide an answer (see also DePaulo et al., 2003). Honest respondents, in contrast, are able to respond automatically and spontaneously, and thus they use fewer cognitive processes than malingerers and their RL is correspondingly shorter. According to these authors, fakers' larger RLs are due to higher levels of arousal, generated by their fear of being detected.

An interesting variant of the *self-schema model* was introduced by Holden (1995). The author found shorter RLs when items were congruent to the faking scheme: if subjects were asked to describe themselves in the best possible way (i.e., comply with a *fake-good scheme*), they registered shorter RLs on items describing socially desirable behaviors. A reverse pattern was observed for items incongruent with the scheme. Similar results were obtained by Holden and Lambert (2015) using the NEO-PF inventory (Costa and McCrae, 1992) and by Brunetti et al. (1998) using MMPI-2 (Hathaway and McKinley, 1989; Butcher et al., 2001). These authors showed that subjects required significantly more time to respond to items that were incongruent with their response set.

Some studies on RL have also evaluated the *pressure of time* effect on faking behavior with personality inventories. Khorramdel and Kubinger (2006) found that faking when responding to dichotomous items was accentuated under time pressure, and thus a time limitation may drive people to increase their faking behavior in the direction required by the instructions (data also reported by Holden et al., 2001). Shalvi et al. (2013) showed that subjects lie more frequently when they have little time to reflect; when they have more time at their disposal, they reflect more deeply on their response and moderate the simulation. Time pressure, therefore, seems an important factor in faking behavior.

While a theoretical basis may exist for the use of latencies in faking detection, previous research on the association of RL with faking has yielded mixed results and, recently, contradictory findings (fakers are faster, Maricuțoiu and Sârbescu, 2016; fakers are slower, Van Hooft and Born, 2012). Therefore, in the current research, we were interested in increasing the understanding of RL by merging it with a time pressure condition to determine whether the combination of these factors can help detect faking behavior.

Dividing our sample into an honest group (H) and a group instructed to fake-good (FG), we used a common self-administered inventory of personality and psychopathology, together with two conditions of time (speeded [S] and un-speeded [U]), to test the following hypotheses:

- H1: There would be significant differences in the protocol's total completion time. Analysis of these differences could increase our knowledge of fakers' test compiling attitudes, in both unrestricted (U) and speeded (S) time conditions.

- H2: There would be significant differences in completion times for the protocol's parts, both within and between groups. In the H groups, we expected a fatigue effect, resulting in progressively higher completion times. In the FG groups, we expected both a fatigue effect and a learning effect, due to the difficulty of learning the FG response model. Studying the partial time responses within groups and the differences across groups in each section could provide a deeper understanding of the information processing of honest and faking respondents.
- H3: There would be significant differences in RLs between groups in self-presentation measures on the self-administered inventory. We investigated the RLs of the self-presentation scales, in particular, since these were thought to be useful for differentiating between H and F respondents.
- H4: There would be differences between groups in self-presentation scores, with FG groups reporting higher values in positive self-presentation, lower scores in negative self-presentation, and lower values in psychopathology, relative to H groups.
- H5: There would be differences in tension and fatigue levels between the H and FG groups. We wanted to study the influence of these variables on RL and inventory scores.
- H6: The RLs identified in H1, H2, and H3 would be effective for predicting faking behavior.

As introduced in the hypotheses, we chose to restrict the analysis to a comparison between H and FG schemes. We chose FG for this study as it is more common than the fake-bad scheme, and thus the application of results would be more extensive. In other words, it is more likely that a situation will drive a subject to exhibit fake-good behaviors (e.g., during personnel selection or qualifying examinations) than fake-bad behaviors. Regarding the measure used, we chose the Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Ben-Porath and Tellegen, 2008), as it has been extensively used in clinical (see, e.g., Anderson et al., 2015) and selection settings (see, e.g., Tarescavage et al., 2015), but not yet used in latency studies. Furthermore, to the best of our knowledge, no prior study has addressed RL and MMPI scores under time pressure conditions.

MATERIALS AND METHODS

Participants

Subjects were 140 young adult volunteers who participated in the study for a small reward (European breakfast in a cafe). To limit confounding variables, we recruited only subjects who were aged 25–30 years ($M = 26.64$; $SD = 1.88$ years), male, Caucasian, graduates (having completed at least 17 years of education), and non-psychology graduates (i.e., those who had not attended the faculty of psychology). Subjects participated in the trial in the morning and were randomly assigned to one of four instruction groups. Six subjects were excluded from data analysis for one or more of the

following reasons: (a) failure to follow instructions as assessed by the final request ($n = 2$), (b) one or more changes in answers ($n = 3$), or (c) too brief a latency in one or more responses ($n = 1$, 3000 m/s). The remaining 135 subjects composed the research group. No statistically significant differences were observed on age or level of education. Data were collected over a period of 2 months, from October to November 2017.

Materials

MMP-2-RF

The full Italian version of the MMPI-2-RF (Sirigatti and Faravelli, 2012) was used. The MMPI-2-RF (Ben-Porath et al., 2008/2011) is a 51-scale measure of personality and psychopathology with 338 items, selected from the 567 of the MMPI-2 (Tellegen et al., 2003; Ben-Porath and Tellegen, 2008). In particular, this study used the T-scores of the three principal validity scales (L-r, F-r, and K-r) and the nine restructured clinical (RC) scales (to assess H4). We chose these scales as they represent the test's core evaluative measures and because our sample was not sufficiently large to guarantee a reliable analysis of all 51 scales (see **Table 1** for a brief description of the 12 selected scales). For our study, we added a Total scale, which was the sum of the T-scores of each of the nine RC scales. This Total scale was similar to the MMPI-2's "total elevation of protocol." T-scores ($M = 50$, $SD = 10$) are the traditional unit of measurement in the MMPI-2 (Tellegen and Ben-Porath, 1992), and they are also used in the MMPI-2-RF. The T-scores classification is: 45–54 (*average*), 55–69 (*slightly high*), 60–64 (*moderately high*), 65–69 (*high*), and 70–79 (*very high*) (Butcher et al., 2001).

We also assessed the completion time for the entire protocol (to assess H1) and the completion times for each of the three consecutive parts, which were composed of a similar number of items (112 for the first part, 112 for the second, and 114 for the

TABLE 1 | Selected MMPI-2-RF scales.

Scale	Title	What is measured
L-r	Uncommon Virtues	Infrequent and therefore improbable virtues
F-r	Infrequent Responses	Infrequent symptomatology
K-r	Adjustment Validity	Adaptation to life
RCd	Demoralization	Unhappiness and dissatisfaction with life
RC1	Somatic Complaints	Pattern of somatic complaints
RC2	Low Positive Emotions	Depressive symptoms
RC3	Cynicism	Negative view of human nature
RC4	Antisocial Behavior	Antisocial behavior and related family conflict
RC6	Ideas of Persecution	Persecutory beliefs
RC7	Dysfunctional Negative Emotions	Various negative emotional experiences
RC8	Aberrant Experiences	Thinking disorders
RC9	Hypomanic Activation	High level of activation and engagement

third, in order to assess H2); and the RL of the three principal validity scales (to assess H3).

Visual Analog Scale (VAS)

The VAS is a simple technique for measuring subjective experience (McCormack et al., 1988). It consists of a 10 cm line segment with two extreme polarities. Subjects must place a single mark on the line to indicate the current level of their experience (0 = the best possible condition, 10 = the worst possible condition). In our experiment, VAS was used to assess subjects' levels of tension (anxiety) (VAS-T) and fatigue (VAS-F), both before (T0) and after (T1) the MMPI-2-RF evaluation. The difference between VAS at T1 and T0 was used to understand changes in subjects' levels of tension and fatigue.

Software Application

We implemented an application for Android devices, with all 338 items loaded onto the platform. Participants used their dominant hand (126 right-handed, 9 left-handed) to press the virtual key F (false, on the bottom left) or V (true, on the bottom right) on the application. Following this response, the next item would appear immediately on the screen. At the top of the screen a red virtual button would offer subjects the possibility to return to the previous question. The program simultaneously recorded subjects' responses (V or F) and RL (measuring the time between the appearance of an item to the subject's tap of the virtual key) for each item. The same device was used for all uses of the application, and the application was stored on the device (rather than accessed online), so that Internet speed would not influence RL.

Research Design

A 2×2 between-subjects design was used. The two manipulated factors were instruction (H vs. FG) and time pressure (U vs. S). Participants were randomly assigned to one of four experimental groups of 35 persons: H/U, FG/U, H/S, and FG/S. The four instructions were:

- (1) H/U: "We are interested in some characteristics of your personality. We want you to take this test in a totally sincere fashion. After reading each item you should take all the time you need to respond in the best way."
- (2) FG/U: "We are interested in some characteristics of your personality. Imagine you are applying for a desired job. In this situation, it would be to your advantage to appear as if you were completely normal and psychologically healthy. Stated differently, we want you to take this test and deliberately fake good. Pay attention, because the questionnaire contains features designed to detect faking, and your intent is to respond in a way that your deception cannot be detected. After reading each item you should take all the time you need to respond in the best way, according to this instruction."
- (3) H/S: "We are interested in some characteristics of your personality. We want you to take this test in a totally honest fashion. After reading each item you should respond as quickly as possible. Short response time is an important factor in this test."
- (4) FG/S: "We are interested in some characteristics of your personality. Imagine you are applying for a desired job. In this situation it would be to your advantage to appear as if you were completely normal and psychologically healthy. Stated differently, we want you to take this test and deliberately fake good. Pay attention, because the questionnaire contains features designed to detect faking, and your intent is to respond in a way that your deception cannot be detected. After reading each item you should respond as quickly as possible. A short response time will enable you to stand out positively from other candidates."

Procedures

The subject, placed in front of a device on a 70 cm high desk with an adjustable height chair set at a distance of about 40 cm (with the back straight on the chair), received the following information and questions: (a) an explanation of the research and procedure, (b) a consent form, (c) a demographic questionnaire, (d) the T0 VAS (on white paper), (e) a brief introduction to the platform, (f) 10 training questions on the device, (g) 10 neutral questions (for which the average response time was collected), (h) instructions for the task, (i) the MMPI-2-RF test, (j) the T1 VAS (on white paper), and (k) a final check of their understanding of the instructions, as follows: after the trial, subjects performed two tasks designed to test their understanding: (1) write briefly on the card next to the device the initial instructions, and (2) write whether they thought they had followed the instructions when completing the protocol. Two participants proved not to have understood the task (1) and one subject declared not to have followed instructions during the test (2).

Statistical Analyses

In order to assess potentially noisy variables between the four groups (such as motor speed and reading speed) at the beginning, we ran an ANOVA to test for significant differences in RL in the 10 neutral questions (procedure point g).

Multivariate analyses of variance (MANOVAs) were run with the two attitudes toward the test conditions (H vs. FG) and the two speed groups (U vs. S) used as independent variables. Times of fulfillment, RL in the selected scale, T-scores, and VAS measures served as the dependent measures. Scheffé's (1959) method was used to assess *post hoc* pair differences ($p < 0.05$). Effect size was calculated using partial eta squared. Values of 0.02, 0.13, and 0.26 were considered indicative of small, medium, and large effects, respectively (Pierce et al., 2004). Binomial logistic regression was run to evaluate the discriminatory power of the variables related to time (dependent variable), with respect to the H condition (fixed factor).

RESULTS

The ANOVA showed a non-significant difference between groups [$F(3,131) = 1.585$; $p = 196$] on RL in the 10 neutral questions. No differences between groups were found on verbal ability or motor speed. We decided, however, to run MANCOVAs with the 10 neutral questions as covariates. As no significant covariate effect

TABLE 2 | Means and SDs of the four experimental groups for MMPI-2-RF completion time and RL in the three validity scales, with *post hoc* test results.

	H/U (n = 35) M (SD)	FG/U (n = 33) M (SD)	H/S (n = 33) M (SD)	FG/S (n = 34) M (SD)
Completion time (min)				
Items 1–112	7.46 (0.99) A	11.56 (1.30) B	5.68 (1.06) C	8.12 (1.27) A
Items 113–224	10.57 (1.19) A	11.07 (1.25) A	7.21 (0.65) B	8.76 (1.12) C
Items 225–338	12.64 (1.23) A	13.87 (1.67) B	9.74 (1.31) C	11.12 (2.32) D
Total time	30.67 (1.76) A	36.49 (2.19) B	22.64 (1.82) C	28.01 (1.78) D
Validity scales completion time (sec.)				
L-r	3.43 (1.14) A	5.52 (1.69) B	2.47 (0.40) C	4.52 (0.82) D
F-r	3.71 (1.52) A, B	4.47 (2.48) A	3.23 (1.56) C	2.75 (0.58) B, C
K-r	4.86 (1.36) B	6.23 (1.58) A	3.83 (0.38) C	6.28 (0.81) A

H, honest; U, un-speeded; FG, faking-good; S, speeded; L-r, Lie scale; K-r, Correction scale; F-r, Frequencies scale. For each line, different letters indicate a significant difference between columns.

was found, we decided not to include the neutral questions in the final analysis.

Variables Related to Time

In the seven variables related to completion time and RL, the MANOVAs revealed a significant effect of honesty [Wilks' lambda $F(3,125) = 91.503$; $p < 0.001$; $\eta_p^2 = 0.813$], speed [Wilks' lambda $F(3,125) = 125.583$; $p < 0.001$; $\eta_p^2 = 0.853$], and the interaction of honesty and time [Wilks' lambda $F(3,128) = 8.472$; $p < 0.001$; $\eta_p^2 = 0.287$]. **Table 2** shows the descriptive values of the four groups for the protocol and validity scale fulfillment times.

Regarding total completion time, the H/S group was fastest, followed by the FG/S, H/U, and FG/U groups. Therefore, both FG groups were slower than the H groups in the same speed condition (U: H = 30.67 min vs. FG = 36.49 min; S: H = 22.64 min vs. FG = 28.01 min). Contrasting the three partial completion times between the four groups, the results showed that the S condition—in both the H and FG groups—always reduced execution time by 2 or 3 min per section, relative to the U condition. The FG/S group was slower than the H/S group in completing all three sections. In the U condition, H was faster than FG in the first and third sections, while both conditions showed equal means in the second section.

Analyzing within-group differences, subjects of the H groups (in both time conditions) showed progressive and significant increases in completion times from the first to the third section. FG groups showed a different pattern, with quite similar completion times for the first two sections and a shorter time for the third section.

In the L-r scale, H groups were faster (first H/S, then H/U) than FG groups. FG groups showed a significant difference of about 1 second between FG/S (faster) and FG/U. In the K-r scale, FG groups showed the same RL and were slower than H groups, for whom the H/S group showed the fastest times. In the F-r scale, groups diverged in the S condition (with H/S faster than H/U and FG/S faster than FG/U), though the average value of the H/U group did not significantly differ from that of the FG/S group (see **Figure 1**).

T-Scores in MMPI-2-RF

In the 12 variables related to *T*-scores, MANOVAs revealed a significant effect of honesty [Wilks' lambda $F(3,125) = 27.308$; $p < 0.001$; $\eta_p^2 = 0.732$] and speed [Wilks' lambda $F(3,125) = 3.209$; $p < 0.001$; $\eta_p^2 = 0.243$], and a non-significant effect of the interaction between honesty and time [Wilks' lambda $F(3,128) = 1.726$; $p = 0.069$; $d = 0.147$]. **Table 3** reports the descriptive *T*-scores for the four groups for the selected MMPI-2-RF scales.

In the L-r and K-r scales, a *post hoc* test showed that the FG/S group obtained significantly higher *T*-scores than the other three groups. The FG/U group obtained the second highest values (significantly different from those of the other three groups), while both H groups (U and S) obtained similar results. It is interesting to underline that the *T*-scores of the L-r and K-r scales were in the normal range in the two H groups, while the FG/S group showed a very high range in the L-r scale and the FG/U group showed a moderately high range in the same scale. The FG/U group showed a tendentially high range in the K-r scale and the FG/S group showed a moderately high range in the same scale. In the F Scale, scores for the H/U (higher) and FG/S (lower) groups significantly differed.

In the RC scales, all scores were in the normal range. In RC1 and RC2, no significant differences were found between groups. Results showed the same trend, with the FG/S group achieving the lowest value, followed by the H/S, FG/U, and H/U groups. Similarly, no significant differences between groups were found in RC3, with the difference between all four groups bounded within 2.7 points. In RC4, RC6, RC7, and RC8, only the H/U (highest scores) and FG/S (lowest scores) groups differed markedly. In RCd and RC9, the H groups differed from the FG groups in the S condition. In the Total scale, the FG/S group reported lower scores than the other three groups.

Subjective Psychological Being

In the two VAS, MANOVA results revealed a significant effect for honesty [Wilks' lambda $F(2,130) = 71.170$; $p < 0.001$; $\eta_p^2 = 0.523$], speed [Wilks' lambda $F(2,130) = 45.257$; $p < 0.001$; $\eta_p^2 = 0.410$], and the interaction between honesty and time [Wilks' lambda

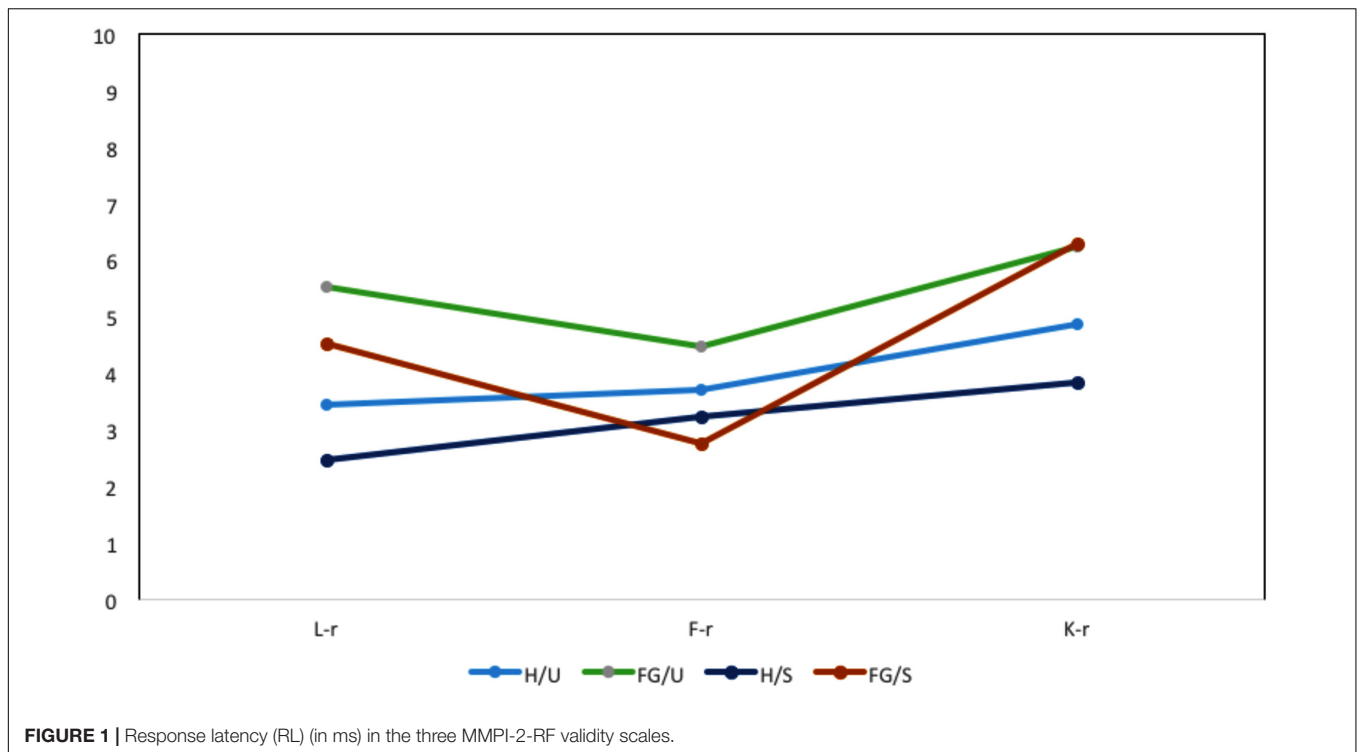


TABLE 3 | Means and SD in the four experimental groups for T-scores in the selected MMPI-2-RF scales, with *post hoc* test results.

	H/U (n = 35) M (SD)	FG/U (n = 33) M (SD)	H/S (n = 33) M (SD)	FG/S (n = 34) M (SD)
Validity scales				
L-r	48.54 (8.33) A	64.21 (12.68) B	48.03 (7.10) A	70.50 (6.09) C
F-r	47.09 (11.39) A	42.36 (8.09) A, B	44.85 (11.70) A, B	39.12 (7.45) B
K-r	52.80 (4.70) A	57.58 (6.33) B	51.67 (6.56) A	62.12 (4.13) C
RC scales				
RCd Demoralization	52.57 (6.29) A	48.57 (9.08) A, B	52.39 (8.41) A	43.88 (9.49) B
RC1 Somatic Complaints	53.40 (12.79) A	50.72 (13.90) A	48.93 (9.43) A	46.65 (13.97) A
RC2 Low Positive Emotions	49.14 (8.15) A	48.39 (8.92) A	47.24 (8.77) A	44.50 (5.96) A
RC3 Cynicism	53.74 (9.26) A	51.03 (6.59) A	51.45 (7.22) A	51.15 (10.86) A
RC4 Antisocial Behavior	53.80 (8.90) A	46.79 (9.33) B, C	52.97 (10.39) A, B	44.26 (11.26) C
RC6 Ideas of Persecution	48.57 (6.16) A	44.85 (9.41) A, B	45.73 (7.62) A, B	41.12 (6.65) B
RC7 Dysfunctional Negative Emotion	52.74 (6.70) A	48.97 (6.59) A	49.82 (6.71) A, B	44.11 (9.38) B
RC8 Aberrant Experiences	45.28 (11.45) A	40.88 (10.12) A, B	43.45 (9.54) A, B	38.42 (6.18) B
RC9 Hypomanic Activation	53.86 (8.24) A, B	47.76 (11.32) A, C	54.70 (6.66) B	47.26 (9.68) C
Total RC T-score	463.11 (32.34) A	427.93 (35.27) B	446.70 (37.12) A, B	401.47 (43.94) C

H, honest; U, un-speeded; FG, faking-good; S, speeded; L-r, Lie scale; K-r, Correction scale; F-r, Frequencies scale. For each line, different letters indicate a significant difference between columns.

$F(2, 130) = 4.030; p = 0.020; \eta_p^2 = 0.058$]. **Table 4** reports the descriptive values of the four groups for the VAS, with *post hoc* results.

A *post hoc* test revealed that tension was higher in the FG/S group. For fatigue, the H/U group was lowest while the FG/S group was highest. We also examined the correlation between the sum of the two VAS (VAS-T and VAS-F) and the RLs in the three validity scales. The results showed a positive correlation with the L-r scale ($rs = 0.357; p < 0.01$) and the K-r scale ($rs = 0.426;$

$p < 0.01$). No significant correlation was found in the F-r scale ($rs = 0.191$).

Regression Analyses

A test of the full model against a constant only model was statistically significant (**Table 5**), indicating that the set of predictors reliably distinguished between the presence or absence of honesty [$\chi^2(6) = 121.075, p < 0.001$]. Nagelkerke's R^2 of 0.790 indicated a moderately strong relationship between

TABLE 4 | Means and SD in the four experimental groups for VAS, with *post hoc* test results.

VAS	H/U (<i>n</i> = 35) M (SD)	FG/U (<i>n</i> = 33) M (SD)	H/S (<i>n</i> = 33) M (SD)	FG/S (<i>n</i> = 34) M (SD)
Tension	2.83 (1.76) A	4.84 (1.86) B	3.76 (2.08) A, B	7.50 (1.78) C
Fatigue	3.54 (1.42) A	5.67 (1.31) B	5.61 (1.30) B	7.29 (1.40) C

VAS, Visual Analog Scale; H, honest; U, un-speeded; FG, faking-good; S, speeded.

TABLE 5 | Binomial logistic regression.

Completion time	B	Exp(B)	Chi-squared Wald test	
			Test	<i>p</i>
First part	-1.243	0.288	16.843	<0.001
Second part	0.666	1.946	4.844	0.028
Third part	-0.113	0.893	0.380	NS
L-r	-0.438	0.646	0.998	NS
F-r	0.447	1.564	2.897	NS
K-r	-1.016	0.362	5.293	0.021

L-r, Lie scale; K-r, Correction scale; F-r, Frequencies scale; NS, not significant.

prediction and grouping. Prediction success, overall, was 93.3% (94.1% for the H condition and 92.5% for the FG condition). The Wald criterion demonstrated that three variables made a significant contribution to the prediction (first and second part of the inventory, K-r, and RL). The Exp(B) value indicated that when these three variables raised by one point, the possibility of faking-good behavior increased 0.28, 1.95, and 0.36, respectively.

DISCUSSION

The principal aim of this study was to assess whether RL in a self-administered questionnaire could discriminate between honest and faking-good respondents, with particular attention given to the effect of time pressure (in a speeded condition) on faking-good. We were interested in gaining insight into this relationship in order to test the use of time as another variable of validity in self-reported inventories, particularly in cases where subjects could be motivated to represent themselves in a better light (e.g., personnel selection). While this topic has been researched since the 1970s, the results have been mixed.

Overall, our results found that H respondents were faster than FG ones. In more detail, our data confirmed H1 (relating to different completion times between groups). Briefly, there was a faster group (H/S) and a slower group (FG/U), and FG groups were always slower than H groups under the same speed conditions. H2 was also mostly confirmed by our data. We found a clear progression of completion times in the two H groups, while in both FG groups only the completion time of the third part was higher than that of the first two. H3 was partially confirmed. In the two scales of positive self-representation (L-r and K-r), FG groups registered longer completion times than H groups. The L-r scale produced a clearer result, differentiating between all four groups, while in K-r, FG groups took a similar

time to respond, suggesting that the reasoning was complex and required more difficult choices. After all—with respect to the L-r scale—the K-r scale assesses more complex behaviors, concerning live adaptation and the ability to control one's own reactions (Friedman et al., 2014). In the F-r scale, results were confused and did not confirm our hypothesis; this may have been due to the fact that we tested normal subjects (as discussed further, below).

However, the finding that emerged most clearly was the shorter completion times and RLs of H groups, relative to FG groups. How should this finding be explained? Over time, researchers have developed various interpretive models. Markus (1977) and Kuiper (1981) hold that schema-relevant characteristics are more difficult to determine than self-schema characteristics. According to Holden et al. (1992), the larger RL of fakers can be attributed to their greater use of cognitive processes relative to honest responders: dishonest respondents must evaluate schematic information before they choose not to provide self-schematic information. On the other hand, self-schematic information is sufficient for honest respondents, who answer more quickly. According to Vasilopoulos et al. (2000) the larger RL of fakers is produced by higher emotional arousal caused by the fear of detection. According to DePaulo et al. (2003), fakers take longer to respond because the schema of an ideal respondent is less accessible than the self-schema of an honest respondent. The present results relating to the L-r scale support Holden's (1995) theory, as the L-r scale comprises 14 items (11 false and only 3 true). Similar to the findings of Holden et al. (1992), we found that FG groups took more time to respond to this scale, as it prevalently scores false. A similar interpretation applies to the K-r scale, which is composed of 16 items (14 false and only 2 true).

The specific pattern of RL and completion time found in our data suggest that, while H groups showed progressive fatigue over the full execution of the test, FG groups' fatigue was interpolated with a longer latency, probably due to the effort required to provide good (and perhaps false) self-information. In the first part of the questionnaire, FG groups reported slower completion times, probably because they were learning a model of FG. In other words: (a) FG respondents may have taken more time to fill in the different sections of the questionnaire than H respondents because they needed more time to think before answering; (b) the natural effect of fatigue in FG respondents may have been amplified by an initial difficulty in learning the FG response model, and this may have increased the completion time in the first part of the test; and (c) the influence of tension and anxiety may have muddled FG respondents' thoughts. The data underlines that the mental task and cognitive process of FG respondents were more complicated than those of H respondents.

The results of the MMPI-2-RF validity scales support H4. FG groups reported higher values on the positive self-presentation scales (L-r and K-r), as also found by others (e.g., Brunetti et al., 1998). The F-r data were complex and only partially satisfied H4. We believe that this occurred because we tested a normal sample, and thus the "floor effect" described by Peterson et al. (1989) was high (honest respondents endorsed so few psychopathology-related items that, when asked to fake good, few differences could

be noted). The results also stressed that speed induced FGs to significantly improve their self-representation in the L-r and K-r scales. This is an interesting outcome, which we attributed to the S condition leading respondents to drastically reduce their consideration of the appropriateness of lying on items about a virtuous attitude. In H subjects, however, speed did not produce differences in L-r, F-r, and K-r scores; this suggests that answering honestly at speed does not lessen scores relative to answering at leisure. The data thus confirm the work of Khorramdel and Kubinger (2006), who found that faking in responding to dichotomous items was accentuated under time pressure. Scores of the RC scales did not reach clinical significance. However, this outcome should take into account the fact that the sample did not belong to a clinical population.

With respect to the variables of tension during the trial and fatigue after the trial (H5), the FG/S group achieved the highest scores, followed by the FG/U and H/S groups. It seems that both the fake good request and the speed request required additional psychological effort on the part of respondents. In other words, the H/S group had to think only about being fast, the FG/U group had to think about only reflecting themselves in the best light, while the FG/S group faced both challenges: going fast and showing their best face. Our results substantiate previous data (see McDaniel and Timm, 1990) showing increased emotional arousal experienced by subjects making an impression managed response under time restriction (Temple and Geisinger, 1990).

With regard to H6, increased completion time in the first part and the K-r scale decreased the probability of honest responding; in contrast, increased completion time in the central part of the test increased the probability of honest responding. These results align with our previous interpretations: in the first part, FG respondents had to learn a schema of dishonesty, and so longer completion times in this section could lead us to believe that subjects were fakers. Further, the K-r scale required complex answers, and thus a long RL may have been associated with malingering behaviors. Moreover, if completion of the second part did not increase significantly relative to the first, there was a greater possibility of dishonest responding.

In conclusion, our data were consistent with the findings of McDaniel and Timm (1990), Walczyk et al. (2005), Foerster et al. (2013), and Maricuțoiu and Sârbescu (2016), which point to an increased response time among FG groups. Moreover, the S condition might more accurately enable the detection of dishonesty, as also found by Khorramdel and Kubinger (2006) and Shalvi et al. (2013), using other questionnaires.

REFERENCES

- Anastasi, A. (1988). *Psychological Testing*, 6th Edn. New York, NY: Macmillan.
- Anderson, J. L., Sellbom, M., Pymont, C., Smid, W., De Saeger, H., and Kamphuis, J. H. (2015). Measurement of DSM-5 section II personality disorder constructs using the MMPI-2-RF in clinical and forensic samples. *Psychol. Assess.* 27, 786–800. doi: 10.1037/pas0000096
- Ben-Porath, Y. S., and Tellegen, A. (2008). Empirical correlates of the MMPI-2 restructured clinical (RC) scales in mental health, forensic, and nonclinical settings: an introduction. *J. Pers. Assess.* 90, 119–121. doi: 10.1080/00223890701845120
- Ben-Porath, Y. S., and Tellegen, A. (2008/2011). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2 Restructured Form) Manual for Administration, Scoring, And Interpretation*. Minneapolis, MN: University of Minnesota Press.
- Brunetti, D. G., Schlottmann, R. S., Scott, A. B., and Hollrah, J. L. (1998). Instructed faking and MMPI-2 response latencies: the potential for assessing response validity. *J. Clin. Psychol.* 54,

Strengths and Limitations

The present study adds useful insight to the debate over the response times of fakers, while examining variables that have not yet been considered in the literature (e.g., completion times for individual sections of a questionnaire). Furthermore, to the best of our knowledge, this study was the first to jointly evaluate honesty conditions and time pressure in the MMPI-2-RF.

Nevertheless, there are two important limitations of this study that require additional research to overcome: (a) the analyzed group was selected for specificity (graduate males aged 25–30 years), and this reduced the generalizability of the findings; and (b) the sample size was small. Moreover, in the future, it would be useful to study a sample of subjects in an ecological condition (e.g., psycho-aptitude or forensic evaluation) and to examine RL differences according to item content. Future studies could investigate whether RLs are associated with particular scales of personality inventories within specific assessment settings in which malingerers must fake good to achieve certain goals.

CONCLUSION

The results suggest that, in computerized self-administered personality and psychopathology tests, RL and completion times could be usefully treated as additional indexes of falsification in self-representation. Furthermore, as speed increases our ability to identify falsifying subjects, time conditions could be applied to selection contexts in which self-reports are often used.

ETHICS STATEMENT

This study was carried out with written informed consent by all subjects and was approved by the local ethics committee (Board of the Department of Human Neuroscience, Faculty of Medicine and Dentistry, Sapienza University of Rome).

AUTHOR CONTRIBUTIONS

All authors helped to conceive and plan the study and prepared and approved the final manuscript. PR conducted the data collection and produced the first draft of the final manuscript. SF, MV, and DM supervised the data collection. PR and CM conducted the analyses and wrote the manuscript. MV and DM carefully read the final version of the manuscript.

- 143–153. doi: 10.1002/(SICI)1097-4679(199802)54:2<143::AID-JCLP3>3.0.CO;2-T
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahstrom, W. G., and Kaemmer, B. (2001). *MMPI-2. Manual for Administration and Scoring*. Minneapolis, MN: University of Minnesota Press.
- Costa, P. T., and McCrae, R. R. (1992). *NEO PI-R Professional Manual: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Harris, C. (2003). Cues to deception. *Psychol. Bull.* 129, 74–112. doi: 10.1037/0033-2909.129.1.74
- Dunn, T. G., Lushene, R. E., and O'Neil, H. F. (1972). Complete automation of the MMPI and a study of its response latencies. *J. Consult. Clin. Psychol.* 39, 381–387. doi: 10.1037/h0033855
- Fluckinger, C. D., McDaniel, M. A., and Whetzel, D. L. (2008). “Review of faking in personnel selection,” in *In Search of the Right Personnel*, ed. M. Mandal (New Delhi: McMillan).
- Foerster, A., Pfister, R., Schmidts, C., Dignath, D., and Kunde, W. (2013). Honesty saves time (and justifications). *Front. Psychol.* 4:473. doi: 10.3389/fpsyg.2013.00473
- Friedman, A. F., Bolinsky, P. K., Levak, R. W., and Nichols, D. S. (2014). *Psychological Assessment with the MMPI-2/MMPI-2-RF*. New York, NY: Routledge.
- Hathaway, S. R., and McKinley, J. C. (1951). *Minnesota Multiphasic Personality Inventory*. Minneapolis, MN: University of Minnesota Press.
- Hathaway, S. R., and McKinley, J. C. (1989). *Minnesota Multiphasic Personality Inventory – 2*. Minneapolis, MN: University of Minnesota Press.
- Holden, R. R. (1995). Response latency detection of fakers on personnel tests. *Can. J. Behav. Sci.* 27, 343–355. doi: 10.1037/0008-400X.27.3.343
- Holden, R. R., and Kroner, D. G. (1992). Relative efficacy of differential response latencies for detecting faking on a self-report measure of psychopathology. *Psychol. Assess.* 4, 170–173. doi: 10.1037/1040-3590.4.2.170
- Holden, R. R., Kroner, D. G., Fekken, G. C., and Popham, S. M. (1992). A model of personality test item response dissimulation. *J. Pers. Soc. Psychol.* 63, 272–279. doi: 10.1037/0022-3514.63.2.272
- Holden, R. R., and Lambert, C. E. (2015). Response latencies are alive and well for identifying fakers on a self-report personality inventory: a reconsideration of van Hooft and Born (2012). *Behav. Res. Methods* 47, 1436–1442. doi: 10.3758/s13428-014-0524-5
- Holden, R. R., Wood, L. L., and Tomashewski, L. (2001). Do response time limitations counteract the effect of faking on personality inventory validity. *J. Pers. Soc. Psychol.* 81, 160–169. doi: 10.1037/0022-3514.81.1.160
- Hsu, L. M., Santelli, J. E., and Hsu, J. R. (1989). Faking detection validity and incremental validity of response latencies to MMPI subtle and obvious items. *J. Pers. Assess.* 53, 278–295. doi: 10.1207/s15327752jpa5302_6
- Khorramdel, L., and Kubinger, K. D. (2006). The effect of speediness on personality questionnaires: an experiment on applicants within a job recruiting procedure. *Psychol. Sci. Q.* 48, 378–397.
- Kuiper, N. A. (1981). Convergent evidence for the self as a prototype: the “inverted-U RT effect” for self and other judgments. *Pers. Soc. Psychol. Bull.* 7, 438–443. doi: 10.1177/014616728173012
- Kuncel, R. B. (1973). Response processes and relative location of subject and item. *Educ. Psychol. Meas.* 33, 545–563. doi: 10.1177/001316447303300302
- Maricuțoiu, L. P., and Sârbescu, P. (2016). The relationship between faking and response latencies: a meta-analysis. *Eur. J. Psychol. Assess.* 1–11. doi: 10.1027/1015-5759/a000361
- Markus, H. (1977). Self-schemata and processing information about the self. *J. Pers. Soc. Psychol.* 35, 63–78. doi: 10.1037/0022-3514.35.2.63
- McCormack, H. M., Horne, D. J., and Sheather, S. (1988). Clinical applications of visual analogue scales: a critical review. *Psychol. Med.* 18, 1007–1019. doi: 10.1017/S0033291700009934
- McDaniel, M. A., and Timm, T. (1990). “Lying takes time: predicting deception in biodata using response latency,” in *Symposium Presented at the 98th Annual Conference of the American Psychological Association*, Boston, MA.
- Nowakowska, M. (1970). A model of answering to a questionnaire item. *Acta Psychol.* 34, 420–439. doi: 10.1016/0001-6918(70)90036-3
- Peterson, G. W., Clark, D. A., and Bennett, B. (1989). The utility of MMPI subtle, obvious scales for detecting fake good and fake bad response sets. *J. Clin. Psychol.* 45, 575–582. doi: 10.1002/1097-4679(198907)45:4<575::AID-JCLP2270450412>3.0.CO;2-F
- Pierce, C. A., Block, R. A., and Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educ. Psychol. Meas.* 64, 916–924. doi: 10.1177/0013164404264848
- Rogers, T. B. (1971). The process of responding to personality items: some issues, a theory, and some research. *Multivariate Behav. Res. Monogr.* 6, 1–65.
- Rogers, T. B. (1977). Self-reference in memory: recognition of personality items. *J. Res. Pers.* 11, 295–305. doi: 10.1016/0092-6566(77)90038-1
- Scheffé, H. (1959). *The Analysis of Variance*. New York, NY: John Wiley & Sons.
- Shalvi, S., Eldar, O., and Bereby-Meyer, Y. (2013). Honesty requires time: a reply to Foerster et al. (2013). *Front. Psychol.* 4:634. doi: 10.3389/fpsyg.2013.00634
- Sirigatti, S., and Faravelli, C. (2012). *MMPI-2 RF: Adattamento Italiano. Taratura, Proprietà Psicometriche e Correlati Empirici*. Florence: Giunti O.S. Organizzazioni Speciali.
- Tarescavage, A. M., Corey, D. M., Gupton, H. M., and Ben-Porath, Y. S. (2015). Criterion validity and practical utility of the Minnesota multiphasic personality inventory–2–restructured form (MMPI–2–RF) in assessments of police officer candidates. *J. Pers. Assess.* 97, 382–394. doi: 10.1080/00223891.2014.995800
- Tellegen, A., and Ben-Porath, Y. S. (1992). The new uniform T scores for the MMPI-2: rationale, derivation, and appraisal. *Psychol. Assess.* 4, 145–155. doi: 10.1037/1040-3590.4.2.145
- Tellegen, A., Ben-Porath, Y. S., McNulty, J. L., Arbisi, P. A., Graham, J. R., and Kaemmer, B. (2003). *RC Scales Test Monograph*. Minneapolis, MN: University of Minnesota Press.
- Temple, D. E., and Geisinger, K. F. (1990). Response latency to computer-administered inventory items as an indicator of emotional arousal. *J. Pers. Assess.* 54, 289–297. doi: 10.1207/s15327752jpa5401&2_27
- Van Hooft, E. A. J., and Born, M. P. (2012). Intentional response distortion on personality tests: using eye-tracking to understand response processes when faking. *J. Appl. Psychol.* 97, 301–316. doi: 10.1037/a0025711
- Vasilopoulos, N. L., Reilly, R. R., and Leaman, J. A. (2000). The influence of job familiarity and impression management on self-report measure scale score and response latencies. *J. Appl. Psychol.* 85, 50–64. doi: 10.1037/0021-9010.85.1.50
- Walczyk, J. J., Roper, K. S., Seemann, E., and Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: response time as a cue to reception. *Appl. Cogn. Psychol.* 17, 755–774. doi: 10.1002/acp.914
- Walczyk, J. J., Schwartz, J. P., Clifton, R., Barrett, A., Wei, M., and Zha, P. (2005). Lying person to person about life events: a cognitive framework for lie detection. *Pers. Psychol.* 58, 141–170. doi: 10.1111/j.1744-6570.2005.00484.x
- Ziegler, M., MacCann, C., and Roberts, R. D. (eds). (2012). “Faking: knowns, unknowns, and points of contention,” in *New Perspectives on Faking in Personality Assessment* (New York, NY: Oxford University Press), 3–16.

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

The handling Editor declared a past co-authorship with two of the authors PR and SF.

Copyright © 2018 Roma, Verrocchio, Mazza, Marchetti, Burla, Cinti and Ferracuti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Using Item Response Theory for the Development of a New Short Form of the Eysenck Personality Questionnaire-Revised

*Daiana Colledani**, Pasquale Anselmi and Egidio Robusto

Department of Philosophy, Sociology, Education and Applied Psychology, School of Psychology, University of Padova, Padova, Italy

OPEN ACCESS

Edited by:

Marco Innamorati,
Università Europea di Roma, Italy

Reviewed by:

Davide Marengo,
Università degli Studi di Torino, Italy
Pietro Cipresso,
Istituto Auxologico Italiano (IRCCS),
Italy

*Correspondence:

Daiana Colledani
daianacolledani@gmail.com

Specialty section:

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

Received: 15 March 2018

Accepted: 07 September 2018

Published: 02 October 2018

Citation:

Colledani D, Anselmi P and Robusto E
(2018) Using Item Response Theory
for the Development of a New Short
Form of the Eysenck Personality
Questionnaire-Revised.
Front. Psychol. 9:1834.
doi: 10.3389/fpsyg.2018.01834

The present work aims at developing a new version of the short form of the Eysenck Personality Questionnaire-Revised, which includes Psychoticism, Extraversion, Neuroticism, and Lie scales (48 items, 12 per scale). The work consists of two studies. In the first one, an item response theory model was estimated on the responses of 590 individuals to the full-length version of the questionnaire (100 items). The analyses allowed the selection of 48 items well discriminating and distributed along the latent continuum of each trait, and without misfit and differential item functioning. In the second study, the functioning of the new form of the questionnaire was evaluated in a different sample of 300 individuals. Results of the two studies show that reliability of the four scales is better than, or equal to that of the original forms. The new version outperforms the original one in approximating scores of the full-length questionnaire. Moreover, convergent validity coefficients and relations with clinical constructs were consistent with literature.

Keywords: short Eysenck personality questionnaire-revised, item response theory, 2PL, ESEM, DIF

INTRODUCTION

In the view of Eysenck (see Eysenck and Eysenck, 1975, 1991), the structure of personality may be effectively described by three main traits: psychoticism (P), extraversion (E), and neuroticism (N). These dimensions are also known as the “Giants Three” and represent basic, independent, and biologically founded traits. They characterize all subjects, with varying degrees, and allow for effectively describing behavioral, emotional, and individual differences among adults and young people. According to the authors, PEN traits do not represent pathological dimensions in themselves, but could lead to the development of abnormal conditions only in particular situations (Eysenck and Eysenck, 1991). In this perspective, neurosis and psychosis should be conceived as pathological exaggerations of the underlying traits of neuroticism and psychoticism (Eysenck and Eysenck, 1991; Mor, 2010).

Extraversion and neuroticism have been the first two dimensions included in the Eysenck’s model and were conceptualized as orthogonal continua (Eysenck and Eysenck, 1964, 1991). The neuroticism dimension describes a trait opposed to emotional stability, and defines the degree to which a person is predisposed to experience negative affect (Eysenck and Eysenck, 1964, 1991; Mor, 2010). Individuals with high levels of this trait tend to be worried, apprehensive, moody, fed-up, and irritable (Eysenck and Eysenck, 1991; Eysenck and Barrett, 2013). Extraversion is the second dimension included in the model and depicts sociable, carefree,

friendly, convivial, easygoing, and impulsive individuals. This trait is opposed to introversion which, in contrast, defines individuals introspective, quiet, serious, and reserved (Eysenck and Eysenck, 1975, 1991; Eysenck and Barrett, 2013). The third dimension included in the Eysenck's model has been psychoticism, or toughmindedness. The typical toughminded is an individual hostile, aggressive, untrusting, cold, unemotional, rude, lacking in human feelings, and unfriendly. On the opposite pole of the continuum, there are individuals with well-adjusted personality, agreeable, empathic, tolerant, conscientious, open-minded, friendly, and warm (Eysenck and Eysenck, 1975, 1991; Eysenck and Barrett, 2013).

Over the years, a series of instruments has been developed for the assessment of PEN traits on both young and adult people (e.g., Eysenck and Eysenck, 1964, 1975; Eysenck et al., 1985). These instruments also included a Lie (L) scale, which measures dissimulation and the tendency to deceive (Eysenck and Eysenck, 1964). Several contributions have been offered for the refinement of the psychometric properties of Eysenck's questionnaires, as well as for the development of brief versions (Eysenck et al., 1985; Francis and Pearson, 1988; Corulla, 1990; Francis et al., 1992; Francis, 1996). The psychometric properties and factor structure of all these instruments have been investigated in cross cultural research (e.g., Hosokawa and Ohyama, 1993; Maltby and Talley, 1998; Forrest et al., 2000; Qian et al., 2000; Scholte and De Bruyn, 2001; Aluja et al., 2003; Alexopoulos and Kalaitzidis, 2004; Dazzi et al., 2004; Francis et al., 2006; Karanci et al., 2006; Tiwari et al., 2009; Picconi et al., 2018). Unidimensionality of N and L scales has been widely supported in literature (e.g., Lajunen and Scherler, 1999; Ferrando, 2001; Ferrando and Chico, 2001; Ferrando and Anguiano-Carrasco, 2009; Dazzi, 2011). Contrasting results have been found concerning E scale: There are several studies supporting the unidimensionality of this scale (e.g., Rocklin and Revelle, 1981; Ferrando and Chico, 2001; Dazzi, 2011), but there is also some evidence suggesting the presence of two dimensions (Eysenck and Eysenck, 1963; Vidotto et al., 2008). Finally, there is large agreement in the literature that P scale comprises different facets (e.g., Howarth, 1986; Roger and Morris, 1991), which nevertheless contribute to a unique dimension (Chico and Ferrando, 1995; Dazzi, 2011).

Eysenck's instruments have been extensively employed for clinical, forensic, educational, and organizational purposes (e.g., Nyborg, 1997; Judge et al., 2000; Wood and Newton, 2003; Laidra et al., 2007; Smillie et al., 2009; Almiro et al., 2016), and all scales showed significant relations with a variety of psychologically and clinically relevant constructs and behaviors. Research, for instance, suggests that individuals with high levels of neuroticism may experience symptoms of anxiety and depression (e.g., Eysenck, 1991; Saklofske et al., 1995; del Barrio et al., 1997; Dazzi et al., 2004; Jylhä and Isometsä, 2006), and may also be more likely exposed to stress and health problems (e.g., Denney and Frisch, 1981; Huang et al., 2015; Bergomi et al., 2017). In contrast, extraversion appears to be mainly linked to adaptive social behavior, mental well-being, happiness, and life satisfaction (e.g., Lu, 1995; Mor, 2010; Gale et al., 2013). Moreover, this trait has been found to be negatively related to symptoms of anxiety and depression, to self-reported mental disorder and to health

care use for psychiatric reasons (e.g., del Barrio et al., 1997; Jylhä and Isometsä, 2006). Finally, psychoticism has been often cited in relation to inappropriate social behaviors, such as unsafe sexual habits, heavy drinking, criminal behavior, dysfunctional impulsivity, gambling, and drug abuse (e.g., Barnes et al., 1984; Blaszczyński et al., 1985; Bogaert, 1993; Lodhi and Thakur, 1993; Francis, 1996; Conrad et al., 1997; Grau and Ortet, 1999; Hoyle et al., 2000; Chico et al., 2003; Heaven et al., 2004; Gudgeon et al., 2005; Colledani, 2018).

The short form of the Eysenck Personality Questionnaire-Revised (EPQ-R; Eysenck et al., 1985; Eysenck and Eysenck, 1991) includes 48 items (out of 100 of the EPQ-R), 12 per each of the four dimensions. This version of the instrument has been translated in several languages and is widely used, across different countries, for scientific and clinical purposes (Hosokawa and Ohyama, 1993; Aluja et al., 2003; Alexopoulos and Kalaitzidis, 2004; Dazzi et al., 2004; Francis et al., 2006; Tiwari et al., 2009; Sanavio et al., 2013). However, it suffers from the same drawbacks of the full-length version. In particular, P scale exhibited poor reliability with a restricted range of scores and a strong positive skewness (Bishop, 1977; Block, 1977; Claridge, 1981; Hosokawa and Ohyama, 1993; Katz and Francis, 2000; Alexopoulos and Kalaitzidis, 2004). In addition, several items showed differential item functioning (DIF) across gender (Eysenck et al., 1985; Eysenck and Eysenck, 1991; Lynn and Martin, 1997; Forrest et al., 2000; Karanci et al., 2006; Escorial and Navas, 2007), which makes the comparison between groups questionable.

A better selection of the items from the full-length version of the instrument could allow for reducing some of the aforementioned drawbacks. The present work aims at developing a new version of the short form of the EPQ-R with improved psychometric properties.

Item response theory (IRT; Bock, 1997; Thissen and Steinberg, 2009) is one of the most promising approaches to this aim. There are several successful applications of IRT for the development and validation of measurement scales (see, Da Dalt et al., 2013, 2015; Balsamo et al., 2014; Anselmi et al., 2015; Zanon et al., 2016; Sotgiu et al., 2018). Moreover, compared with classical test theory, IRT was found to provide more diagnostic information useful for the development of brief scales (Spence et al., 2012; Bortolotti et al., 2013; Petrillo et al., 2015). IRT allows for identifying the items that are best at discriminating different levels of the latent trait of interest, while ensuring that the entire trait continuum is covered. Selecting these items can result in a brief version of the scale that produces scores very similar to those obtained with the full-length scale and has the same external validity (i.e., the same correlations with other constructs; Reise and Henson, 2000; Spence et al., 2012). Moreover, IRT allows for detecting items that are unclear, ambiguous, or which exhibit DIF. These items should be not included in the brief scale. Despite advantages offered by IRT, only a few studies employed this approach for the refinement of Eysenck's instruments (e.g., Ferrando, 2001; Ferrando and Chico, 2001; Escorial and Navas, 2007; Maij-de Meij et al., 2008). Recently, Colledani et al. (2018) used IRT for developing a new version of the abbreviated form of the Junior EPQ-R (6 items per scale). The new version outperformed the original one on several aspects.

This work includes two main studies. In Study 1, a series of analyses were performed on the responses to the full-length version of the EPQ-R in order to select the 48 items (12 per each scale) with the best psychometric properties. In Study 2, the functioning of the new short form was tested in a new data sample. Reliability, validity and factor structure were examined. Relationships of the new scales with social desirability, the dimensions of the Five Factor Model (FFM), and clinically relevant constructs were verified.

STUDY 1

Participants

A total of 590 participants took part in the study (mean age = 36.69 years, $SD = 14.16$; from 18 to 75 years; 55.8% females). They were recruited from different Italian regions through convenience sampling. All participants were native Italian speakers and completed the questionnaire anonymously and voluntarily. All standards for research with human subjects were respected. Written informed consent was obtained from the participants. The project has been approved, now as later, by the Ethical Committee for the Psychological Research of the University of Padova since a prospective ethics approval was not required at the time when the research was conducted (Protocol n. 2622).

Instruments

The participants were presented with the Italian version of the EPQ-R (Dazzi et al., 2004; Dazzi, 2011). The instrument consists of 100 dichotomous items (yes/no), 32 for P scale (e.g., “Should people always respect the law?”; “Do you enjoy hurting people you love?”), 23 for E scale (e.g., “Do you enjoy meeting new people?”; “Can you get a party going?”), 24 for N scale (e.g., “Would you call yourself a nervous person,” “Are you often troubled about feelings of guilt?”), and 21 for L scale (e.g., “Are all your habits good and desirable ones?”; “Have you ever cheated at a game?”). Administration of the questionnaire was individual and paper-and-pencil.

The Italian version of the questionnaire has good reliability and the four-factor structure was confirmed ($\alpha = 0.67, 0.78, 0.85,$ and 0.75 for P, E, N, and L scales, respectively; Dazzi et al., 2004; Dazzi, 2011). The reliability found in the current sample ($\alpha = 0.60, 0.79, 0.85,$ and 0.77 for P, E, N, and L scales) is in line with literature.

Studies in the Italian context aimed also to test the factor structure and the psychometric characteristics of the short version of the instrument (Dazzi et al., 2004). Consistently with cross-cultural findings, results supported the four-factor structure of the instrument and showed reliability coefficients satisfactory for E, N, and L scales, while lower for P ($\alpha = 0.37, 0.77, 0.83,$ and 0.70 for P, E, N, and L, respectively; Dazzi et al., 2004). The reliability found in the current sample ($\alpha = 0.40, 0.73, 0.83,$ and 0.73 for P, E, N, and L scales) is in line with literature.

Analysis Strategy

The two-parameter logistic (2PL) model (see Thissen and Steinberg, 2009) was separately estimated on the responses to each of the four scales of the questionnaire. This model describes

the probability that a subject endorses a certain item as a function of the latent trait level of the subject (parameter θ), the “endorsability” level of the item (i.e., the ease of providing a “yes” response to that item; parameter ϵ), and the capability of the item in differentiating subjects with different trait levels (parameter δ). In the case of the P scale, for instance, the greater the value of parameter θ , the greater the level of psychoticism of the subject; the greater the value of parameter ϵ , the greater the ease of responding “yes” to the item (i.e., of providing a response that is indicative of the presence of psychoticism); the greater the value of parameter δ , the greater the capability of the item in differentiating between subjects with different levels of psychoticism. All the analyses were run using the packages “difR” (Magis et al., 2016) and “ltm” (Rizopoulos, 2012) for the statistical environment R (R Core Team, 2016).

The 2PL assumes unidimensionality of the scales. Confirmatory factor analyses were run on the data of each of the four scales (for a reasonable fit, $CFI \geq 0.90$, $RMSEA < 0.08$; see Hu and Bentler, 1999; Marsh et al., 2004; Brown, 2006). These analyses confirmed the unidimensionality of N [$\chi^2_{(252)} = 1046.791, p \leq 0.001$; $RMSEA = 0.073$; $CFI = 0.919$] and L [$\chi^2_{(189)} = 532.901, p \leq 0.001$; $RMSEA = 0.056$; $CFI = 0.900$]. Fit indices of E scale were close to acceptance [$\chi^2_{(230)} = 808.417, p \leq 0.001$; $RMSEA = 0.065$; $CFI = 0.890$]. The unidimensional model did not fit the data of P scale [$\chi^2_{(464)} = 1841.233, p \leq 0.001$; $RMSEA = 0.071$; $CFI = 0.467$]. An exploratory factor analysis on this scale suggests a four-factor solution with 7 items out of 32 exhibiting cross-loadings. In line with literature (e.g., Howarth, 1986; Roger and Morris, 1991; Chico and Ferrando, 1995; Dazzi, 2011), this result confirms that P scale defines a complex and multifaceted construct.

Item Selection for the New Short Scales

DIF and item fit statistics were used to identify the items with the poorest psychometric properties that were not included in the new short scales.

Three item fit statistics were used: *infit*, *outfit* (Wright and Masters, 1982), and the index suggested by Bock (1972). *Infit* and *outfit* are two χ^2 -based statistics, the former being effective in detecting unexpected responses to items close to a subject's trait level, the latter being effective in detecting unexpected responses to items far from the subject's trait level. In this work, items with *infit* and/or *outfit* higher than 1.4 (Wright and Linacre, 1994) were considered misfitting and not included in the new short scales. The index suggested by Bock involves grouping subjects into n categories on the basis of their latent trait level, and observed and expected proportions of subjects endorsing the item for each group are compared (Bock, 1972; Reise, 1990). In this work, subjects were grouped into four categories and the items which displayed a medium ($0.3 \leq \Phi < 0.5$) to large ($\Phi \geq 0.5$) effect size (Cohen, 1988) were not selected for inclusion in the new questionnaire.

Items exhibiting gender DIF were also excluded from the new questionnaire. Both uniform and non-uniform DIF were considered. The former is a systematic bias expressing a different probability of endorsing an item for the members of a specific group. The latter is a non-systematic bias which varies with the

latent trait level. Females were used as reference group. Effect sizes of uniform and non-uniform DIF were evaluated by the R^2 difference test (Nagelkerke, 1991; Gómez-Benito et al., 2009), with values higher than 0.035 denoting moderate DIF and values higher than 0.07 denoting strong DIF (Jodoin and Gierl, 2001; Magis et al., 2016).

Parameters ϵ and δ were examined to select, among the remaining items, those that allow for covering the entire trait continuum and with the greatest discrimination level.

Assessment of the Psychometric Characteristics of the New Short Scales

Reliability and validity of the newly developed PEN-L scales were evaluated and compared with those of the original short

scales. Reliability was evaluated through Cronbach's α and test information function (TIF). TIF tells us how well the test measures the latent trait levels over the entire range of interest (Baker, 2001; Petrillo et al., 2015). The larger the value of TIF, the greater the accuracy with which the latent trait levels are measured. TIF depends on the latent trait range under consideration and on the number of items in the test (Baker, 2001). In this work, the old and new short scales had the same length (12 items), and TIF was defined on the same range of latent trait levels (-5 to 5). Validity was evaluated using a bias index and the correlation between scores obtained with full-length and short scales. The bias index was computed as the average difference (in absolute terms) between the parameters θ estimated on the full-length scales and those estimated on the short scales.

TABLE 1 | Easiness (ϵ) and discrimination (δ) parameters for the 32 items of the Psychoticism scale.

New form	Original form	Item n.	Text	ϵ	δ
		99	Would you feel very sorry for an animal caught in a trap?	-11.873	-0.121
		56	Do most things taste the same to you?	-10.789	-0.195
	✓	79	Do you try not to be rude to people?	-7.722	-0.354
		95	Do people tell you a lot of lies?	-2.538	-0.200
✓	✓	29	Do you prefer to go your own way rather than act by the rules?	-0.855	0.939
		9	Do you give money to charities?	-0.395	0.373
✓	✓	88	Is it better to follow society's rules than go your own way?	-0.045	1.247
✓		42	Have you often gone against your parents' wishes?	0.466	0.703
	✓	5	Do you take much notice of what people think?	0.502	0.150
	✓	75	Do you think people spend too much time safeguarding their future with savings and insurance?	0.713	0.313
✓		18	Should people always respect the law?	0.750	1.260
✓		64	Do you like to arrive at appointments in plenty of time?	0.953	0.846
✓		21	Are good manners very important?	0.993	1.580
✓	✓	41	Do good manners and cleanliness matter much to you?	1.439	1.763
		2	Do you stop to think things over before doing anything?	1.448	0.569
✓		81	Do you generally 'look before you leap'?	1.807	0.708
✓		96	Do you believe one has special duties to one's family?	2.404	1.249
✓	✓	25	Would you take drugs which may have strange or dangerous effects?	2.725	1.138
	✓	48	Do you think marriage is old-fashioned and should be done away with?	3.343	0.552
		34	Do you have enemies who want to harm you?	3.734	0.367
✓		30	Do you enjoy hurting people you love?	3.775	0.910
	✓	59	Does it worry you if you know there are mistakes in your work?	4.024	0.538
	✓	7	Would being in debt worry you?	4.028	0.446
		73	Are there several people who keep trying to avoid you?	4.119	0.596
		68	Is (or was) your mother a good woman?	4.774	0.527
		85	Can you on the whole trust people to tell the truth?	4.986	-0.128
		12	Would it upset you a lot to see a child or an animal suffer?	5.307	0.750
	✓	54	Do you enjoy co-operating with others?	5.517	0.446
		14	Would it upset you a lot to see a child or an animal suffer?	5.623	-0.044
	✓	91	Would you like other people to be afraid of you?	6.036	0.242
✓		37	Do you have many friends?	6.638	0.515
		50	Are you more easy-going about right and wrong than most people?	8.580	-0.089

The items are ordered by increasing easiness. The items included in the new and in the original short forms are marked by "✓."

Low biases suggest that the latent trait estimates obtained with the short scales approximate those of the full-length versions. In addition, the correlations between scores obtained with the full-length and short scales were computed and corrected for common items using the Levy's (1967) method.

Results

Three of the 32 items of P scale exhibited uniform and non-uniform gender DIF of moderate (Items 68 and 91) or strong (Item 12) size. Fit statistics were adequate for all the items. From the remaining 29 items, 12 were selected taking into account their parameters ϵ and δ . This resulted in a new short scale, that differed from the original one for eight items (see **Table 1**). Specifically, Item 91 was changed because it showed uniform and non-uniform gender DIF of moderate size. These modifications allowed for obtaining a new scale with increased reliability (α increased from 0.40 to 0.62; TIF increased from 8.13 to 12.86) and with scores that better approximate those obtained with the full-length scale (bias decreased from 0.37 to 0.18, corrected correlation increased from 0.47 to 0.52). It is worth noting that Cronbach's α of the new 12-item scale (0.62) largely resembles that of the full 32-item scale (0.60).

Regarding the 23 items of E scale, only Item 55 exhibited uniform gender DIF of moderate size and no item showed misfit. Selecting 12 items upon the basis of their parameters ϵ and δ , we obtained a new E scale that differed from the original one for three items (see **Table 2**). The differences in reliability and validity of the new and original scales were small in size,

nevertheless in favor of the new version (α increased from 0.73 to 0.75; TIF increased from 16.62 to 16.83; bias decreased from 0.21 to 0.19; corrected correlation increased from 0.74 to 0.77).

Concerning N and L scales, no one item exhibited gender DIF or misfit. Therefore, items were selected considering their ϵ and δ parameters. For both scales, the new versions differed from the original ones for two items (see **Tables 3, 4**). Item 35 was present in the previous version of the N scale but it has not been included in the new one because of its redundant content. Reliability of the new scales largely resembles that of the original versions ($\alpha = 0.83, 0.82$; TIF = 20.86, 20.80 for original and new N scale, respectively; $\alpha = 0.73, 0.74$; TIF = 13.86, 14.15 for original and new L scale, respectively). Concerning N scale, a slight decrease of bias was observed (from 0.22 to 0.16). The other indexes remained substantially unchanged (bias = 0.20, 0.18 for original and new L scale, respectively; corrected correlation = 0.74, 0.75 for original and new L scale, respectively; 0.83, 0.84, for original and new N scale, respectively).

Discussion

This study aimed at developing a new short version of the EPQ-R with improved psychometric characteristics. IRT based statistics allowed the identification of 48 items without gender DIF or misfit, well discriminating, and well distributed along the four latent traits continua. The new version of the P scale differs from the original one for eight items (out of 12), E scale for three, and N and L only for two. The largest improvement was reached for P scale, which in literature was found to perform less well than the

TABLE 2 | Easiness (ϵ) and discrimination (δ) parameters for the 23 items of the Extraversion scale.

New form	Original form	Item n.	Text	ϵ	δ
	✓	47	Are you mostly quiet when you are with other people?	-5.623	-0.395
✓	✓	20	Do you enjoy meeting new people?	-1.914	1.613
	✓	11	Are you rather lively?	-1.605	1.311
✓	✓	58	Do you like mixing with people?	-1.592	2.182
		33	Do you prefer reading to meeting people?	-1.246	0.973
	✓	94	Do other people think of you as being very lively	-1.180	0.805
✓		28	Do you like going out a lot?	-1.105	1.193
		72	Do you often take on more activities than you have time for?	-0.832	0.570
		69	Do you often make decisions on the spur of the moment?	-0.755	0.489
		36	Do you have many friends?	-0.652	1.525
✓	✓	78	Can you get a party going?	-0.636	1.418
✓		67	Do you like doing things in which you have to act quickly?	-0.608	0.783
✓	✓	16	Can you usually let yourself go and enjoy yourself at a lively party?	-0.536	1.676
✓	✓	6	Are you a talkative person?	-0.468	1.211
		55	Do you like telling jokes and funny stories to your friends?	-0.365	0.705
		63	Do you nearly always have a 'ready answer' when people talk to you?	-0.230	0.401
✓	✓	45	Do you usually take the initiative in making new friends?	-0.171	1.741
✓	✓	24	Do you tend to keep in the background on social occasions?	0.151	1.658
✓	✓	51	Can you easily get some life into a rather dull party?	0.402	1.729
		40	Would you call yourself happy-go-lucky?	0.753	0.944
✓		61	Have people said that you sometimes act too rashly?	0.934	0.582
		1	Do you have many different hobbies?	0.974	0.519
✓	✓	90	Do you like plenty of bustle and excitement around you?	1.648	0.884

The items are ordered by increasing easiness. The items included in the new and in the original short forms are marked by "✓."

TABLE 3 | Easiness (ϵ) and discrimination (δ) parameters for the 24 items of the Neuroticism scale.

New form	Original form	Item n.	Text	ϵ	δ
		87	Are you easily hurt when people find fault with you of the work you do?	-1.652	0.930
		97	Are you touchy about some things?	-1.528	0.696
✓	✓	22	Are your feelings easily hurt?	-1.486	1.143
		92	Are you sometimes bubbling over with energy and sometimes very sluggish?	-1.186	1.025
✓	✓	80	Do you worry too long after an embarrassing experience?	-1.139	1.128
		13	Do you often worry about things you should not have done or said?	-1.027	1.084
		43	Do you worry about awful things that might happen?	-0.819	0.497
		74	Do you worry a lot about your looks?	-0.460	0.484
✓	✓	3	Does your mood often go up and down?	-0.319	1.785
		100	When your temper rises, do you find it difficult to control?	-0.164	0.848
✓	✓	31	Are you often troubled about feelings of guilt?	-0.114	1.445
✓		65	Have you often felt listless and tired for no reason?	-0.003	1.720
	✓	8	Do you ever feel "just miserable" for no reason?	0.007	1.392
✓	✓	17	Are you an irritable person?	0.225	1.527
	✓	35	Would you call yourself a nervous person?	0.250	2.582
✓	✓	46	Would you call yourself tense or "highly-strung"?	0.308	2.928
✓	✓	84	Do you often feel lonely?	0.308	1.669
✓	✓	83	Do you suffer from "nerves"?	0.394	2.833
✓	✓	26	Do you often feel "fed-up"?	0.591	1.713
		76	Have you ever wished that you were dead?	0.976	1.102
✓		70	Do you often feel life is very dull?	0.990	1.273
		60	Do you suffer from sleeplessness?	2.443	0.691
✓	✓	38	Are you a worrier?	2.948	0.721
		52	Do you worry about your health?	8.462	-0.132

The items are ordered by increasing easiness. The items included in the new and in the original short forms are marked by "✓."

other three scales (e.g., Bishop, 1977; Block, 1977; Claridge, 1981). In particular, the new version is not affected by gender DIF and outperforms the original one for reliability and approximation of the scores obtained with the full-length form. The new versions of the other three scales performed as well as, or slightly better than the original ones. Although small in size, these improvements are valuable taking into account that were obtained by substituting a small number of items and reducing content redundancy.

STUDY 2

This study aimed at investigating the functioning of the new version of the short EPQ-R on a new data set. Other to reliability and factor structure, construct validity was evaluated by taking into account relationships with social desirability, the dimensions of the FFM, and measures of anxiety and depression.

Participants

Participants were 300 native Italian speakers aged between 18 and 65 (mean age = 29.28, $SD = 10.38$; 60.2% females). They were recruited from different Italian regions using convenience sampling. All participants were presented with the new version of the short EPQ-R, whereas a subsample of 158 participants (mean

age = 34.73, $SD = 9.88$; 68.7% females) also received the other measures. The participation to the study was anonymous and voluntary, and all standards for research with human subjects were respected. Written informed consent was obtained from the participants. The project has been approved, now as later, by the Ethical Committee for the Psychological Research of the University of Padova since a prospective ethics approval was not required at the time when the research was conducted (Protocol n. 2622).

Instruments

The new form of the short EPQ-R devised in Study 1 was administered to all participants.

The five traits of the FFM of personality (i.e., extraversion, agreeableness, conscientiousness, emotional stability, and openness) were measured through the Italian version (Ubbiali et al., 2013; Chiorri et al., 2016) of the Big Five Inventory (BFI; John et al., 2008). The questionnaire consists of 44 items answered on a five-point Likert scale (from 1 "Strongly disagree" to 5 "Strongly agree"; e.g., "I see myself as someone who is full of energy" for extraversion; "I see myself as someone who is helpful and unselfish with others" for agreeableness; "I see myself as someone who perseveres until the task is finished"

TABLE 4 | Easiness (ϵ) and discrimination (δ) parameters for the 21 items of the Lie scale.

New form	Original form	Item n.	Text	ϵ	δ
		98	Are you always willing to admit it when you have made a mistake?	-4.705	0.427
	✓	57	As a child were you ever cheeky to your parents?	-2.153	0.536
		62	Do you always wash before a meal?	-2.120	0.614
✓		4	Have you ever taken the praise for something you knew someone else had really done?	-2.057	0.678
✓	✓	19	Have you ever blamed someone for doing something you knew was really your fault?	-1.687	0.869
✓	✓	71	Have you ever taken advantage of someone?	-0.909	1.638
✓		32	Do you sometimes talk about things you know nothing about?	-0.733	1.188
	✓	15	If you say you will do something, do you always keep your promise no matter how inconvenient it might be?	-0.693	0.898
✓	✓	66	Have you ever cheated at a game?	-0.508	1.183
✓	✓	44	Have you ever broken or lost something belonging to someone else?	-0.172	1.299
✓	✓	27	Have you ever taken anything (even a pin or button) that belonged to someone else?	-0.149	1.382
		49	Do you sometimes boast a little?	0.164	0.996
✓	✓	10	Were you ever greedy by helping yourself to more than your share of anything?	0.303	1.201
✓	✓	53	Have you ever said anything bad or nasty about anyone?	0.419	1.513
		77	Would you dodge paying taxes if you were sure you could never be found out?	0.421	0.473
✓	✓	86	Do you always practice what you preach?	0.423	1.052
		89	Have you ever been late for an appointment or work?	0.451	1.094
		39	As a child did you do as you were told immediately and without grumbling?	0.691	0.717
✓	✓	23	Are all your habits good and desirable ones?	0.712	1.080
✓	✓	93	Do you sometimes put off until tomorrow what you ought to do today?	1.155	1.212
		82	Have you ever insisted on having your own way?	5.265	0.173

The items are ordered by increasing easiness. The items included in the new and in the original short forms are marked by "✓."

for conscientiousness; "I see myself as someone who worries a lot" for emotional stability; "I see myself as someone who is ingenious, a deep thinker" for openness). Convincing evidence was found concerning construct validity, factor structure, gender invariance, and reliability (α from 0.75 to 0.86; Ubbiali et al., 2013; Chiorri et al., 2016; α from 0.73 to 0.83 in the current sample).

The Impression Management (IM) scale of the Italian brief version (Bobbio and Manganelli, 2011) of the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1991) was also administered. The scale comprises 8 items answered on a six-point Likert scale (from 1 "Strongly disagree" to 6 "Strongly agree") and assesses the conscious tendency of individuals to provide positively inflated self-descriptions (e.g., "I have never dropped litter on the street"). Internal consistency of the scale ranges from 0.73 to 0.81 (Bobbio and Manganelli, 2011; in the current sample, $\alpha = 0.75$).

The trait scale of the State-Trait Anxiety Inventory (STAI-Y; Spielberger et al., 1983; Pedrabissi and Santinello, 1989)

was used to evaluate anxiety. The scale comprises 20 items answered on a four-point Likert scale (from 1 "Not at all" to 4 "Very much"). The instrument evaluates the tendency of people to experience general anxiety and the relatively stable predisposition to view stressful situations as threatening (e.g., "I am regretful"). The Italian version of the questionnaire showed adequate validity and reliability (α from 0.85 and 0.90; Pedrabissi and Santinello, 1989; in the current sample, $\alpha = 0.92$).

Finally, the Italian version of the Patient Health Questionnaire-9 (PHQ-9; Spitzer et al., 1999; Kroenke et al., 2001) was used to evaluate depressive symptoms. The questionnaire is a self-administered instrument and assesses the nine DSM-IV (American Psychiatric Association, 2000) criteria for depression. Respondents are asked to evaluate the presence of depressive symptoms over the last 2 weeks through nine items scored on a four-point Likert scale (from 0 "Not at all" to 3 "Nearly every day"; e.g., "Feeling tired or having little energy"). This instrument showed adequate reliability (α from

0.86 to 0.89), and good sensitivity and specificity (see Kroenke et al., 2001). In the current sample, α equals 0.81.

Analysis Strategy

Reliability of the new version of the short EPQ-R was tested through Cronbach's α . Construct validity was evaluated by computing convergent validity coefficients and by analyzing the factor structure of the instrument.

Convergent validity was evaluated considering correlations between the four PEN-L traits, the five dimensions of FFM, social desirability, and indexes of depression and trait anxiety. According with literature, L scores are expected to positively correlate with the IM scale of the BIDR (e.g., Gillings and Joseph, 1996), while PEN traits are expected to correlate with BFI scales, depression and trait anxiety. In particular, positive correlations are expected between E scores of the EPQ-R and the extraversion measure of the BFI, while negative correlations are expected between P scale and agreeableness and conscientiousness. Positive correlations are also expected between N scale of the EPQ-R and the neuroticism measure of the BFI (e.g., McCrae and Costa, 1985; Draycott and Kline, 1995; Saggino, 2000; Barbaranelli et al., 2003; Scholte and De Bruyn, 2004; Heaven et al., 2013). Neuroticism, in addition, is expected to positively correlate with indexes of anxiety and depression (STAI-Y; Spielberger et al., 1983; PHQ-9; Spitzer et al., 1999; Kroenke et al., 2001). In contrast, extraversion is expected to negatively correlate with these two clinical indexes.

An Exploratory Structural Equation Model (ESEM; Asparouhov and Muthén, 2009) was run to evaluate the factor structure. The ESEM framework represents an integration of confirmatory factor analysis (CFA), structural equation modeling (SEM), and exploratory factor analysis (EFA). ESEMs give access to all the common statistics of SEM/CFA but, at the same time, overcome the restrictions associated with the confirmatory approach. CFA fixes non-target loadings to zero and, therefore, it may be inadequate to handle complex and multifaceted constructs where many cross-loadings may be expected (Marsh et al., 2009, 2010, 2011, 2014). When this is the case, fit problems and upward-biased estimates of correlations between factors can be observed (Cole et al., 2007; Marsh and Hau, 2007; Marsh et al., 2010). As in EFA, ESEMs allow for the free estimation of cross-loadings between items and non-target

factors. In this work, ESEM was run using Mplus7 (Muthén and Muthén, 2012), and the WLSMV as estimator (weighted least squares mean and variance-adjusted). This method is recommended for binary or ordinal observed data (e.g., Flora and Curran, 2004; Brown, 2006) such as the dichotomous items of the EPQ-R. In the model, the 48 items were the indicators and four factors were modeled. The GEOMIN oblique rotation was used. To evaluate the goodness of fit of the model, several fit indexes were considered: χ^2 , Comparative Fit Index (CFI; Bentler, 1990), Weighted Root Mean Square Residual (WRMR; Yu, 2002), and Root Mean Square Error of Approximation (RMSEA; Browne and Cudeck, 1993) with its 90% confidence interval (90% CI) and the test of close fit (CFit; Browne and Cudeck, 1993). A solution fits the data well when χ^2 is non-significant ($p \geq 0.05$). Since this statistic is sensitive to sample size, the other fit measures were also considered. In particular, a solution fits the data well when CFI is close to 0.95 (0.90 to 0.95 for reasonable fit), WRMR is close to 1.0, and RMSEA is smaller than 0.06 (0.06 to 0.08 for reasonable fit) with CFit non-significant (see Hu and Bentler, 1999; Marsh et al., 2004; Brown, 2006).

Results

Cronbach's α coefficients were 0.55, 0.80, 0.81, and 0.70 for P, E, N, and L scales, respectively. These values were consistent with those of Study 1. Compared with the original version, the largest improvement was reached for P scale, as observed in Study 1.

Convergent validity coefficients are reported in **Table 5**. All the four PEN-L traits correlated in the expected direction with the considered constructs. E scale showed a strong positive relation with the extraversion measure of the BFI (0.727). P scale was negatively related to agreeableness (-0.323) and conscientiousness (-0.321). N scale was strongly correlated with neuroticism (0.709). Relations with anxiety and depression were also in the expected directions. N scale showed positive relations with scores of PHQ-9 (0.619) and STAI-Y (0.697), while moderate negative relations were found between these two indexes and E scale ($r = -0.409, -0.405$ for PHQ-9 and STAI-Y, respectively). Finally, L scale showed a strong positive correlation with the IM scale of the BIDR.

Results of the ESEM supported the four-factor structure of the instrument ($\chi^2_{(942)} = 1122.686, p < 0.001$; RMSEA = 0.025

TABLE 5 | Cronbach's α s and correlations between the four PEN-L traits, STAI-Y, PHQ-9, BIDR-IM, and the five BFI dimensions.

	α	Psychoticism	Extraversion	Neuroticism	Lie
STAI-Y	0.92	0.092	-0.409^{***}	0.697***	-0.259^{**}
PHQ-9	0.81	0.099	-0.405^{***}	0.619***	-0.202^*
BIDR-IM	0.75	-0.429^{***}	0.195*	-0.293^{***}	0.561***
BFI-Extraversion	0.79	0.058	0.727***	-0.369^{***}	0.132
BFI-Agreeableness	0.73	-0.323^{***}	0.317***	-0.241^{**}	0.242**
BFI-Conscientiousness	0.83	-0.321^{***}	0.220**	-0.385^{***}	0.334***
BFI-Neuroticism	0.80	0.021	-0.363^{***}	0.709***	-0.110
BFI-Openness	0.78	0.094	0.262**	-0.126	0.065

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

TABLE 6 | Exploratory structural equation modeling.

Item	Psychoticism	Extraversion	Neuroticism	Lie
29	0.279**	-0.058	0.303**	-0.089
88	0.317***	0.066	0.101	-0.123
96	0.444***	0.065	0.017	0.050
41	0.913***	-0.076	-0.058	0.032
25	0.384**	-0.193	-0.119	-0.551***
42	0.312**	0.045	0.049	-0.304
37	0.596***	0.084	0.141	-0.255
64	0.457***	0.118	0.051	0.117
21	0.728***	-0.148	-0.050	0.058
30	0.743***	-0.236	-0.352*	-0.001
81	0.546***	0.223	0.286**	0.004
18	0.531***	-0.075	-0.063	-0.342**
16	-0.065	0.796***	-0.152	-0.141
78	0.083	0.482***	0.037	0.151
20	-0.081	0.775***	-0.067	0.176
51	0.182	0.605***	-0.014	0.014
67	0.075	0.317***	-0.052	-0.027
6	0.090	0.374***	-0.060	0.030
28	0.180	0.678***	0.033	0.154
58	-0.077	0.865***	-0.201	0.060
90	-0.004	0.732***	-0.017	-0.242
45	0.100	0.600***	0.020	0.100
24	-0.067	0.817***	-0.164	-0.094
61	0.396***	0.499***	0.414***	-0.038
65	-0.034	-0.162	0.540***	-0.184*
17	0.065	0.136	0.701***	0.187*
46	0.024	0.019	0.844***	0.245**
3	0.045	0.176	0.798***	-0.007
38	-0.217*	-0.151	0.658***	-0.016
80	-0.262*	-0.296*	0.313***	-0.040
26	-0.201*	-0.131	0.683***	-0.107
70	0.002	-0.191	0.619***	-0.070
31	-0.135	-0.140	0.605***	-0.029
22	-0.168	-0.114	0.479***	0.199*
38	-0.156	-0.210	0.558***	0.048
83	0.045	-0.034	0.910***	0.323
66	-0.094	-0.089	0.011	0.664***
27	-0.385*	0.047	-0.109	0.318**
32	0.018	-0.319**	-0.374***	0.235*
86	0.110	-0.057	-0.243**	0.365***
4	-0.402**	-0.005	-0.049	0.483***
93	0.000	-0.301**	-0.201*	0.449***
23	-0.144	0.069	-0.167	0.369***
53	0.060	0.043	-0.107	0.639***
10	0.054	0.050	-0.011	0.566***
44	-0.290*	-0.113	-0.066	0.502***
71	-0.359***	0.060	0.026	0.476***
19	-0.302**	-0.054	-0.326**	0.191*
CORRELATION WITH				
Extraversion	0.043			
Neuroticism	0.04	-0.198***		
Lie	-0.125	0.081	-0.258**	

Standardized factor loadings and factor correlations (N = 300).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Bolded coefficients are target loadings.

[0.019, 0.031]; CFit \cong 1.000; CFI = 0.930; WRMR = 0.864]. The model is represented in **Table 6**. All items loaded on the intended factor and cross-loadings were, in general, lower than those observed on the target-factor.

Discussion

The analyses performed in this study provide further evidence concerning the adequate psychometric properties of the new short form of the EPQ-R. Concerning reliability, results are in line with those of Study 1 and confirm that, compared with the original version, the largest improvement was observed for P scale. Concerning validity, both the factor structure of the instrument and its convergent validity are supported.

FINAL REMARKS

This work aimed at developing a new and improved version of the short form of the EPQ-R. This instrument is well-known and widely used in different settings. However, some weaknesses have been pointed out, especially for P scale (e.g., Bishop, 1977; Block, 1977; Claridge, 1981). IRT approach was used to develop the new instrument. This approach allowed for removing items with misfit or gender DIF, and for identifying items that were best at discriminating different levels of traits, while ensuring that the respective continua were covered. As suggested in literature, following these criteria for item selection should lead to a short scale with the same psychometric properties of the full-length instrument (Reise and Henson, 2000; Spence et al., 2012). In fact, results of this work show that the new short form of the EPQ-R approximated the scores obtained with the full-length form better than the original short version. In addition, convergent validity of the new scale was consistent with literature (e.g., Saklofske et al., 1995; Gillings and Joseph, 1996; del Barrio et al., 1997; Dazzi et al., 2004; Jylhä and Isometsä, 2006; Mor, 2010). The moderate to strong relationships between Eysenck's traits and clinical constructs provide further evidence toward the usefulness of assessing these traits in clinical settings.

A strength of the present work is that it provides a solution to some well-known drawbacks of the full-length EPQ-R and of its short form existing in the literature (Eysenck et al., 1985; Eysenck and Eysenck, 1991). The largest improvement was obtained for P scale. The new version is not affected by gender DIF and outperforms the original one for reliability and approximation of the full-length form. The new versions of the other three scales performed as well as the original ones, or slightly better. These improvements are small in size, yet notable considering that were obtained by substituting a small number of items and reducing content redundancy.

In the present work, separate analyses have been performed on each of the four scales by using a unidimensional IRT model. An alternative could have been examining the four scales at once through a multidimensional IRT (MIRT) model (see Haberman et al., 2008; Reckase, 2009). MIRT models offer some advantages over unidimensional IRT models. They could allow for better understanding the traits measured by an

instrument and how well individual items measure each of them (Ackerman, 1994). Moreover, MIRT models could provide a more precise estimation of scale reliability (Cheng et al., 2009) and item parameters (Finch, 2010). In the present work, some of these advantages are not very relevant. On the one hand, the factor structure of the EPQ-R has been widely tested and validated in the literature (e.g., Hosokawa and Ohyama, 1993; Maltby and Talley, 1998; Forrest et al., 2000; Qian et al., 2000; Scholte and De Bruyn, 2001; Aluja et al., 2003; Alexopoulos and Kalaitzidis, 2004; Dazzi et al., 2004; Francis et al., 2006; Karanci et al., 2006; Tiwari et al., 2009; Picconi et al., 2018). On the other hand, for scales whose length is analogous to that of the four EPQ-R scales (i.e., from 21 to 32 items), the unidimensional IRT models have been found to provide item parameter estimates whose precision exceeds or equals that of the estimates produced by the MIRT models (Finch, 2010). Finch (2010) investigated the precision of MIRT estimates on tests measuring a number of traits as small as two. For larger numbers of traits (e.g., the four traits of the EPQ-R), the number of parameters of a MIRT model increases considerably. Thus, the sample size of Study 1 (590 individuals) could have not been appropriate for performing a multidimensional analysis.

Concerning P scale, despite notable improvements, reliability remains rather low. This result, however, was expected. P scale, in fact, maybe because of its complex and clinical nature, is the most

problematic and controversial of the instrument (e.g., Eysenck et al., 1985). Future research, therefore, should try to develop a new pool of items effective in capturing the multifaced aspects of this trait.

In the present work, a new short version of the EPQ-R has been devised, which consists of 12 items per each of the four scales. An abbreviated form exists also in literature (Francis et al., 1992) that consists of only 6 items per scale. This abbreviated form suffers of the same weaknesses that have been pointed out for the other Eysenck's questionnaires. Future research should try to devise a new version of the abbreviated form by using the IRT approach.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

DC contributed to the conception and design of the study, conducted the research, performed the statistical analyses, and wrote the first draft of the manuscript. DC and PA wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

REFERENCES

- Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Appl. Meas. Educ.* 7, 255–278. doi: 10.1207/s15324818ame0704_1
- Alexopoulos, D. S., and Kalaitzidis, I. (2004). Psychometric properties of Eysenck Personality Questionnaire-Revised (EPQ-R) short scale in Greece. *Pers. Individ. Differ.* 37, 1205–1220. doi: 10.1016/j.paid.2003.12.005
- Almiro, P. A., Moura, O., and Simões, M. R. (2016). Psychometric properties of the European Portuguese version of the Eysenck Personality Questionnaire—Revised (EPQ-R). *Pers. Individ. Differ.* 88, 88–93. doi: 10.1016/j.paid.2015.08.050
- Aluja, A., Garcia, Ó., and Garcia, L. F. (2003). A psychometric analysis of the revised Eysenck Personality Questionnaire short scale. *Pers. Individ. Differ.* 35, 449–460. doi: 10.1016/S0191-8869(02)00206-4
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th Ed.). Washington, DC: American Psychiatric Association.
- Anselmi, P., Vidotto, G., Bettinardi, O., and Bertolotti, G. (2015). Measurement of change in health status with Rasch models. *Health Qual. Life Out.* 13:16. doi: 10.1186/s12955-014-0197-x
- Asparouhov, T., and Muthén, B. (2009). Exploratory structural equation modeling. *Struct. Equ. Modeling* 16, 397–438. doi: 10.1080/10705510903008204
- Baker, F. B. (2001). *The Basics of Item Response Theory* (2nd Ed.). Maryland: ERIC Clearinghouse on Assessment and Evaluation.
- Balsamo, M., Giampaglia, G., and Saggino, A. (2014). Building a new Rasch-based self-report inventory of depression. *Neuropsych. Dis. Treat.* 10, 153–165. doi: 10.2147/NDT.S53425
- Barbaranelli, C., Caprara, G. V., Rabasca, A., and Pastorelli, C. (2003). A questionnaire for measuring the Big Five in late childhood. *Pers. Individ. Differ.* 34, 645–664. doi: 10.1016/S0191-8869(02)00051-X
- Barnes, G. E., Malamuth, N. M., and Check, J. V. (1984). Personality and sexuality. *Pers. Individ. Differ.* 5, 159–172.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychol. Bull.* 107, 238–246.
- Bergomi, M., Modenese, A., Ferretti, E., Ferrari, A., Licitra, G., Vivoli, R., et al. (2017). Work-related stress and role of personality in a sample of Italian bus drivers. *Work* 57, 433–440. doi: 10.3233/WOR-172581
- Bishop, D. V. (1977). The P scale and psychosis. *J. Abnorm. Psychol.* 86, 127–134.
- Blaszczynski, A. P., Buhrich, N., and McConaghy, N. (1985). Pathological gamblers, heroin addicts and controls compared on the EPQ 'Addiction Scale'. *Br. J. Addict.* 80, 315–319.
- Block, J. (1977). P scale and psychosis: continued concerns. *J. Abnorm. Psychol.* 86, 431–434.
- Bobbio, A., and Manganelli, A. M. (2011). Measuring social desirability responding. A short version of Paulhus' BIDR 6. *TPM* 18, 117–135. doi: 10.4473/TPM.18.2.4
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* 37, 29–51. doi: 10.1007/BF02291411
- Bock, R. D. (1997). A brief history of item response theory. *Educ. Meas. Issues Pract.* 16, 21–33. doi: 10.1111/j.1745-3992.1997.tb00605.x
- Bogaert, A. F. (1993). Personality, delinquency, and sexuality: data from three Canadian samples. *Pers. Individ. Differ.* 15, 353–356.
- Bortolotti, S. L. V., Tezza, R., de Andrade, D. F., Bornia, A. C., and de Sousa Júnior, A. F. (2013). Relevance and advantages of using the item response theory. *Qual. Quant.* 47, 2341–2360. doi: 10.1007/s11135-012-9684-5
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Press.
- Browne, M. W., and Cudeck, R. (1993). "Alternative ways of assessing model fit," in *Testing Structural Equation Models*, eds. K. A. Bollen and J. S. Long (Newbury Park, CA: SAGE), 136–162.
- Cheng, Y. Y., Wang, W. C., and Ho, Y. H. (2009). Multidimensional Rasch analysis of a psychological test with multiple subtests: a statistical solution for the bandwidth–fidelity dilemma. *Educ. Psychol. Meas.* 69, 369–388. doi: 10.1177/0013164408323241

- Chico, E., and Ferrando, P. J. (1995). A psychometric evaluation of the revised P scale in delinquent and non-delinquent Spanish samples. *Pers. Individ. Differ.* 18, 331–337.
- Chico, E., Tous, J. M., Lorenzo-Seva, U., and Vigil-Colet, A. (2003). Spanish adaptation of Dickman's impulsivity inventory: its relationship to Eysenck's personality questionnaire. *Pers. Individ. Differ.* 35, 1883–1892. doi: 10.1016/S0191-8869(03)00037-0
- Chiorri, C., Marsh, H. W., Ubbiali, A., and Donati, D. (2016). Testing the factor structure and measurement invariance across gender of the Big Five inventory through exploratory structural equation modeling. *J. Pers. Assess.* 98, 88–99. doi: 10.1080/00223891.2015.1035381
- Claridge, G. S. (1981). "Psychoticism," in *Dimensions of Personality. Papers in Honour of H.J. Eysenck*, eds. H. J. Eysenck, and R. Lynn (Oxford, UK: Pergamon Press), 79–110.
- Cohen, J. (1988). *Statistical Power and Analysis for the Behavioral Sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Cole, D. A., Ciesla, J. A., and Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychol. Methods* 12, 381–398. doi: 10.1037/1082-989X.12.4.381
- Colledani, D. (2018). Psychometric properties and gender invariance for the Dickman Impulsivity Inventory. *Test. Psychometr. Methodol. Appl. Psychol.* 25, 49–61. doi: 10.4473/TPM25.1.3
- Colledani, D., Robusto, E., and Anselmi, P. (2018). Development of a new abbreviated form of the Junior Eysenck Personality Questionnaire-Revised. *Pers. Individ. Differ.* 120, 159–165. doi: 10.1016/j.paid.2017.08.037
- Conrad, P. J., Petersen, J. B., and Pihl, R. O. (1997). Disinhibited personality and sensitivity to alcohol reinforcement: independent correlates of drinking behavior in sons of alcoholics. *Alcohol. Clin. Exp. Res.* 21, 1320–1332.
- Corulla, W. J. (1990). A revised version of the psychoticism scale for children. *Pers. Individ. Differ.* 11, 65–76. doi: 10.1016/0191-8869(90)90169-R
- Da Dalt, L., Anselmi, P., Bressan, S., Carraro, S., Baraldi, E., Robusto, E., et al. (2013). A short questionnaire to assess pediatric resident's competencies: the validation process. *Ital. J. Pediatr.* 39:41. doi: 10.1186/1824-7288-39-41
- Da Dalt, L., Anselmi, P., Furlan, S., Carraro, S., Baraldi, E., Robusto, E., et al. (2015). Validating a set of tools designed to assess the perceived quality of training of pediatric residency programs. *Ital. J. Pediatr.* 41:2. doi: 10.1186/s13052-014-0106-2
- Dazzi, C. (2011). The Eysenck personality questionnaire-Revised (EPQ-R): A confirmation of the factorial structure in the Italian context. *Pers. Individ. Differ.* 50, 790–794. doi: 10.1016/j.paid.2010.12.032
- Dazzi, C., Pedrabissi, L., and Santinello, M. (2004). *Adattamento Italiano delle Scale di Personalità Eysenck per Adulti [Italian Adaptation of the Eysenck Personality Scales for Adults]*. Firenze, IT: Organizzazioni Speciali.
- del Barrio, V., Moreno-Rosset, C., López-Martínez, R., and Olmedo, M. (1997). Anxiety, depression and personality structure. *Pers. Individ. Differ.* 23, 327–335.
- Denney, D. R., and Frisch, M. B. (1981). The role of neuroticism in relation to life stress and illness. *J. Psychosom. Res.* 25, 303–307.
- Draycott, S. G., and Kline, P. (1995). The big three or the big five—The EPQ-R vs the NEOPI: a research note, replication and elaboration. *Pers. Individ. Differ.* 18, 801–804. doi: 10.1016/0191-8869(95)00010-4
- Escorial, S., and Navas, M. J. (2007). Analysis of the gender variable in the Eysenck Personality Questionnaire-revised scales using differential item functioning techniques. *Educ. Psychol. Meas.* 67, 990–1001. doi: 10.1177/0013164406299108
- Eysenck, H. J. (1991). Neuroticism, anxiety, and depression. *Psychol. Inq.* 2, 75–76.
- Eysenck, H. J., and Eysenck, S. B. G. (1964). *Manual of the Eysenck Personality Inventory*. London, UK: University of London Press.
- Eysenck, H. J., and Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (Adult and Junior)*. London, UK: Hodder and Stoughton.
- Eysenck, H. J., and Eysenck, S. B. G. (1991). *Manual of the Eysenck Personality Scale (Adults)*. London, UK: Hodder and Stoughton.
- Eysenck, S., and Barrett, P. (2013). Re-introduction to cross-cultural studies of the EPQ. *Pers. Individ. Differ.* 54, 485–489. doi: 10.1016/j.paid.2012.09.022
- Eysenck, S. B., Eysenck, H. J., and Barrett, P. (1985). A revised version of the psychoticism scale. *Pers. Individ. Differ.* 6, 21–29. doi: 10.1016/0191-8869(85)90026-1
- Eysenck, S. B. G., and Eysenck, H. J. (1963). On the dual nature of extraversion. *Brit. J. Clin. Psychol.* 2, 46–55.
- Ferrando, P. J. (2001). The measurement of neuroticism using MMQ, MPI, EPI and EPQ items: a psychometric analysis based on item response theory. *Pers. Individ. Differ.* 30, 641–656. doi: 10.1016/S0191-8869(00)00062-3
- Ferrando, P. J., and Anguiano-Carrasco, C. (2009). The interpretation of the EPQ Lie scale scores under honest and faking instructions: a multiple-group IRT-based analysis. *Pers. Individ. Differ.* 46, 552–556. doi: 10.1016/j.paid.2008.12.013
- Ferrando, P. J., and Chico, E. (2001). Detecting dissimulation in personality test scores: a comparison between person-fit indices and detection scales. *Educ. Psychol. Meas.* 61, 997–1012. doi: 10.1177/00131640121971617
- Finch, H. (2010). Item parameter estimation for the MIRT model: Bias and precision of confirmatory factor analysis-based models. *Appl. Psych. Meas.* 34, 10–26. doi: 10.1177/0146621609336112
- 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
- Forrest, S., Lewis, C. A., and Shevlin, M. (2000). Examining the factor structure and differential functioning of the Eysenck Personality Questionnaire Revised-abbreviated. *Pers. Individ. Differ.* 29, 579–588. doi: 10.1016/S0191-8869(99)00220-2
- Francis, L. J. (1996). The relationship between Eysenck's personality factors and attitude towards substance use among 13–15-year-olds. *Pers. Individ. Differ.* 21, 633–640. doi: 10.1016/0191-8869(96)00125-0
- Francis, L. J., Brown, L. B., and Philipchalk, R. (1992). The development of an abbreviated form of the revised Eysenck Personality Questionnaire (EPQR-A): its use among students in England, Canada, the USA and Australia. *Pers. Individ. Differ.* 13, 443–449. doi: 10.1016/0191-8869(92)90073-X
- Francis, L. J., Lewis, C. A., and Ziebertz, H. G. (2006). The short-form revised Eysenck personality questionnaire (EPQR-S): a German edition. *Soc. Behav. Personal.* 34, 197–204. doi: 10.2224/sbp.2006.34.2.197
- Francis, L. J., and Pearson, P. R. (1988). Religiosity and the short-scale EPQ-R indices of E, N and L, compared with the JEPI, JEPQ and EPQ. *Pers. Individ. Differ.* 9, 653–651. doi: 10.1016/0191-8869(88)90162-6
- Gale, C. R., Booth, T., Möttus, R., Kuh, D., and Deary, I. J. (2013). Neuroticism and Extraversion in youth predict mental wellbeing and life satisfaction 40 years later. *J. Res. Pers.* 47, 687–697. doi: 10.1016/j.jrp.2013.06.005
- Gillings, V., and Joseph, S. (1996). Religiosity and social desirability: Impression management and self-deceptive positivity. *Pers. Individ. Differ.* 21, 1047–1050.
- Gómez-Benito, J., Hidalgo, M. D., and Padilla, J. L. (2009). Efficacy of effect size measures in logistic regression: an application for detecting DIF. *Methodology* 5, 18–25. doi: 10.1027/1614-2241.5.1.18
- Grau, E., and Ortet, G. (1999). Personality traits and alcohol consumption in a sample of non-alcoholic women. *Pers. Individ. Differ.* 27, 1057–1066.
- Gudgeon, E. T., Connor, J. P., Young, R. M., and Saunders, J. B. (2005). The relationship between personality and drinking restraint in an alcohol dependent sample. *Pers. Individ. Differ.* 39, 885–893. doi: 10.1016/j.paid.2005.03.009
- Haberman, S. J., von Davier, M., and Lee, Y.-H. (2008). *Comparison of Multidimensional Item Response Models: Multivariate Normal Ability Distributions Versus Multivariate Polytomous Ability Distributions (RR-08-45)*. Educational Testing Service.
- Heaven, P. C., Ciarrochi, J., Leeson, P., and Barkus, E. (2013). Agreeableness, conscientiousness, and psychoticism: distinctive influences of three personality dimensions in adolescence. *Brit. J. Psychol.* 104, 481–494. doi: 10.1111/bjop.12002
- Heaven, P. C., Newbury, K., and Wilson, V. (2004). The Eysenck psychoticism dimension and delinquent behaviours among non-criminals: changes across the lifespan? *Pers. Individ. Differ.* 36, 1817–1825. doi: 10.1016/j.paid.2003.07.003
- Hosokawa, T., and Ohyama, M. (1993). Reliability and validity of a Japanese version of the short-form Eysenck Personality Questionnaire—Revised. *Psychol. Rep.* 72, 823–832.
- Howarth, E. (1986). What does Eysenck's psychoticism scale really measure? *Brit. J. Psychol.* 77, 223–227.
- Hoyle, R. H., Fejfar, M. C., and Miller, J. D. (2000). Personality and sexual risk taking: a quantitative review. *J. Pers.* 68, 1203–1231. doi: 10.1111/1467-6494.00132
- Hu, L. T., 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.

- Huang, L., Zhou, D., Yao, Y., and Lan, Y. (2015). Relationship of personality with job burnout and psychological stress risk in clinicians. *Chin. J. Industr. Hyg. Occup. Dis.* 33, 84–87.
- Jodoin, M. G., and Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Appl. Meas. Educ.* 14, 329–349. doi: 10.1207/S15324818AME1404_2
- John, O. P., Naumann, L. P., and Soto, C. J. (2008). “Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues,” in *Handbook of Personality: Theory and Research*, eds. O.P. John, R.W. Robins, and L.A. Pervin (New York, NY: Guilford), 114–158.
- Judge, T. A., Bono, J. E., and Locke, E. A. (2000). Personality and job satisfaction: the mediating role of job characteristics. *J. Appl. Psychol.* 85, 237–249. doi: 10.1037/0021-9010.85.2.237
- Jylhä, P., and Isometsä, E. (2006). The relationship of neuroticism and extraversion to symptoms of anxiety and depression in the general population. *Depress. Anxiety* 23, 281–289. doi: 10.1002/da.20167
- Karancı, A. N., Dirik, G., and Yorulmaz, O. (2006). [Reliability and validity studies of Turkish translation of Eysenck personality questionnaire revised-Abbreviated]. *Turk. J. Psichiatri* 18, 254–261.
- Katz, Y. J., and Francis, L. J. (2000). Hebrew revised Eysenck Personality Questionnaire: Short form (EPQR-S) and abbreviated form (EPQR-A). *Soc. Behav. Personal.* 28, 555–560. doi: 10.2224/sbp.2000.28.6.555
- Kroenke, K., Spitzer, R. L., and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16, 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x
- Laidra, K., Pullmann, H., and Allik, J. (2007). Personality and intelligence as predictors of academic achievement: a cross-sectional study from elementary to secondary school. *Pers. Individ. Differ.* 42, 441–451. doi: 10.1016/j.paid.2006.08.001
- Lajunen, T., and Scherler, H. R. (1999). Is the EPQ Lie Scale bidimensional? Validation study of the structure of the EPQ Lie Scale among Finnish and Turkish university students. *Pers. Individ. Differ.* 26, 657–664.
- Levy, P. (1967). The correction for spurious correlation in the evaluation of short-form tests. *J. Clin. Psychol.* 23, 84–86. doi: 10.1002/1097-4679(196701)23:1<84::AID-JCLP2270230123>3.0.CO;2-2
- Lodhi, P. H., and Thakur, S. (1993). Personality of drug addicts: Eysenckian analysis. *Pers. Individ. Differ.* 15, 121–128.
- Lu, L. (1995). The relationship between subjective well-being and psychosocial variables in Taiwan. *J. Soc. Psychol.* 135, 351–357.
- Lynn, R., and Martin, T. (1997). Gender differences in extraversion, neuroticism, and psychoticism in 37 nations. *J. Soc. Psychol.* 137, 369–373.
- Magis, D., Beland, S., and Raiche, G. (2016). *Package 'difR'. R package version 4.7.*
- Maij-de Meij, A. M., Kelderman, H., and Flier, V. D. H. (2008). Fitting a mixture item response theory model to personality questionnaire data: characterizing latent classes and investigating possibilities for improving prediction. *Appl. Psych. Meas.* 32, 611–631. doi: 10.1177/0146621607312613
- Maltby, J., and Talley, M. (1998). The psychometric properties of an abbreviated form of the revised Junior Eysenck Personality Questionnaire (JEPQR-A) among 12–15-year old US young persons. *Pers. Individ. Differ.* 24, 891–893. doi: 10.1016/S0191-8869(97)00234-1
- Marsh, H. W., and Hau, K. T. (2007). Applications of latent-variable models in educational psychology: the need for methodological-substantive synergies. *Contemp. Educ. Psychol.* 32, 151–170. doi: 10.1016/j.cedpsych.2006.10.008
- Marsh, H. W., Hau, K. T., and Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Struct. Equ. Modeling* 11, 320–341. doi: 10.1207/s15328007sem1103_2
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J., Trautwein, U., et al. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychol. Assess.* 22, 471–491. doi: 10.1037/a0019227
- Marsh, H. W., Morin, A. J., Parker, P. D., and Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annu. Rev. Clin. Psycho.* 10, 85–110. doi: 10.1146/annurev-clinpsy-032813-153700
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J., et al. (2009). Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. *Struct. Equ. Modeling* 16, 439–476. doi: 10.1080/10705510903008220
- Marsh, H. W., Nagengast, B., Morin, A. J., Parada, R. H., Craven, R. G., and Hamilton, L. R. (2011). Construct validity of the multidimensional structure of bullying and victimization: an application of exploratory structural equation modeling. *J. Educ. Psychol.* 103, 701–732. doi: 10.1037/a0024122
- McCrae, R. R., and Costa, P. T. (1985). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Pers. Individ. Differ.* 6, 587–597. doi: 10.1016/0191-8869(85)90008-X
- Mor, N. (2010). “Eysenck personality questionnaire,” in *The Corsini Encyclopedia of Psychology*, eds. I. B. Weiner and W. E. Craighead (Hoboken, NJ: John Wiley & Sons, Inc.), 1–2.
- Muthén, L. K., and Muthén, B. O. (2012). *Mplus Version 7 User's Guide*. Los Angeles, CA: Muthén & Muthén.
- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika* 78, 691–692. doi: 10.1093/biomet/78.3.691
- Nyborg, H. E. (1997). *The Scientific Study of Human Nature: tribute to Hans J. Eysenck at Eighty*. New York, NY: Elsevier.
- Paulhus, D. L. (1991). “Measurement and control of response bias,” in *Measures of Personality and Social Psychological Attitudes*, Vol. 1, eds. J. P. Robinson, P. R. Shaver, and L. S. Wrightsman (San Diego, CA: Academic Press), 17–59.
- Pedrabissi, L., and Santinello, M. (1989). *Inventario per Lansia di «Stato» e di «Tratto»: Nuova Versione Italiana Dello Stai Forma Y: Manuale*. Firenze, IT: Organizzazioni Speciali.
- Petrillo, J., Cano, S. J., McLeod, L. D., and Coon, C. D. (2015). Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. *Value Health* 18, 25–34. doi: 10.1016/j.jval.2014.10.005
- Picconi, L., Jackson, C. J., Balsamo, M., Tommasi, M., and Saggino, A. (2018). Factor structure and measurement invariance across groups of the Italian Eysenck Personality Questionnaire-Short form (EPP-S). *Pers. Individ. Differ.* 123, 76–80. doi: 10.1016/j.paid.2017.11.013
- Qian, M., Wu, G., Zhu, R., and Zhang, S. (2000). Development of the revised Eysenck personality questionnaire short scale for Chinese (EPQ-RSC). *Acta Psychol. Sinica* 32, 317–323.
- R Core Team (2016). *A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. Available online at: <http://www.R-project.org/>
- Reckase, M. D. (2009). *Multidimensional Item Response Theory: Statistics for Social and Behavioral Sciences*. New York, NY: Springer. doi: 10.1007/978-0-387-89976-3
- Reise, S. P. (1990). A comparison of item-and person-fit methods of assessing model-data fit in IRT. *Appl. Psych. Meas.* 14, 127–137. doi: 10.1177/014662169001400202
- Reise, S. P., and Henson, J. M. (2000). Computerization and adaptive administration of the NEO PI-R. *Assessment* 7, 347–364. doi: 10.1177/107319110000700404
- Rizopoulos, D. (2012). *Package 'ltm'. R package version 1.1-0.*
- Rocklin, T., and Revelle, W. (1981). The measurement of extroversion: a comparison of the Eysenck Personality Inventory and the Eysenck Personality Questionnaire. *Brit. J. Soc. Psychol.* 20, 279–284.
- Roger, D., and Morris, J. (1991). The internal structure of the EPQ scales. *Pers. Individ. Differ.* 12, 759–764.
- Saggino, A. (2000). The big three or the big five? A replication study. *Pers. Individ. Differ.* 28, 879–886. doi: 10.1016/S0191-8869(99)00146-4
- Saklofske, D. H., Kelly, I. W., and Janzen, B. L. (1995). Neuroticism, depression, and depression proneness. *Pers. Individ. Differ.* 18, 27–31.
- Sanavio, E., Bertolotti, G., Bettinardi, O., Michielin, P., Vidotto, G., and Zotti, A. M. (2013). The cognitive behavioral assessment (CBA) project: presentation and proposal for international collaboration. *Psychol. Community Health* 2, 362–380. doi: 10.5964/pch.v2i3.61
- Scholte, R. H., and De Bruyn, E. E. (2001). The Revised Junior Eysenck Personality Questionnaire (JEPQ-R): Dutch replications of the full-length, short, and abbreviated forms. *Pers. Individ. Differ.* 31, 615–625. doi: 10.1016/S0191-8869(00)00166-5
- Scholte, R. H., and De Bruyn, E. E. (2004). Comparison of the Giant Three and the Big Five in early adolescents. *Pers. Individ. Differ.* 36, 1353–1371. doi: 10.1016/S0191-8869(03)00234-4

- Smillie, L. D., Bhairo, Y., Gray, J., Gunasinghe, C., Elkin, A., McGuffin, P. et al. (2009). Personality and the bipolar spectrum: normative and classification data for the Eysenck Personality Questionnaire-Revised. *Compr. Psychiat.* 50, 48–53. doi: 10.1016/j.comppsy.2008.05.010
- Sotgiu, I., Anselmi, P., and Meneghini, A. M., (2018). Investigating the psychometric properties of the Questionnaire for Eudaimonic Well-Being: A Rasch analysis. *Test. Psychometr. Methodol. Appl. Psychol.* (in press).
- Spence, R., Owens, M., and Goodyer, I. (2012). Item response theory and validity of the NEO-FII in adolescents. *Pers. Individ. Differ.* 53, 801–807. doi: 10.1016/j.paid.2012.06.002
- Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., and Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory STAI*. Palo Alto, CA: Consulting Psychologists Press.
- Spitzer, R. L., Kroenke, K., and Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* 282, 1737–1744.
- Thissen, D., and Steinberg, L. (2009). “Item response theory,” in *The Sage Handbook of Quantitative Methods in Psychology*, eds. R. Millsap and A. Maydeu-Olivares (London: Sage), 148–177.
- Tiwari, T., Singh, A. L., and Singh, I. L. (2009). The short-form revised Eysenck personality questionnaire: a Hindi edition (EPQRS-H). *Ind. Psychiatry J.* 18, 27–31. doi: 10.4103/0972-6748.57854
- Ubbiali, A., Chiorri, C., and Hampton, P. (2013). Italian Big Five Inventory. Psychometric properties of the Italian adaptation of the Big Five Inventory (BFI). *Appl. Psychol. Bull.* 266, 37–48.
- Vidotto, G., Cioffi, R., Saggino, A., and Wilson, G. (2008). The Italian version of the Junior Eysenck Personality Questionnaire: a confirmatory factor analysis. *Psychol. Rep.* 103, 715–726. doi: 10.2466/pr0.103.3.715-726
- Wood, J., and Newton, A. K. (2003). The role of personality and blame attribution in prisoners' experiences of anger. *Pers. Individ. Differ.* 34, 1453–1465. doi: 10.1016/S0191-8869(02)00127-7
- Wright, B. D., and Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Meas. Trans.* 8:370.
- Wright, B. D., and Masters, G. N. (1982). *Rating Scale Analysis: Rasch Measurement*. Chicago: Mesa Press.
- Yu, C. Y. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes (Doctoral dissertation, University of California Los Angeles)*. Available online at: http://www.researchgate.net/profile/Brian_Goldman3/publication/222578792_A_Multicomponent_Conceptualization_of_Authenticity_Theory_and_Research/links/00b49520552576cb3b000000.pdf
- Zanon, C., Hutz, C. S., Yoo, H., and Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica* 29:18. doi: 10.1186/s41155-016-0040-x

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

Copyright © 2018 Colledani, Anselmi and Robusto. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Using Overt and Covert Items in Self-Report Personality Tests: Susceptibility to Faking and Identifiability of Possible Fakers

Giulio Vidotto^{1*}, Pasquale Anselmi², Luca Filipponi³, Marco Tommasi⁴ and Aristide Saggino⁴

¹ Department of General Psychology, School of Psychology, University of Padova, Padova, Italy, ² Department of Philosophy, Sociology, Education and Applied Psychology, School of Psychology, University of Padova, Padova, Italy, ³ Department of Developmental Psychology and Socialization, School of Psychology, University of Padova, Padova, Italy, ⁴ Department of Psychological, Humanistic and Territorial Sciences, Università degli Studi "G. d'Annunzio" Chieti-Pescara, Chieti, Italy

OPEN ACCESS

Edited by:

Dorian A. Lamis,
Emory University School of Medicine,
United States

Reviewed by:

Kathy Ellen Green,
University of Denver, United States
Elisa Pedroli,
Istituto Auxologico Italiano (IRCCS),
Italy

*Correspondence:

Giulio Vidotto
giulio.vidotto@unipd.it

Specialty section:

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

Received: 09 March 2018

Accepted: 11 June 2018

Published: 03 July 2018

Citation:

Vidotto G, Anselmi P, Filipponi L,
Tommasi M and Saggino A (2018)
Using Overt and Covert Items
in Self-Report Personality Tests:
Susceptibility to Faking
and Identifiability of Possible Fakers.
Front. Psychol. 9:1100.
doi: 10.3389/fpsyg.2018.01100

Self-report personality tests widely used in clinical, medical, forensic, and organizational areas of psychological assessment are susceptible to faking. Several approaches have been developed to prevent or detect faking, which are based on the use of faking warnings, ipsative items, social desirability scales, and validity scales. The approach proposed in this work deals with the use of overt items (the construct is clear to test-takers) and covert items (the construct is obscure to test-takers). Covert items are expected to be more resistant to faking than overt items. Two hundred sixty-seven individuals were presented with an alexithymia scale. Two experimental conditions were considered. Respondents in the faking condition were asked to reproduce the profile of an alexithymic individual, whereas those in the sincere condition were not asked to exhibit a particular alexithymia profile. The items of the scale were categorized as overt or covert by expert psychotherapists and analyzed through Rasch models. Respondents in the faking condition were able to exhibit measures of alexithymia in the required direction. This occurred for both overt and covert items, but to a greater extent for overt items. Differently from overt items, covert items defined a latent variable whose meaning was shared between respondents in the sincere and faking condition, and resistant to deliberate distortion. Rasch fit statistics indicated unexpected responses more often for respondents in the faking condition than for those in the sincere condition and, in particular, for the responses to overt items by individuals in the faking condition. More than half of the respondents in the faking condition showed a drift rate (difference between the alexithymia levels estimated on the responses to overt and covert items) significantly larger than that observed in the respondents in the sincere condition.

Keywords: faking, overt, covert, psychological assessment, personality tests, Rasch models

INTRODUCTION

Self-report personality tests, such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher et al., 1989), the Eysenck Personality Questionnaire (EPQ; Eysenck and Eysenck, 1975), the Millon Clinical Multiaxial Inventory-IV (MCMI-IV; Millon et al., 2015), and the Sixteen Personality Factor Questionnaire (16PF; Cattell et al., 1970), are widely used in clinical, medical,

forensic, and organizational areas of psychological assessment (see, e.g., Domino and Domino, 2006; Rothstein and Goffin, 2006; Kaplan and Saccuzzo, 2009). An important limitation of these measures is that people can fake or distort responses. Faking occurs when respondents (a) engage in presentation behavior, framing a presentation of truth in a positive way; (b) lie; or (c) use only expediency as the criterion for making representations, without regard for either truth or falsehood (Levin and Zickar, 2002).

Several approaches have been developed to prevent or detect faking. Faking warning comprises a warning to test-takers that advanced approaches exist for detecting faking on the personality test that is being used. It may also include the information that adverse consequences will result for those who have been found to fake (Fluckinger et al., 2008). Literature supports faking warning as a viable approach to reducing, although not completely eliminating, faking (Goffin and Woods, 1995; Rothstein and Goffin, 2006). A meta-analysis by Dwight and Donovan (2003) indicated that faking warning may reduce faking by 30% on average, with larger reductions accompanying warnings that include mention of the consequences of faking detection. In addition, faking warning is inexpensive to add to an assessment program and can be easily combined with other approaches to faking reduction. However, there are some concerns associated with the use of this strategy for reducing faking. The validity of personality measures can be reduced by test-takers trying too hard to appear as though they are not faking (Dwight and Donovan, 2003). Faking warning has been found to increase the cognitive loading of personality trait scores (Vasilopoulos et al., 2005), that is the extent to which cognitive ability is assessed by the personality test. Cognitive loading may decrease the validity of personality measures because a given personality test score might be, to some extent, indicative of the test-taker's level of cognitive ability as well as of his/her personality (Rothstein and Goffin, 2006).

Social desirability is the tendency of respondents to answer questions in a manner that will be viewed favorably by others, rather than how they truly feel or believe (King and Bruner, 2000). Elevate scores to social desirability scales have been taken as an indication of possible faking (van de Mortel, 2008), and "corrections" have been proposed that remove the effects of social desirability from personality test scores (Goffin and Christiansen, 2003; Sjöberg, 2015). However, there is evidence in the literature that social desirability is a poor indicator of faking (Zickar and Robie, 1999; Peterson et al., 2011), and that correcting personality test scores on the basis of social desirability does not improve the validity of measures (Christiansen et al., 1994; Ones et al., 1996; Ellingson et al., 1999).

The ipsative approach (or forced-choice approach) aims at obtaining more honest, self-descriptive responses to personality items by reducing the effect of perceived desirability of response options. This is achieved by presenting statements in pairs, triplets or quartets that have been equated with respect to perceived desirability (Rothstein and Goffin, 2006). The test-taker is instructed to choose the statement that best describes him/her. Because all the options have the same perceived desirability, there is no clear benefit to distort responses. Performance on one or

more ipsative measures that falls below change to a statistically significant degree indicates biased responding. There is not clear evidence that tests with ipsative items reduce faking (Fluckinger et al., 2008), whereas they could increase the cognitive loading of trait scores, with a detrimental effect on the validity of measures (Christiansen et al., 2005). Moreover, test-taker reactions to these tests may be less positive than reactions to traditional tests (Harland, 2003).

The validity scales aim at measuring the extent to which respondents endorse items in a forthright manner. The validity scales of the Minnesota Multiphasic Personality Inventory (MMPI, Hathaway and McKinley, 1940, 1943), and those of its revisions, are among the most relevant examples. A type of validity scales are the lie scales, which aim at detecting attempts by respondents to present themselves in a favorable light. The logic beyond these scales is that only people who are high on social deception would endorse very improbable and trivial statements such as "I have never stolen anything, not even a hairpin." Professionals have been warned against the use of validity scales for detecting faking. If a person is highly motivated to present an average, yet different profile, he/she is likely to be able to accomplish that simulation without the validity scales detecting faking (Streicher, 1991). Respondents are able to reproduce without detection a specific profile (e.g., a creative artist), provided that they possess an accurate conception of the role to be simulated (Kroger and Turnbull, 1975).

The approach presented in this article takes into account whether the construct measured by the items is clear to test-takers or not. An item is called "overt" when the respondents immediately understand what the item is intended to measure. An item is called "covert" when the respondents (at least those without a thorough knowledge of the construct under investigation) are unaware of what the item measures. Covert items are expected to be more resistant to faking than overt items. Whenever test-takers have no idea about what the items are measuring, they cannot distort the responses in such a manner to present themselves in the desired way. Covert items have less face validity than overt items (Loewenthal, 2001). As a consequence, they demand a non-trivial knowledge of the construct to be correctly distorted in the desired direction.

The influence of faking on overt and covert items has been poorly investigated in the literature. Alliger et al. (1996) compared an overt and a covert integrity test in terms of their susceptibility to faking. The test scores of respondents who were asked to appear as honest as possible (faking condition) were compared with the test scores of respondents who were asked to answer the questions as candidly as possible (sincere condition). In the overt test, the respondents in the faking condition showed greater integrity than those in the sincere condition. No difference between the two conditions was found in the covert test.

The present study aims at investigating the influence of faking on overt and covert items, and the identifiability of possible fakers. The comparison between overt and covert is carried out at the level of the items, instead of being at the level of the different test (i.e., an overt test and a covert test). An overt test and a covert test measuring the same construct might differ with respect to

the way in which the construct is defined. Conversely, the overt and covert items belonging to the same test derive from the same definition of the construct. Therefore, differences between the functioning of overt and covert items can be more easily attributed to the different clarity of the underlying construct, rather than to the different definition of the construct itself. Moreover, using one test instead of two reduces time and costs of the psychological assessment.

An analysis procedure is used, which is based on Rasch models (Rasch, 1960; Andrich, 1988; Bond and Fox, 2001). Rasch models characterize the responses of persons to items as a function of person and item measures, which, respectively, pertain to the level of a quantitative latent trait possessed by the persons or by the items. The specific meaning of these measures relies on the subject of the psychological assessment. In cognitive assessment, for instance, person measures denote the ability of persons, and item measures denote the difficulty of items. In this area, the higher the ability of a person relative to the difficulty of an item, the higher the probability that the person will give a correct response to the item. In health status assessment, person measures denote the health of persons, and item measures denote the severity of items. In this area, the higher the health of a person relative to the severity of an item, the higher the probability that the person will give to the item a response denoting absence of symptoms (e.g., a response “Not at all” to an item asking the person if he/she has trouble falling asleep). Applications of Rasch models for psychological assessment are well documented in the literature (see, e.g., Cole et al., 2004; Shea et al., 2009; Thomas, 2011; Anselmi et al., 2013, 2015; Da Dalt et al., 2013, 2015; Colledani et al., 2018; Sotgiu et al., 2018).

Several advantages derive from a Rasch analysis of faking. Rasch models allows for the transformation of non-linear, ordinal raw scores into linear, interval measures. Differently from ordinal scores, interval measures are characterized by measurement units that maintain the same size over the entire domain, so that measurement is more precise. Misusing ordinal raw scores as they were interval measures (e.g., calculating means and variances) is a common malpractice that can lead to erroneous conclusions (Merbitz et al., 1989; Kahler et al., 2008; Grimby et al., 2012). The measurement units constructed by Rasch models are called log-odds units or logits (Wright, 1993).

In the framework of Rasch models, the measures of respondents quantify the level of latent trait possessed by them. We expect the measures estimated on covert items to be less susceptible to faking than the measures estimated on overt items.

In addition to persons, Rasch models parameterize the items of the test. The location of the items on the latent trait defines the meaning of the variable which the items are intended to implement and, hence, its construct validity (Wright and Stone, 1999; Smith, 2001). Differently from overt items, we expect the covert items to implement a latent variable whose meaning is resistant to deliberate distortion. This means that the latent variables resulting by the responses of sincere respondents and fakers to covert items are expected to be similar, whereas the latent variables resulting by their responses to overt items are expected to be not.

In the framework of Rasch analysis, fit statistics are computed for each person and each item, that express the adherence between observed and expected responses. The fit statistics of a person quantify the extent to which his/her response behavior is consistent with that of the majority of people. These statistics might suggest, for instance, that the person has responded randomly or idiosyncratically, or that he/she has employed a particular response strategy (Smith, 2001; Linacre, 2009). Faking is a kind of response strategy (Frederiksen and Messick, 1959). We expect the fit statistics to reveal unexpected response behaviors more often for fakers than for sincere respondents. This is expected to occur more often for overt items, which should be more susceptible to faking.

MATERIALS AND METHODS

In the present work, a scenario was set up that concerns the faking of an alexithymia scale in personnel selection. Alexithymia is the inability to recognize, express and verbalize emotions. This construct was chosen because it is relatively little-known and, therefore, it is unlikely that individuals know how to distort their responses to covert items in the desired direction. Personnel selection was chosen because it is a high-stake setting in which individuals are highly motivated to fake. The occurrence of faking in personnel selection is well documented in the literature (see, e.g., Hough et al., 1990; Barrick and Mount, 1996; Ones et al., 1996; Hough, 1998; Rosse et al., 1998).

Respondents

Two hundred sixty-seven university students, recruited from various degree courses at the University of Padova, took part in the study on a voluntary basis. Their mean age was 25.58 years ($SD = 4.15$), and 196 (73.41%) were female. All respondents gave written informed consent in accordance with the Declaration of Helsinki and anonymized for the analyses. The project has been approved, now as later, by the Ethical Committee for the Psychological Research of the University of Padova since a prospective ethics approval was not required at the time when the research was conducted (Protocol n. 2616).

Measure of Alexithymia and Procedure

The Roman Alexithymic Scale (RAS; Baiocco et al., 2005) consists of 27 items, which are evaluated on a 4-point scale (Never-1, Sometimes-2, Often-3, and Always-4). Thirteen items are reverse. Greater scores indicate greater alexithymia.

The RAS was administered in individual sessions. All the respondents were asked to consider that they were applying for a job in which they were very interested. The respondents were randomly assigned to one of two conditions. The respondents in the faking condition were asked to reproduce the profile of an alexithymic individual. The instructions given to respondents in this condition were:

“Imagine you have responded to a job posting for a job that is prestigious, well-paid, and very important to you. The ideal candidate must be a person with a solid basic training and good skills in the use of computer programs.

Good organizational skills, task-oriented objectives, emotional detachment, self-control, imperturbability, and no emotional involvement complete the profile. The received CVs will be selected on the basis of the requested requirements. Now, answer the questionnaire that I will present to you in such a way as to satisfy the conditions to be the ideal candidate.”

Conversely, the respondents in the sincere condition were not asked to exhibit a particular alexithymia profile. The instructions given to respondents in this condition were:

“Imagine you have responded to a job posting for a job that is prestigious, well-paid, and very important to you. The ideal candidate must be a person with a solid basic training and good skills in the use of computer programs. Good organizational skills and spontaneity complete the profile. The received CVs will be selected on the basis of the requested requirements. Now, answer the questionnaire that I will present to you in such a way as to satisfy the conditions to be the ideal candidate.”

Categorization of the Items of the Roman Alexithymia Scale as “Overt” or “Covert”

Twenty-four expert psychotherapists were instructed about the meaning of “overt” and “covert” items, and were asked to categorize each of the 27 items of the RAS as overt or covert. The

psychotherapists worked individually. Their evaluations were based on the content of the items and not on the response data.

For each item, **Table 1** presents the number of psychotherapists who categorized it as overt or covert. Twenty-one items were identified as overt (e.g., “I clearly recognize the emotions I feel”) and 6 as covert (e.g., “My physical sensations confuse me”). The agreement among psychotherapists was high for all the items. The lowest percentage of agreement was 87.50, and it was only observed for 2 items out of 27. There was perfect agreement (100%) for 17 items. The average agreement was 97.53%.

Cohen’s k (Cohen, 1968) was computed on all the $\frac{24!}{2!(24-2)!} = 276$ pairs of psychotherapists. The lowest agreement ($k = 0.57$) was observed in one pair only. Perfect agreement ($k = 1$) was observed in 68 pairs. The average agreement was $\bar{k} = 0.87$ ($SD = 0.10$). Kendall’s W (Kendall and Babington Smith, 1939) confirmed the high agreement among psychotherapists ($W = 0.88$, $df = 26$, $p < 0.001$).

Data Analyses

Among the Rasch models, the rating scale model (RSM; Andrich, 1978) was chosen because the response scale of the RAS is polytomous and equal for all the items. The analyses were run

TABLE 1 | Categorization of the items of the Roman Alexithymic Scale as “overt” or “covert”.

Item	Categorized as overt		Categorized as covert		Overall judgment
	N	%	N	%	
1	24	100	0	0	OVERT
2	0	0	24	100	COVERT
3	23	95.83	1	4.17	OVERT
4	22	91.67	2	8.33	OVERT
5	21	87.50	3	12.50	OVERT
6	24	100	0	0	OVERT
7	1	4.17	23	95.83	COVERT
8	24	100	0	0	OVERT
9	24	100	0	0	OVERT
10	23	95.83	1	4.17	OVERT
11	0	0	24	100	COVERT
12	23	95.83	1	4.17	OVERT
13	2	8.33	22	91.67	COVERT
14	24	100	0	0	OVERT
15	24	100	0	0	OVERT
16	21	87.50	3	12.50	OVERT
17	24	100	0	0	OVERT
18	23	95.83	1	4.17	OVERT
19	24	100	0	0	OVERT
20	24	100	0	0	OVERT
21	24	100	0	0	OVERT
22	1	4.17	23	95.83	COVERT
23	24	100	0	0	OVERT
24	0	0	24	100	COVERT
25	24	100	0	0	OVERT
26	24	100	0	0	OVERT
27	24	100	0	0	OVERT

using the computer program Facets 3.66.0 (Linacre, 2009). The responses to the reverse items were rescored prior to the analyses.

The functioning of the items and that of the response scale, as well as the internal consistency of the RAS were evaluated in all the analyses. The functioning of the items was evaluated through the infit and outfit mean-square statistics of the items. Their expected value is 1. Values greater than 2 (Wright and Linacre, 1994; Linacre, 2002b) for a specific item suggest that the item is badly formulated and confusing, or that it may measure a construct that is different from that measured by the other items (Smith, 2001; Linacre, 2009).

Likert scale structure requires that increasing levels of latent trait in a respondent correspond to increasing probabilities that he/she will choose higher response categories (Linacre, 2002a). The functioning of the response scale was assessed by determining whether the step calibrations (the points on the latent trait where two adjacent response categories are equally probable) were ordered or not (Linacre, 2002a; Tennant, 2004). If they were not ordered (i.e., if they did not increase monotonically while going up the response scale), then there would be discordance between the alexithymia level of respondents and the choice of the response categories. This would be interpreted as an indication that the response scale is not adequate for measuring alexithymia.

The internal consistency of the RAS was evaluated through the separation reliability (R) of respondents (Fisher, 1992; Linacre, 2009). R is the Rasch equivalent of Cronbach's α , but it is considered to be a better estimate of internal consistency for two main reasons (Wright and Stone, 1999; Smith, 2001). First, Cronbach's α assumes that the level of measurement error is uniform across the entire range of test scores. Actually, the level of measurement error is generally larger for high and low scores than for scores in the middle of the range. This is due to the fact that, usually, there are more items designed to measure medium levels of the trait than items designed to measure extreme levels. In Rasch models, the estimate of each person measure has an associated standard error of measurement, thus differences in the level of measurement error among individuals are taken into account. Second, Cronbach's α uses test scores for calculating the sample variance. Since test scores are not linear representations of the variable they are intended to indicate, the calculation of variance from them is always incorrect to some degree. Conversely, if the data fit the Rasch model, the measures estimated for each respondent are on a linear scale. Therefore, these measures are numerically suitable for calculating the sample variance.

Unidimensionality of the RAS was evaluated through infit and outfit mean-square statistics of the items, Wright's unidimensionality index (WUI; Wright, 1994), and confirmatory factor analysis (CFA). Infit, outfit, and WUI are Rasch-based indicators of unidimensionality. Values of infit and/or outfit greater than 2 for a particular item suggest that the item may measure a construct that is different from that measured by the other items (Smith, 2001; Linacre, 2009). WUI is the ratio between the separation reliability of respondents based on asymptotic standard errors and the separation reliability of respondents based on misfit-inflated standard errors

(Wright, 1994; Tennant and Pallant, 2006). Values above 0.9 are indicative of unidimensionality. CFA was run using Lisrel 8.71 (Jöreskog and Sörbom, 2005). According to Schermelleh-Engel et al. (2003), fit is reasonable when χ^2 is smaller than $3 \times df$ (where df is the number of degrees of freedom), root mean square error of approximation (RMSEA) is smaller than 0.08, comparative fit index (CFI) is larger than 0.95, normed fit index (NFI) and goodness of fit index (GFI) are larger than 0.90.

Investigating the Influence of Faking on Overt and Covert Items

Three RSM analyses were run to investigate the influence of faking on overt and covert items. The first analysis was performed on the overall sample of respondents ($N = 267$). The responses to the overt items were considered separately from those to the covert items. This provided us with two measures for each respondent (parameters β), one denoting his/her alexithymia level estimated on the responses to overt items and the other denoting his/her alexithymia level estimated on the responses to covert items. It is worth noting that the estimates of parameters β are not influenced by the number of items. The estimates relative to overt and covert items were anchored to the same mean. Greater measures (i.e., larger logits) indicate higher alexithymia levels.

A 2×2 mixed factorial ANOVA was conducted, in which the condition (sincere, faking) was the between factor, and the item type (overt, covert) was the within factor. The dependent variables were the β estimates based on overt and covert items. We expect respondents in the faking condition to show greater alexithymia than respondents in the sincere condition. Since covert items are assumed to be less susceptible to manipulation than overt items, we expect the difference between the two conditions to decrease when the responses to covert items are considered.

The location of the items on the latent variable defines the meaning of the variable itself, thus providing information about construct validity (Wright and Stone, 1999; Smith, 2001). Two separate RSM analyses were conducted on respondents in the sincere and faking condition. These analyses provided us with two measures for each item (parameters δ), one estimated from the responses in the sincere condition and the other one estimated from the responses in the faking condition. Greater measures (i.e., larger logits) indicate items with fewer responses denoting alexithymia. The item measures concerning the two conditions were correlated. Since covert items should be more resistant to manipulation than overt items, we expect to find a stronger positive correlation between the measures of covert items than between the measures of overt items.

Investigating the Identifiability of Possible Fakers

This section presents two methods that could allow for identifying possible fakers. The first method is based on the infit and outfit mean-square statistics of the respondents. The expected value of both statistics is 1. Values greater than 2 (Wright and Linacre, 1994; Linacre, 2002b) for a specific respondent suggest that his/her response behavior is unexpected,

given that exhibited by the majority of respondents. For example, he/she could have responded randomly or idiosyncratically, or he/she could have employed a particular response strategy (Smith, 2001; Linacre, 2009). Since faking is a kind of response strategy (Frederiksen and Messick, 1959), the fit statistics of respondents could allow the identification of possible fakers.

Five-hundred samples were generated, each one including all the 134 respondents in the sincere condition, plus 5 respondents randomly sampled from the faking condition. Therefore, the 500 samples differed from each other with respect to the 5 respondents sampled from the faking condition. Fit statistics allow the identification of respondents whose response behavior differs from that of the majority of respondents. For this reason, in each of the 500 samples, the number of possible fakers was kept small (5) compared to that of respondents who were asked to be sincere (134). The RSM was estimated on each sample, separately for the responses given to overt and covert items. We obtained, for each respondent of each sample, fit statistics based on the responses to overt items and fit statistics based on the responses to covert items. We expect the fit statistics to exceed the critical value of 2 more often for respondents in the faking condition than for respondents in the sincere condition. Overt items are assumed to be more susceptible to faking attempts than covert items. For respondents in the faking condition, we expect the fit statistics pertaining the responses to overt items to exceed 2 more often than those pertaining the responses to covert items. The z test was used for testing the statistical difference in the percentages of fit statistics greater than 2 between respondents of the two conditions, as well as between overt and covert items. Effect size of the z statistics was evaluated through odd ratio (OR). For each fit statistic (FS; infit or outfit) and each item type (overt or covert), an OR was computed as $(P_{\text{faking FS} > 2} \times P_{\text{sincere FS} < 2}) / (P_{\text{faking FS} < 2} \times P_{\text{sincere FS} > 2})$. For the respondents in the faking condition, an OR was computed for each fit statistic as $(P_{\text{overt FS} > 2} \times P_{\text{covert FS} < 2}) / (P_{\text{overt FS} < 2} \times P_{\text{covert FS} > 2})$.

The second method is based on computing a *drift rate* for each respondent, that is defined as the difference between his/her alexithymia level estimated on overt items and that estimated on covert items. For each respondent in the faking condition, it is tested if his/her drift rate is statistically larger than the average of the drift rates pertaining to the respondents in the sincere condition. The one sample t -test was used for this purpose. The rejection of the null hypothesis suggests that the respondent does not belong to the same population of the respondents in the sincere condition.

RESULTS

The Rasch-based statistics infit, outfit (smaller than 2 for all the items) and WUI (0.95), as well as the CFA-based statistics χ^2 ($739.39 < 3 \times 281$) and RMSEA (0.07) suggested that the RAS is substantially unidimensional. Conversely, the CFA-based statistics CFI (0.93), NFI (0.89), and GFI (0.72)

suggested that there could be more than one dimension. These results do not allow for drawing certain conclusions about the unidimensionality of the RAS.

The step calibrations were ordered (the step calibrations Never-Sometimes, Sometimes-Often, Often-Always were -1.34 , 0.52 , 0.83 in the analysis on the overall sample; -1.93 , 0.40 , 1.53 in the analysis on respondents in the sincere condition; -0.99 , 0.18 , 0.81 in the analysis on respondents in the faking condition). This suggests that the response scale is adequate for measuring alexithymia.

The RAS has an adequate internal consistency (see Table 2). No relevant differences were found when the overall sample was considered, the respondents in the sincere condition only, or those in the faking condition only. The statistics R and α are affected by the number of items. The Spearman–Brown prophecy formula (Brown, 1910; Spearman, 1910) was used to predict the internal consistency of the covert items if their number was equal to that of the overt items (i.e., 21 items). Under this condition, the internal consistency of covert items largely resembled that of overt items.

Influence of Faking on Overt and Covert Items

Figure 1 depicts the average alexithymia level of respondents in the sincere and faking condition, estimated on overt and covert items separately. Respondents in the faking condition showed greater alexithymia than those in the sincere condition, both on the overt items [$\bar{\beta}_{\text{faking-overt}} = 0.49$, $\bar{\beta}_{\text{sincere-overt}} = -0.52$, $SE_{\text{faking-overt}} = 0.06$, $t(265) = 11.90$, $p < 0.001$, Cohen's $d = 1.46$] and on the covert items [$\bar{\beta}_{\text{faking-covert}} = 0.01$, $\bar{\beta}_{\text{sincere-covert}} = -0.58$, $SE_{\text{faking-covert}} = 0.11$, $t(265) = 3.79$, $p < 0.001$, Cohen's $d = 0.47$]. The interaction between condition and item type was significant, with the difference in alexithymia between respondents in the two conditions decreasing when responses to covert items were considered [$F(1,265) = 7.65$, $p < 0.01$, $\eta_p^2 = 0.03$]. Respondents in the faking group showed higher alexithymia on overt items than on covert items [$t(132) = 3.83$, $p < 0.001$, Cohen's $d = 0.40$]. Respondents in the sincere group showed the same alexithymia on overt and covert items [$t(133) = 0.46$, $p = 0.65$, Cohen's $d = 0.04$].

When the item measures estimated for the sincere condition and for the faking condition were correlated, a significant correlation was found between the measures of covert items ($r = 0.92$, $p = 0.01$) but not between those of overt items ($r = 0.40$, $p = 0.07$). The former correlation was significantly stronger than the latter (Fisher's $z = 1.87$, $p < 0.05$). This result suggests that, differently from overt items, covert items define a latent variable whose meaning is shared between respondents in the sincere and faking condition, and resistant to deliberate distortion.

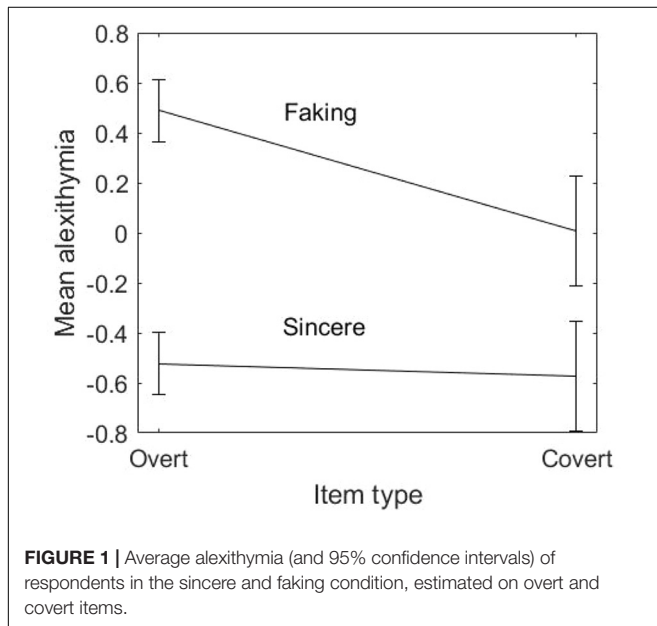
Identifying Possible Fakers

About 5% of respondents in the sincere condition gave unexpected responses (infit/outfit > 2) to overt items (5.97% infit; 4.48% outfit) or covert items (5.22% infit; 4.48% outfit). In our study, this 5% can be taken as a benchmark for the percentage

TABLE 2 | Internal consistency of the Roman Alexithymic Scale.

Items	Overall sample (N = 267)		Sincere condition (N = 134)		Faking condition (N = 133)	
	R	α	R	α	R	α
Entire scale (N = 27)	0.87	0.87	0.82	0.81	0.85	0.85
Overt items (N = 21)	0.86	0.87	0.82	0.81	0.84	0.84
Covert items (N = 6)	0.60	0.63	0.56	0.63	0.63	0.62
Covert items (N = 21) – predicted ^a	0.84	0.86	0.82	0.86	0.86	0.85

R, Rasch separation reliability; α , Cronbach's alpha. ^aPredicted with the Spearman–Brown prophecy formula (Brown, 1910; Spearman, 1910).



of respondents with unexpected response behavior that can be encountered among respondents who are expected to be sincere.

Across the 500 samples, about 35% of respondents in the faking condition gave unexpected responses to overt items (35.56% infit; 35.72% outfit), and about 19% to covert items (18.56% infit; 19.68% outfit). These percentages are greater than those observed in respondents in the sincere condition ($z = 7.04$, $OR = 2.27$ for infit on overt items; $z = 7.43$, $OR = 11.85$ for outfit on overt items; $z = 3.93$, $OR = 4.14$ for infit on covert items; $z = 4.38$, $OR = 5.22$ for outfit on covert items; $p < 0.001$ for all). For the respondents in the faking condition, unexpected responses were more frequent to overt items than to covert items ($z = 13.53$, $OR = 2.42$ for infit; $z = 12.67$, $OR = 2.27$ for outfit). These results suggest that both overt and covert items are susceptible to faking attempts, with overt items being to a greater extent.

The average drift rate of respondents in the sincere condition was 0.05 ($SD = 1.32$). Seventy-two respondents in the faking condition (out of 133; 54.14%) showed a drift rate significantly larger than 0.05 (Type-1 error probability = 0.05; Cohen's d from 0.15 to 3.04), suggesting that they could belong to a population different from that of the respondents in the sincere condition.

DISCUSSION

The present study investigated the influence of faking on overt and covert items, and the identifiability of possible fakers. The investigations have been conducted on an alexithymia scale. The results were in line with expectations. Experimentally induced fakers were able to exhibit measures of alexithymia in the required direction. This occurred for both overt and covert items, but to a greater extent for overt items. Differently from overt items, covert items defined a latent variable whose meaning was shared between respondents in the sincere and faking condition, and resistant to deliberate distortion. Rasch fit statistics indicated unexpected responses more often for respondents in the faking condition than for those in the sincere condition and, in particular, for the responses to overt items by individuals in the faking condition. More than half of the respondents in the faking condition showed a drift rate (difference between the alexithymia levels estimated on the responses to overt and covert items) significantly larger than that observed in the respondents in the sincere condition.

We found that also covert items were susceptible to faking, although to a lesser extent than overt items. This is not in line with Alliger et al. (1996), who found no difference between the scores of the respondents who were asked to fake and those of the respondents who were asked to be sincere in a covert integrity test. Alliger et al. (1996) used an integrity test specifically developed as covert test. Differently, the items of the RAS were *a posteriori* categorized as overt and covert, instead of being specifically developed as overt or covert. The covert items of the RAS may be not as “covert” as items that are appositely thought to be covert. The Rasch fit statistics indicated more unexpected responses to covert items by respondents asked to fake than by respondents asked to be sincere. This confirms the small, yet existing influence of faking on covert items, that has been found in the present study.

Two methods for identifying possible fakers have been proposed, which are based on the fit statistics of the respondents and on the computation of a drift rate. Results of the present study provide moderate evidence for the effectiveness of the two methods. It is worth noting that, once the Rasch model has been calibrated on unbiased data, it can be used for testing possible fakers without having to collect data on a new sample. Moreover, drift rate and fit statistics can be used for identifying possible fakers without having to add further tests (e.g., validity scales, social desirability scales) to the assessment program.

Limitations and Suggestions for Future Research

Rasch models assume unidimensionality of the scale. A limitation of the present study is that unidimensionality of the RAS has not been supported with certainty. Multidimensionality, if present, could have influenced the estimation of person measures (Henning, 1988), with a detrimental effect on the functioning of the proposed approach. Future studies could investigate the functioning of the approach with scales whose unidimensionality is well-established.

Another limitation of the present study is that respondents in the faking condition were not asked about their perceived success in simulating the required profile. Future studies could investigate the relationship between the perceived success in simulating a profile and the responses to overt and covert items.

The items considered in the present study were *a posteriori* categorized as overt and covert, instead of being specifically developed as overt or covert. This could represent another limitation of the study, even if it is worth noting that psychotherapists agreed to a very large extent in categorizing the items. Future studies could investigate the functioning of items that are specifically developed as overt or covert.

A relatively little-known construct (alexithymia) was chosen to reduce the probability that individuals know how to distort

their responses to covert items in the desired direction. Future studies should investigate the resistance of covert items to faking when the construct under evaluation is well-known.

A high-stake setting has been considered (personnel selection) in which individuals are highly motivated to fake. Future studies should investigate the functioning of overt and covert items in other areas of psychological assessment, such as clinical, medical, and forensic areas, which are affected by faking.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

GV contributed conception and design of the study. LF conducted the research. PA and LF performed the statistical analyses. PA wrote the first draft of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

REFERENCES

- Alliger, G. M., Lilienfeld, S. O., and Mitchell, K. E. (1996). The susceptibility of overt and covert integrity tests to coaching and faking. *Psychol. Sci.* 7, 32–39. doi: 10.1111/j.1467-9280.1996.tb00663.x
- Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika* 43, 561–573. doi: 10.1007/BF02293814
- Andrich, D. (1988). *Rasch Models for Measurement*. Sage: Beverly Hills. doi: 10.4135/9781412985598
- Anselmi, P., Vianello, M., and Robusto, E. (2013). Preferring thin people does not imply derogating fat people. A Rasch analysis of the implicit weight attitude. *Obesity* 21, 261–265. doi: 10.1002/oby.20085
- Anselmi, P., Vidotto, G., Bettinardi, O., and Bertolotti, G. (2015). Measurement of change in health status with Rasch models. *Health Qual. Life Out.* 13:16. doi: 10.1186/s12955-014-0197-x
- Baiocco, R., Giannini, A. M., and Laghi, F. (2005). *SAR. Scala Alessitimica Romana. Valutazione delle Capacità di Riconoscere, Esprimere e Verbalizzare le Emozioni. Manuale e Protocolli [RAS. Roman Alexithymic Scale. Evaluation of the Ability to Recognize, Express and Verbalize Emotions. Manual and Protocols]*. Trento: Erickson.
- Barrick, M. R., and Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *J. Appl. Psychol.* 81, 261–272. doi: 10.1037/0021-9010.81.3.261
- Bond, T. G., and Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *Br. J. Psychol.* 3, 296–322. doi: 10.1111/j.2044-8295.1910.tb00207.x
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., and Kaemmer, B. (1989). *The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for Administration and Scoring*. Minneapolis, MN: University of Minnesota Press.
- Cattell, R. B., Eber, H. W., and Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire*. Champaign, IL: Institute for Personality and Ability Testing.
- Christiansen, N. D., Burns, G. N., and Montgomery, G. E. (2005). Reconsidering forced-choice item format for applicant personality assessment. *Hum. Perform.* 18, 267–307. doi: 10.1207/s15327043hup1803_4
- Christiansen, N. D., Goffin, R. D., Johnston, N. G., and Rothstein, M. G. (1994). Correcting the 16PF for faking: effects on criterion-related validity and individual hiring decisions. *Pers. Psychol.* 47, 847–860. doi: 10.1111/j.1744-6570.1994.tb01581.x
- Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol. Bull.* 70, 213–220. doi: 10.1037/h0026256
- Cole, J. C., Rabin, A. S., Smith, T. L., and Kaufman, A. S. (2004). Development and validation of a Rasch-derived CES-D short form. *Psychol. Assess.* 16, 360–372. doi: 10.1037/1040-3590.16.4.360
- Colledani, D., Robusto, E., and Anselmi, P. (2018). Development of a new abbreviated form of the junior Eysenck personality questionnaire-revised. *Pers. Individ. Differ.* 120, 159–165. doi: 10.1016/j.paid.2017.08.037
- Da Dalt, L., Anselmi, P., Bressan, S., Carraro, S., Baraldi, E., Robusto, E., et al. (2013). A short questionnaire to assess pediatric resident's competencies: the validation process. *Ital. J. Pediatr.* 39:41. doi: 10.1186/1824-7288-39-41
- Da Dalt, L., Anselmi, P., Furlan, S., Carraro, S., Baraldi, E., Robusto, E., et al. (2015). Validating a set of tools designed to assess the perceived quality of training of pediatric residency programs. *Ital. J. Pediatr.* 41:2. doi: 10.1186/s13052-014-0106-2
- Domino, G., and Domino, M. L. (2006). *Psychological Testing: An Introduction*, 2nd Edn. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511813757
- Dwight, S. A., and Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Hum. Perform.* 16, 1–23. doi: 10.1207/S15327043HUP1601_1
- Ellingson, J. E., Sackett, P. R., and Hough, L. M. (1999). Social desirability corrections in personality measurement: issues of applicant comparison and construct validity. *J. Appl. Psychol.* 84, 155–166. doi: 10.1037/0021-9010.84.2.155
- Eysenck, H. J., and Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (Adult and Junior)*. Kent: Hodder & Stoughton.
- Fisher, W. P. Jr. (1992). Reliability statistics. *Rasch Meas. Trans.* 6:238.
- Fluckinger, C. D., McDaniel, M. A., and Whetzel, D. L. (2008). "Review of faking in personnel selection," in *Search of the Right Personnel*, ed. M. Mandal (New Delhi: Macmillan), 91–109.
- Frederiksen, N., and Messick, S. (1959). Response set as a measure of personality. *Educ. Psychol. Meas.* 19, 137–157. doi: 10.1177/001316445901900201

- Goffin, R. D., and Christiansen, N. D. (2003). Correcting personality tests for faking: a review of popular personality tests and an initial survey of researchers. *Int. J. Select. Assess.* 11, 340–344. doi: 10.1111/j.0965-075X.2003.00256.x
- Goffin, R. D., and Woods, D. M. (1995). Using personality testing for personnel selection: Faking and test-taking inductions. *Int. J. Select. Assess.* 3, 227–236. doi: 10.1111/j.1468-2389.1995.tb00035.x
- Grimby, G., Tennant, A., and Tesio, L. (2012). The use of raw scores from ordinal scales: Time to end malpractice? *J. Rehabil. Med.* 44, 97–98. doi: 10.2340/16501977-0938
- Harland, L. K. (2003). Using personality tests in leadership development: Test format effects and the mitigating impact of explanations and feedback. *Hum. Resour. Dev. Q.* 14, 285–301. doi: 10.1002/hrdq.1067
- Hathaway, S. R., and McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *J. Psychol.* 10, 249–254. doi: 10.1080/00223980.1940.9917000
- Hathaway, S. R., and McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory*. Minneapolis, MN: University of Minnesota Press.
- Henning, G. (1988). The influence of test and sample dimensionality on latent trait person ability and item difficulty calibrations. *Lang. Test.* 5, 83–99. doi: 10.1177/026553228800500106
- Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Hum. Perform.* 11, 209–244. doi: 10.1080/08959285.1998.9668032
- Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., and McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *J. Appl. Psychol.* 75, 581–595. doi: 10.1037/0021-9010.75.5.581
- Jöreskog, K. G., and Sörbom, D. (2005). *Lisrel 8: User's Reference Guide*. Chicago, IL: Scientific Software International Inc.
- Kahler, E., Rogausch, A., Brunner, E., and Himmel, W. (2008). A parametric analysis of ordinal quality-of-life data can lead to erroneous results. *J. Clin. Epidemiol.* 61, 475–480. doi: 10.1016/j.jclinepi.2007.05.019
- Kaplan, R. M., and Saccuzzo, D. P. (2009). *Psychological Testing: Principles, Applications, and Issues*, 7th Edn. Belmont, CA: Wadsworth.
- Kendall, M. G., and Babington Smith, B. (1939). The problem of m rankings. *Ann. Math. Stat.* 10, 275–287. doi: 10.1214/aoms/1177732186
- King, M. F., and Bruner, G. C. (2000). Social desirability bias: a neglected aspect of validity testing. *Psychol. Market.* 17, 79–103. doi: 10.1002/(SICI)1520-6793(200002)17:2<79::AID-MAR2>3.0.CO;2-0
- Kroger, R. O., and Turnbull, W. (1975). Invalidity of validity scales: the case of the MMPI. *J. Consult. Clin. Psychol.* 43, 48–55. doi: 10.1037/h0076266
- Levin, R. A., and Zickar, M. J. (2002). “Investigating self-presentation, lies, and bullshit: understanding faking and its effects on selection decisions using theory, field research, and simulation,” in *The Psychology of Work: Theoretically Based Empirical Research*, eds J. M. Brett and F. Drasgow (Mahwah, NJ: Lawrence Erlbaum Associates, Inc), 253–276.
- Linacre, J. M. (2002a). Optimizing rating scale category effectiveness. *J. Appl. Meas.* 39, 85–106.
- Linacre, J. M. (2002b). What do Infit and Outfit, mean-square and standardized mean? *Rasch Meas. Trans.* 16:878.
- Linacre, J. M. (2009). *Facets Rasch Measurement Computer Program (Version 3.66.0) [Computer Software]*. Chicago, IL: Winsteps.com.
- Loewenthal, K. M. (2001). *An Introduction to Psychological Tests and Scales*, 2nd Edn. Philadelphia, PA: Psychology Press.
- Merbitz, C., Morris, J., and Grip, J. C. (1989). Ordinal scales and foundations of misinference. *Arch. Phys. Med. Rehabil.* 70, 308–312.
- Millon, T., Millon, C., and Grossman, S. (2015). *Millon Clinical Multiaxial Inventory - IV Manual*, 4th Edn. Bloomington, IN: NCS Pearson, Inc.
- Ones, D. S., Viswesvaran, C., and Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: the red herring. *J. Appl. Psychol.* 81, 660–679. doi: 10.1037/0021-9010.81.6.660
- Peterson, M. H., Griffith, R. L., Isaacson, J. A., O'Connell, M. S., and Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Hum. Perform.* 24, 270–290. doi: 10.1080/08959285.2011.580808
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Test*. Copenhagen: Danish Institute for Educational Research. Chicago, IL: The University of Chicago Press.
- Rosse, J. G., Stecher, M. D., Miller, J. L., and Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *J. Appl. Psychol.* 83, 634–644. doi: 10.1037/0021-9010.83.4.634
- Rothstein, M. G., and Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Hum. Resour. Manage. Rev.* 16, 155–180. doi: 10.1016/j.hrmr.2006.03.004
- Schermelell-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. Online* 8, 23–74.
- Shea, T. L., Tennant, A., and Pallant, J. F. (2009). Rasch model analysis of the depression, anxiety and stress scales (DASS). *BMC Psychiatry* 9:21. doi: 10.1186/1471-244X-9-21
- Sjöberg, L. (2015). Correction for faking in self-report personality tests. *Scand. J. Psychol.* 56, 582–591. doi: 10.1111/sjop.12231
- Smith, E. V. Jr. (2001). Evidence for the reliability of measures and validity of measure interpretation: a Rasch measurement perspective. *J. Appl. Psychol.* 2, 281–311.
- Sotgiu, I., Anselmi, P., and Meneghini, A. M. (2018). Investigating the psychometric properties of the questionnaire for eudaimonic well-being: a Rasch analysis. *Test. Psychom. Methodol. Appl. Psychol.* 14:12.
- Spearmen, C. C. (1910). Correlation calculated from faulty data. *Br. J. Psychol.* 3, 271–295. doi: 10.1111/j.2044-8295.1910.tb00206.x
- Streicher, P. J. (1991). Personality simulation as a response set on the MMPI. *Appl. H.R.M. Res.* 2, 68–83.
- Tennant, A. (2004). Disordered thresholds: an example from the functional independence measure. *Rasch Meas. Trans.* 17, 945–948.
- Tennant, A., and Pallant, J. F. (2006). Unidimensionality matters! (A tale of two Smiths?). *Rasch Meas. Trans.* 20, 1048–1051.
- Thomas, M. L. (2011). The value of item response theory in clinical assessment: a review. *Assessment* 18, 291–307. doi: 10.1177/1073191110374797
- van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Aust. J. Adv. Nurs.* 25, 40–48.
- Vasilopoulos, N. L., Cucina, J. M., and McElreath, J. M. (2005). Do warnings of response verification moderate the relationship between personality and cognitive ability? *J. Appl. Psychol.* 90, 306–322. doi: 10.1037/0021-9010.90.2.306
- Wright, B., and Stone, M. (1999). *Measurement Essentials*, 2nd Edn. Wilmington, DE: Wide Range, Inc.
- Wright, B. D. (1993). Logits? *Rasch Meas. Trans.* 7:288.
- Wright, B. D. (1994). Unidimensionality coefficient. *Rasch Meas. Trans.* 8:385.
- Wright, B. D., and Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Meas. Trans.* 8, 370–371.
- Zickar, M. J., and Robie, C. (1999). Modeling faking good on personality items: an item-level analysis. *J. Appl. Psychol.* 84, 551–563. doi: 10.1037/0021-9010.84.4.551

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

Copyright © 2018 Vidotto, Anselmi, Filippini, Tommasi and Saggino. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



New Perspectives in the Adaptive Assessment of Depression: The ATS-PD Version of the QuEDS

Andrea Spoto^{1*}, Francesca Serra¹, Ivan Donadello², Umberto Granzio¹ and Giulio Vidotto¹

¹ Quantitative Psychology Laboratory, Department of General Psychology, University of Padua, Padua, Italy, ² Fondazione Bruno Kessler, Trento, Italy

OPEN ACCESS

Edited by:

Michela Balsamo,
Università degli Studi G. d'Annunzio
Chieti e Pescara, Italy

Reviewed by:

Davide Marengo,
Università degli Studi di Torino, Italy
Leonardo Adrián Medrano,
Siglo 21 Business University,
Argentina

*Correspondence:

Andrea Spoto
andrea.spoto@unipd.it

Specialty section:

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

Received: 13 March 2018

Accepted: 11 June 2018

Published: 06 July 2018

Citation:

Spoto A, Serra F, Donadello I,
Granzio U and Vidotto G (2018) New
Perspectives in the Adaptive
Assessment of Depression: The
ATS-PD Version of the QuEDS.
Front. Psychol. 9:1101.
doi: 10.3389/fpsyg.2018.01101

Measurement is a crucial issue in psychological assessment. In this paper a contribution to this task is provided by means of the implementation of an adaptive algorithm for the assessment of depression. More specifically, the Adaptive Testing System for Psychological Disorders (ATS-PD) version of the Qualitative-Quantitative Evaluation of Depressive Symptomatology questionnaire (QuEDS) is introduced. Such implementation refers to the theoretical background of Formal Psychological Assessment (FPA) with respect to both its deterministic and probabilistic issues. Three models (one for each sub-scale of the QuEDS) are fitted on a sample of 383 individuals. The obtained estimates are then used to calibrate the adaptive procedure whose performance is tested in terms of both efficiency and accuracy by means of a simulation study. Results indicate that the ATS-PD version of the QuEDS allows for both obtaining an accurate description of the patient in terms of symptomatology, and reducing the number of items asked by 40%. Further developments of the adaptive procedure are then discussed.

Keywords: adaptive psychological assessment, formal psychological assessment, depression, qualitative and quantitative assessment, item response theory (IRT)

1. INTRODUCTION

Measurement in Psychology is a challenging issue that rose since the very beginning of the history of Psychology as a science. The first formalization of measurement in Psychology is due to the empirical research of Weber and to the Psychophysics of Fechner (1860), while Spearman (1904) paved the way for the measurement of theoretical constructs through methodologies such as factor analysis. The lack of a consistent formulation of the measurement problem in psychology was first addressed by Stevens by means of direct methods for psychological measurement (Stevens, 1946, 1951, 1957). An axiomatic definition of the measurement scales appeared within the theoretical framework of the Relational Theory of Measurement (RTM; Suppes and Zinnes, 1963; Suppes et al., 1989; Narens and Luce, 1993).

Currently, in psychological measurement the classical test theory (CTT; Spearman, 1904; Novick, 1965; Gulliksen, 2013) and the item response theory (IRT; Rasch, 1960; Lord, 1980) are the formal and methodological frameworks for the construction of measurement tools. The classical test theory relies on the evaluation of the reliability, validity and factorial structure of a defined psychological measure. An important limit lies in the impossibility to distinguish and compare the parameters related to the individuals (abilities) and those relative to the items (difficulties). On the other hand, the item response theory and the Rasch model (Rasch, 1960) explain the test

performance of individuals by referring to the presence of latent traits. A relation between the latent traits and the observed scores is postulated, so that information on the first ones are inferred starting from the observable performance of the individual in answering the set of items. The relationship between test score and latent trait is expressed by a mathematical model, defined *a priori*. IRT models are a family of mathematical models which describe a wide number of contexts starting from the simple logistic model. The interest in applying these models is growing thanks to the numerous advantages compared to the classical test methods. IRT models represent a fundamental formal tool for applying adaptive measurement in psychology since they allow the definition of precedence relations among items according to their location on the latent trait dimension (Marsman et al., 2018).

In recent years another approach to measurement has been adopted in psychological assessment. This approach allows for expanding the measurement properties obtained through IRT by considering complex precedence relations among items (i.e., beyond the linear order). It refers to an axiomatic formulation of the relations among sets of items and sets of attributes investigated by them. It refers to the possibility of depicting the relations among items according to well established mathematical tools such as *lattices* and *posets* (Birkhoff, 1937, 1940; Davey and Priestley, 2002). It is the Formal Psychological Assessment (FPA; Spoto, 2011; Spoto et al., 2013a; Bottesi et al., 2015; Serra et al., 2015, 2017). It was developed with the aim of providing detailed information about the clinical features endorsed by a patient who answered a specific set of items of a questionnaire.

In the present article, this last methodology is employed to implement an adaptive algorithm for the assessment of depression. More specifically, the main aim of this article is to present the adaptive form of the Qualitative-Quantitative Evaluation of Depressive Symptomatology questionnaire (QuEDS; Serra et al., 2017).

The paper is structured as follows: In section 2, the general concepts concerning recent developments in psychological assessment are addressed; In section 3 a brief outline of the main deterministic and probabilistic issues related to the FPA is presented; section 4 introduces the main concepts of the adaptive algorithm implemented in this study; In sections 5 and 6 a simulation study aimed at testing the accuracy and the efficiency of the adaptive procedure is presented; Final remarks, limitations and future perspectives are explored in section 7.

2. PSYCHOLOGICAL ASSESSMENT: STATE OF THE ART

Measurement is a crucial issue in many fields of psychology. One of them is psychological assessment. The main tools adopted for carrying out measurement in psychological assessment are self-report questionnaires, observation and interviews. Clinical interviews and observation have the capacity for gathering and deepening several information such as nonverbal aspects, which are essential to make a diagnosis (Annen et al., 2012; Fiquier

et al., 2013; Girard et al., 2014). This last is the main aim of clinical assessment, and therefore, the main goal of measurement in clinical psychology. For example, negative emotions and social behaviors are indicators of the severity of depression and relevant predictors of its clinical remission (Philippot et al., 2003; Uhlmann et al., 2012) that are beyond the control and awareness of the patient (Andersen, 1999; Geerts and Brüne, 2009). Nonetheless, both observation and clinical interview are time consuming and prone to inferential errors by clinicians (Strull et al., 1984; Nordgaard et al., 2013).

The self-report questionnaires provide scores that are supposed to indicate the severity of the symptomatology and the impairment level (Groth-Marnat, 2009). The score of a questionnaire is helpful in distinguishing individuals with critical clinical features, but it is not sufficient, in the form so far provided by both CTT and IRT, to differentiate patients with different symptom configurations who obtained similar scores (in the limit, the same score) to the test (Spoto et al., 2013a; Bottesi et al., 2015; Serra et al., 2015, 2017). Moreover, not all the items have the same “weight” from the clinical point of view, since they reflect different symptoms that may be more or less severe (Gibbons et al., 1985; Serra et al., 2017).

Therefore, measurement in clinical psychology should attempt to evaluate individual data in a broad perspective and it should account for individual specific features (Meyer et al., 2001). For example, the construct of depression, as represented by the score, can sometimes be misleading. Indeed, depression can manifest with a variety of different symptoms that may be due to a different culture or a different etiology (Benazzi et al., 2002; Goodwin and Jamison, 2007). Only through personalized assessment it is possible to clearly distinguish among such different manifestations of the disorder (Groth-Marnat, 2009; Serra et al., 2017). For these reasons, an increasing number of self-report assessment tools are validated according to the IRT framework (Gibbons et al., 2008; Embretson and Reise, 2013). In this way the assessment can be much more focused on the objective measure of the uniqueness of a particular clinical configuration. For instance, Balsamo et al. (2014) applied Rasch analysis to the item selection for the Teate Depression Inventory, a self-report depression tool; it has been highlighted by a number of papers that this tool, built according to the IRT methodology, performs better than tools developed within the CTT with respect to many different measurement properties such as, for instance, convergent-divergent validity (e.g., Innamorati et al., 2013; Balsamo et al., 2015a,b).

As mentioned above, IRT is also a crucial stepping stone for implementing adaptive testing, which in turn is an important way to implement the administration of a questionnaire in a personalized fashion. Each individual is administered with different scale items on the basis of the specific answers he/she provided to the previous ones (Wainer, 2000; Fliege et al., 2005). Within this field, Computerized Adaptive Testing (CAT; Wainer, 2000) is an approach which presents multiple advantages. Different studies showed that questionnaires could be shortened without loss of information by means of CAT, achieving a more efficient and equally accurate assessment (Petersen et al., 2006). CAT’s procedure mimics the semi-structured interview

(i.e., clinical interview where only some items, out a list, are posed according to specific adaptive selection criteria), letting the algorithm to carry out inferences by accounting for all the information collected and following a logically correct process (Spoto, 2011). It should be quite easy to understand why the assessment of knowledge is one of the core areas in which such procedures have been developed. For instance, Eggen and Straetmans (2000) combined IRT with statistical procedures, like sequential probability ratio test and weighted maximum likelihood, for classifying people under exam. Other systems use Bayesian statistical techniques instead of IRT in the evaluation of students' knowledge. Examples are EDUFORM (Nokelainen et al., 2001), and PARES (Marinagi et al., 2007). In the field of knowledge assessment the ALEKS (Assessment and Learning in Knowledge Spaces) system implements the theoretical framework of Knowledge Space Theory (KST; Doignon and Falmagne, 1985, 1999; Falmagne and Doignon, 2011) for the adaptive assessment of the so called *knowledge state* of a student, i.e., the set of items that he/she is able to solve about a specific topic (Grayce, 2013).

The formulation of an adaptive algorithm is clearly more difficult in the clinical setting. In fact, the objectivity of the questions and therefore of the answers given by the subject is much more questionable, and the probability of misinterpretations in the answers is increased. Despite this, research has demonstrated that both IRT and CAT (Baek, 1997) can be applied to the measurement of attitudes and personality variables (Reise and Waller, 1990). In the clinical framework, Spiegel and Nenh (2004) developed an expert system, which calculates possible symptom combinations given the answers of a patient and returns all possible diagnoses coherent with such combination. Yong et al. (2007) developed an interactive self-help system for depression diagnosis that provides advice about patients levels of impairment. Simms et al. (2011) developed the CAT for Personality Disorders (CAT-PD) aimed at realizing a computerized adaptive assessment system. CAT has been applied also in developing adaptive classification tests by means of stochastic curtailment using CES-D for depression (Finkelman et al., 2012; Smits et al., 2016). Gibbons et al. (2008) used the combination of item response theory and CAT in mood and anxiety disorder assessment. In particular, they applied a bifactor structure consisting of a primary dimension and four sub-factors (mood, panic-agoraphobia, obsessive-compulsive, and social phobia) to build the CAT version of the Mood and Anxiety Spectrum Scales (MASS; Dell'Osso et al., 2002). Results of this study showed that the adaptive tool allowed to both administer a small set of the items (the most relevant for a given individual) with no loss of information compared to the classical form of the MASS, and strongly reduce time consumption as well as patient and clinician burden. In six patients with mood disorders (three major depressive disorder and three bipolar disorder) who were interviewed by the psychiatrist, many of the CAT items investigating important information, such as a history of manic symptoms, potentially risky behaviors, etc., were both endorsed and not documented in the psychiatric evaluation through SCID-I (First et al., 1996). Gibbons' study is an important example of how adaptive testing can be effective and efficient.

Although there have been several attempts to apply adaptive clinical assessment, as far as we know, no system is able to combine adaptivity, quantitative and qualitative information, and punctual estimates of error parameters. Within clinical psychology, the Formal Psychological Assessment (Spoto et al., 2010, 2013a; Serra et al., 2015; Granzio et al., 2018) represents an important contribution in the improvement of adaptive psychological assessment, allowing to overcome the obstacles encountered up to now in this field. The main deterministic and probabilistic concepts of this methodology are presented in the next section.

3. THE FORMAL PSYCHOLOGICAL ASSESSMENT

An adaptive testing, most of the time, relies on a formal substratum composed by several relations among items (Donadello et al., 2017). The FPA is a methodology which allows defining assessment tools able to detect specific symptoms in several mental disorders, independently of the kind of assessment used, such as self-report (Serra et al., 2015, 2017) or behavioral observations (Granzio et al., 2018). FPA makes it possible applying two theories of Mathematical Psychology in psychological assessment: The Knowledge Space Theory (KST; Doignon and Falmagne, 1999; Falmagne and Doignon, 2011) and the Formal Concept Analysis (FCA; Wille, 1982; Ganter and Wille, 1999). The core characteristic of FPA is the definition of a relation between a set of items and a set of clinical criteria. In the next two subsections are reported separately the deterministic and probabilistic main concepts of FPA.

3.1. FPA Deterministic Concepts

In FPA methodology, a very basic concept is the *clinical domain*, intended as a nonempty set Q of questions that can be asked to a patient for investigating a certain psychopathology. Each item is referred as an *object*. The complete list of the items included in the QuEDS, namely the objects in the present article, are listed in **Table 1**.

For instance, the item QuEDS34 "I feel sad" is an object. For the sake of simplicity, the term item will be preferred to object in the sequel. The subset $K \subseteq Q$ of all the items that are endorsed by a patient is called the *clinical state* of that patient. Each item investigates one or more *attributes*, intended as a diagnostic criteria of a psychopathology selected from either clinical sources like the DSM-5 (American Psychiatric Association, 2013) or the scientific literature, or both. The complete list of the attributes investigated by the QuEDS is displayed in **Table 2**.

For instance, the diagnostic criterion "Depressed mood" of the DSM-5 is investigated by the aforementioned item. The collections of both items and attributes make it possible to build the so called *clinical context*, formally a triple (Q, M, I) , where Q is the set of items, M is a set of attributes and I is a binary relation between the sets Q and M which assigns to each item $q \in Q$ the attributes $m \in M$ it investigates. The clinical context can be represented as a Boolean Matrix, having the items in the rows and the attributes in the columns: whenever an item q investigates an

TABLE 1 | The list of items of the QuEDS grouped by sub-scale.

Item	Description
Cognitive scale	
QuEDS5	I am stressed by feelings of guilt
QuEDS6	I think the world is cruel and unhappy
QuEDS9	I think my life is hell and I only deserve to feel bad
QuEDS10	I feel incapable of facing life's events
QuEDS14	I feel incapable and totally useless
QuEDS19	I have thought about killing myself
QuEDS20	Sometimes I think it would be better if I were dead
QuEDS21	I am really worried about my health
QuEDS24	I am afraid about everything that will happen to me because I am incapable
QuEDS25	I feel like I don't have any more power over my empty and sad life
QuEDS27	Making choices is hard for me
QuEDS30	I feel I am a burden other people and it would be better if I killed myself
QuEDS32	I am disappointed in myself and the choices I have made
QuEDS33	I have problems making decisions
QuEDS41	I often feel like a loser
Somatic scale	
QuEDS1	I feel like I don't have the same energy to have sex
QuEDS2	I often wake up in the middle of the night and I can't go back to sleep
QuEDS3	I feel like that my thinking is slowing down
QuEDS4	I have sleep problems
QuEDS11	I suffer from somatic disorders (e.g., headache, stomach ache)
QuEDS13	I am less interested in sex
QuEDS16	My desire to eat is not the same
QuEDS22	My weight has had significant changes
QuEDS23	I've visibly lost (or gained) weight
QuEDS26	My appetite has changed
QuEDS28	I feel I'm slowing down in my daily routines
QuEDS31	I don't have much energy and I feel tired
QuEDS35	My ability to think and memorize has been reduced
QuEDS39	I feel so tired and without any energy that I need help to look after myself
Affective scale	
QuEDS7	I keep crying very easily
QuEDS8	I get irritated very easily
QuEDS12	I have lost interest in the future which doesn't have anything good for me
QuEDS15	I see the same unhappiness I have now in the future
QuEDS17	I often feel like crying, but I cannot do it
QuEDS18	I can't have interest and pleasure in people and things as before
QuEDS29	I feel helpless and inhibited facing my incapacity to concentrate
QuEDS34	I feel sad
QuEDS36	I don't have any interest and desire to do anything
QuEDS37	I am agitated about the idea that this sadness won't ever leave me
QuEDS38	I feel agitated
QuEDS40	I am better in the evening more than in the morning

TABLE 2 | The list of the attributes investigated by the items of the QuEDS.

Attribute	Description
A1	Depressed mood
A2	Diminished interest and pleasure
A3	Decreased interest in sex
A4	Increase or loss of weight
A5	Gain or loss of appetite
A6	Insomnia or hypersomnia
A7	Agitation
A8	Psycho-motor retardation
A9	Fatigue or energy loss
A10	Feelings of worthlessness (or Beck's negative view of self)
A11	Feelings of guilt
A12	Diminished ability to think and concentrate
A13	Indecision
A14	Recurrent thoughts of death
A15	Suicidal ideation or attempted suicide
A16	Beck's negative view of the world
A17	Beck's negative expectation of the future
A18	Seligman's learned helplessness
A19	Irritability
A20	Apathy
A21	Health concern
A22	Somatic disorders
A23	More positive mood in the evening

attribute m (i.e., whenever the relation qIm holds true), the qm cell contains the value 1, otherwise 0.

Starting from the clinical context, the *clinical structure* \mathcal{K} can be delineated (Spoto et al., 2010, 2016). A clinical structure is a collection of clinical states, containing at least the empty set (\emptyset) and the whole clinical domain (Q) and it represents the implications among the items of the domain. Whenever a clinical structure is closed under set union (i.e., $K_1 \cup K_2 \in \mathcal{K}$ for all $K_1, K_2 \in \mathcal{K}$) it is a *clinical space*. On the other hand, a clinical structure closed under intersection (i.e., $K_1 \cap K_2 \in \mathcal{K}$ for all $K_1, K_2 \in \mathcal{K}$), it is a *clinical closure space*. In order to obtain a structure where all the states K are in a one to one connection with the set of attributes endorsed by all the items in K , it is necessary to modify the clinical context by defining a relation R between items and attributes, which is dual to I :

$$qRm \iff q-Im.$$

According to this relation, a clinical closure space is obtained, where each state is in a one to one correspondence with the set of attributes endorsed by the items in each state (for details refer to Spoto et al., 2010). In other words the relation R allows for representing in the structure a principle often used in clinical and medical practices: if a patient endorses an item, he should present all the attributes investigated by that item. From a practical point of view, a clinical structure is useful since it includes only the clinical states, that is all and only the admissible response patterns given the clinical context. Any state $K \in \mathcal{K}$

is coherent with the theoretical framework and, therefore, it does not violate specific order relation among items. In FPA, the relation required is based on the attributes investigated by each item and it is called *prerequisite relation*, stating that whenever an item q investigates a subset of attributes of another item r , q is a prerequisite for r . For instance, taken the subset of QuEDS's selected attributes composed by {A1, A17} and two items of the QuEDS, namely QuEDS34 which investigates only A1, and QuEDS15, investigating both A1 and A17. The two rows of the clinical context representing QuEDS15 and QuEDS34 will be in an inclusion relation with respect to the attributes investigated (Table 3).

Among the following response patterns:

- (a) $\{\emptyset\}$;
- (b) $\{QuEDS34\}$;
- (c) $\{QuEDS15\}$;
- (d) $\{QuEDS34, QuEDS15\}$,

only the patterns a, b, and d are clinical states. In fact, it is not possible (excluding errors) that a patient who endorses item QuEDS15 does not endorse item QuEDS34 (i.e., the pattern c). A clinical structure can be represented as a complete lattice displaying the partial order among the items of a domain, where each node contains a subset of items (investigating a specific subset of attributes; Granzio et al., 2018). In the case at hand, the clinical structure is a clinical closure space where each node contains the items endorsed by the patient and the uniquely determined set of attributes (symptoms) corresponding to that set of questions.

By delineating a clinical structure, it is possible to obtain the deterministic skeleton for defining a computerized adaptive assessment, which needs to be completed also from a probabilistic point of view. The next section will deepen the probabilistic features needed to implement the algorithm.

3.2. FPA Probabilistic Concepts

A deterministic clinical structure provides a fundamental starting point for the procedure aimed at creating an adaptive assessment. Nonetheless, such a structure is incomplete from both a theoretical and practical point of view. In fact, each state could be present with different frequencies in the population; moreover, the observed response pattern of a subject could not represent his/her real state. The probability of observing each clinical state π_K is then related to both its actual frequency in the population, and to two further parameters, namely the *false negative* (β) and the *false positive* (η). The false negative refers to the probability that the patient does not endorse an item that he/she actually presents. The false positive parameter, on the contrary, refers

to the probability that a patient endorses an item that he/she does not present. By means of all the aforementioned parameters (i.e., π_K for each $K \in \mathcal{K}$; β_q and η_q for each item $q \in Q$) a *probabilistic clinical structure* (Donadello et al., 2017) can be obtained. Formally, it is a triple (Q, \mathcal{K}, π) where (Q, \mathcal{K}) is the clinical structure and π is the probability distribution on \mathcal{K} estimated through a sample of patients (Spoto et al., 2010). The probability distribution for each response pattern $R \subseteq Q$ is obtained by means of a response function assigning to R its conditional probability given a state K (for all states $K \in \mathcal{K}$), as displayed by the unrestricted latent class model represented by Equation 1:

$$P(R) = \sum_{K \in \mathcal{K}} P(R|K)\pi(K). \tag{1}$$

This model is the so called *basic local independence model* (BLIM; Falmagne and Doignon, 1988; Doignon and Falmagne, 1999). Within the probabilistic clinical structure, the responses to the items are assumed to be locally independent. The conditional probability $P(R|K)$ is determined by the probability of false negative (β_q) and false positive (η_q) while answering to q , as displayed by Equation (2):

$$P(R|K) = \left(\prod_{q \in K \setminus R} \beta_q \right) \left(\prod_{q \in K \cap R} (1 - \beta_q) \right) \left(\prod_{q \in R \setminus K} \eta_q \right) \left(\prod_{q \in R \cup \bar{K}} (1 - \eta_q) \right). \tag{2}$$

In the present study, the expectation-maximization algorithm (Dempster et al., 1977) has been used in order to estimate both the β and η parameters and the probability distribution for \mathcal{K} . These estimates have been carried out on a sample of 383 individuals, according to the same procedure employed in, e.g., Spoto et al. (2013a), Bottesi et al. (2015), and Donadello et al. (2017). Such estimates were then used to implement the adaptive algorithm whose general functioning is detailed in the next section.

4. THE ADAPTIVE TESTING SYSTEM FOR PSYCHOLOGICAL DISORDERS ALGORITHM

In this section we aim at introducing the Adaptive Testing System for Psychological Disorders Donadello et al. (ATS-PD; 2017) developed starting from the clinical structure and the parameters' estimate via the BLIM.

Within this framework, the clinical structure is the deterministic skeleton defining the starting point for an adaptive assessment which, if no error is assumed, could reasonably proceed as follows:

- i) Select the item that is closest to be present in 50% of the clinical states of the structure;
- ii) Ask this item to the patient;
- iii) Register the dichotomous answer;

TABLE 3 | The precedence relation between items QuEDS15 and QuEDS34 as depicted in the clinical context.

Items	A1	...	A17	...	A23
QuEDS15	1	...	1	...	0
QuEDS34	1	...	0	...	0

- iv) Exclude all the states not containing the investigated item if the answer is “yes,” or vice versa, all the states that contain the item if the answer is “no.”

These steps are repeated on the remaining states until only one state remains. The output is the clinical state with all the attributes (diagnostic criteria) satisfied by all the items of the state. This procedure applied in a real context would almost surely fail due to the absence of a probabilistic model defining both the probabilities of the different states, and the error probabilities for the answers. As it has been shown in the previous section, the probabilistic model which accounts for both these issues could be the BLIM. Therefore, an adaptive procedure should make an appropriate use of the probabilities of the states (π_K) and of the false positive (η) and false negative (β) rates for each item.

Thus, the above outlined procedure can be therefore modified as follows:

- i) Detect and administer the item q which best splits into two equal parts the probability mass of the states (*questioning rule*), and register the dichotomous answer;
- ii) Update the probability π_K of all the states $K \in \mathcal{K}$ according to the following *updating rule*:
 - if an affirmative answer to q is observed: increase π_K for all $K \in \mathcal{K}$ which contain q , and decrease π_K for the remaining states;
 - if a negative answer to q is observed: decrease π_K for all $K \in \mathcal{K}$ which contain q , and increase π_K for the remaining states.
- iii) Repeat the previous steps until a given condition is satisfied (*stopping rule*).

The algorithm used in this research implements these three main steps. The questioning rule selects the item to ask, i.e., the item $q \in Q$ that is “maximally informative.” This characteristic is satisfied by the item(s) for which the sum of π_K for all the states containing q best approaches 0.50. In other words, this item maximizes the obtainable information irrespectively of its observed answer (i.e., either “yes” or “no”). If many items are equally informative, one of them is chosen at random. We call $L_n(K)$ the probability of the state K at the step n . At each step of the procedure, the subject’s response is collected by the system. Then, the updating rule is applied to obtain the likelihood $L_{n+1}(K)$ for all the states $K \in \mathcal{K}$. More precisely, let us denote an affirmative response with $r = 1$ and a negative one with $r = 0$. It is then possible to formalize the updating rule of the probability $L(K)$ for each $K \in \mathcal{K}$ as follows:

$$L_{n+1}(K) = \frac{\zeta^K L_n(K)}{\sum_{K' \in \mathcal{K}} \zeta^K L_n(K')} \tag{3}$$

where

$$\zeta_{q,r}^K = \begin{cases} \zeta_{q,1} & \text{if } q \in K, r = 1; \\ 1 & \text{if } q \notin K, r = 1; \\ 1 & \text{if } q \in K, r = 0; \\ \zeta_{q,0} & \text{if } q \notin K, r = 0. \end{cases} \tag{4}$$

In this formulation ζ is a parameter always greater than 1 that increases the likelihood and influences the efficiency of the adaptive assessment process. The higher ζ , the more reliable are considered the answers provided by the subject, and therefore, the more efficient (but potentially less accurate) the adaptive procedure. It has been observed by Falmagne and Doignon (2011) that ζ values less than 2 make the assessment redundant. On the contrary, fixing the ζ value to an excessively high number could affect algorithm accuracy. It has been proven that an adequate value of ζ could be 21 (Falmagne and Doignon, 2011). This value allows for an accurate and efficient detection of the state of individuals in several applications, e.g., ALEKS (Falmagne et al., 2013). An alternative way to estimate ζ is based on the η_q and β_q parameters of each item (see Falmagne and Doignon, 2011, p. 265). The estimate is carried out according to the following formulas:

$$\zeta_{q,1} = \frac{1 - \beta_q}{\eta_q} \qquad \zeta_{q,0} = \frac{1 - \eta_q}{\beta_q}$$

This rule is local since it takes into account both the η and β of the last item asked in order to update the probability of the states. According to this method the “weight” of each item in updating the probabilities is a function of its error rates. Namely, an item whose error rates are low (i.e., whose answer is more reliable) will produce a significant modification on the probability distribution of the states, while a less reliable item will have a weaker effect in updating of the probabilities of the states.

In order to further refine the updating of the states probabilities given the pattern observed at a specific step n of the adaptive assessment, a Bayesian rule can be introduced according to what described by Donadello et al. (2017):

$$P(K_i|R) = \frac{P(R|K_i)L_n(K_i)}{\sum_{j=1}^{|K|} P(R|K_j)L_n(K_j)} \tag{5}$$

Where $P(R|K_i)$ is obtained by Equation (2), and $L_n(K)$ is the estimated probability of the state K at the step n .

All these steps are replicated until a given stopping criterion is reached. In the present article, the stopping rule is satisfied whenever $L_n(K_q)$ is outside the interval [0.20, 0.80] for all $q \in Q$. In this way the algorithm stops as soon as any possible item to be asked splits the probability mass in very unequal parts, indicating that it is almost surely either inside or outside the state. This choice is coherent with previous literature (Falmagne and Doignon, 2011, p. 362). When the stopping criterion is matched, the algorithm concludes the assessment and provides as output the response pattern R , the estimated state K (with its estimated probability) and the amount of time needed to complete the assessment.

In the next section, we will present a simulation study aimed at testing the algorithm under different conditions in order to identify the best performing configuration of the procedure. Before conducting such a simulation we estimated the parameters of the BLIM in order to provide the deterministic skeleton with probabilistic weights.

5. A SIMULATION STUDY

Testing the adaptive procedure with respect to both its accuracy and efficiency is a necessary operation in order to guarantee that (i) the information collected through the adaptive form of the questionnaire is reliable, and (ii) that the administration time of the questionnaire is actually reduced. In order to reach these goals, the traditional form of the QuEDS containing 41 dichotomous items (“Yes”/“No”) grouped into three sub-scales (namely: Cognitive, Somatic and Affective) was administered to a sample of 383 individuals. Using the collected data, the parameters of the BLIM were estimated. Then, such values were passed to the adaptive procedure. Finally, the adaptive algorithm, under different conditions, simulated the administration of the test starting from the available 383 response patterns, and the results were analyzed. The details of all these passages are provided below.

5.1. Sample

The sample of 383 Italian individuals included a clinical group consisting of 38 subjects with Major Depressive Episode (who were diagnosed with either major depressive disorder or bipolar disorder). These patients were recruited by the Neurosciences Mental Health and Sensory Organ (NESMOS) Department of La Sapienza University, Rome. The psychiatrists of NESMOS Department evaluated the presence of Major Depressive Episode in participants of the clinical group by means of the clinical interview and the SCID-I. The diagnosis was then formulated according to the DSM-IV-TR nosology classification system. The exclusion criteria were mental retardation and psychotic traits in order to guarantee a correct interpretation of the meaning of QuEDS items. The 47% of participants were males and the remaining 53% were females. The majority of the participants had a high school diploma, and their age ranged between 21 and 69 years old ($M = 33.5$; $SD = 4.8$). The remaining 345 individuals were randomly selected from the general population and recruited in Padova (68% were females). The majority of participants had a high school diploma, and their age ranged between 19 and 58 years old ($M = 27.5$; $SD = 6.4$). Participants of the non-clinical group did not undergo a psychological assessment. Before the beginning of the administration of the test they were asked to indicate whether they were currently under pharmacological or psychotherapeutic treatment for MDE. The exclusion criterion in the non-clinical group was the presence of MDE (i.e., individuals under pharmacological or psychotherapeutic treatment for depression).

The study was conducted in accordance with the Declaration of Helsinki and the research protocol was approved by the Psychology Ethical Committee of the University of Padua. All participants entered the study of their own free will and provided their written informed consent before taking part. They were informed in detail about the aims of the study, the voluntary nature of their participation, and their right to withdraw from the study at any time and without being penalized in any way. Furthermore, participants were allowed for asking the restitution about their own score, providing authors with their own auto generated code, used during the administration phase.

5.2. Procedure

All participants completed informed consent and sociodemographic forms before answering the questionnaire items. All participants completed the written form of QuEDS according to the following instructions: “Please answer “Yes” or “No” to the following statements on the basis of how you felt in the last 15 days.” No time limit was imposed. Clinical participants provided written, informed consent for potential research analysis and anonymous reporting of clinical findings in aggregate form, at clinical intake.

5.3. Parameters Estimate

As mentioned in the previous sections, the estimate of BLIM’s parameters (i.e., π_K for each state, β_q and η_q for each item), as well as of the fit of the model, were performed with a specific version of the Expectation-Maximization Algorithm (Dempster et al., 1977). For the details of the algorithm, refer to Spoto (2011). The tested models were obtained from the formal contexts displayed in **Tables 4–6** according to the methodology described in Spoto et al. (2010).

The fit of each of the three models has been tested by Pearson’s Chi-square. It is well established that for large data matrices (as those used in the present study) the asymptotic distribution of χ^2 is not reliable. Therefore, a p -value for the obtained χ^2 was calculated by parametric bootstrap with 5,000 replications. An important fit index is provided by the estimates of the error rates. In general, they are expected to be low, but it is crucial that for each item the following inequality holds: $\eta_q < 1 - \beta_q$. If this condition is not satisfied, then the assessment loses its meaning, since the probability of observing a false positive (η) on an item q would be greater than the probability of observing an actual affirmative answer to q . Spoto et al. (2012) established a specific connection between the characteristics of the context and the identifiability of the error parameters of the items. In this respect the value of the unidentifiable parameters (Spoto et al., 2013b; Stefanutti et al., 2018) was fixed to a constant corresponding to the maximum possible value of the parameter (Stefanutti et al., 2018) in order to both preserve the computability of the adaptive procedure and adopt the maximally conservative approach from a diagnostic point of view, preferring accuracy to efficiency.

This first step of the study provided the needed parameters for calibrating the adaptive procedure.

5.4. Simulation Design

Six different conditions for testing the adaptive algorithm were generated by manipulating the following two variables:

1. Estimate of ζ parameter (2 levels): (i) 21 in one case; (ii) estimated via the values of η and β in the second;
2. Implementation of Bayesian updating rule (3 levels): (i) on-line (at each step n of the adaptive procedure), (ii) off-line (when the stopping criterion is reached), (iii) absent (no Bayesian updating is applied).

The adaptive procedure was then run according to each of the six conditions described above in order to simulate the 383 response patterns collected in the previous part of the study. The task of

TABLE 4 | The clinical context for the Cognitive sub-scale of the QuEDS.

Items	A1	A10	A11	A12	A13	A14	A15	A16	A17	A18	A21
QuEDS5	0	0	1	0	0	0	0	0	0	0	0
QuEDS6	0	0	0	0	0	0	0	1	0	0	0
QuEDS9	1	1	1	0	0	0	0	1	0	0	0
QuEDS10	0	0	0	0	0	0	0	0	0	1	0
QuEDS14	0	1	1	0	0	0	0	0	0	1	0
QuEDS19	0	0	0	0	0	1	1	0	0	0	0
QuEDS20	0	0	0	0	0	1	0	0	0	0	0
QuEDS21	0	0	0	0	0	0	0	0	0	0	1
QuEDS24	0	0	0	0	0	0	0	0	1	1	0
QuEDS25	1	0	0	0	0	0	0	1	0	1	0
QuEDS27	0	0	0	1	1	0	0	0	0	0	0
QuEDS30	1	1	1	0	0	1	0	1	0	0	0
QuEDS32	0	1	1	0	0	0	0	0	0	0	0
QuEDS33	0	0	0	0	1	0	0	0	0	0	0
QuEDS41	0	1	0	0	0	0	0	0	0	0	0

TABLE 5 | The clinical context for the Somatic sub-scale of the QuEDS.

Items	A3	A4	A5	A6	A7	A8	A9	A12	A18	A20	A22
QuEDS1	1	0	0	0	0	0	1	0	0	0	0
QuEDS2	0	0	0	1	1	0	0	0	0	0	0
QuEDS3	0	0	0	0	0	1	0	1	0	0	0
QuEDS4	0	0	0	1	0	0	0	0	0	0	0
QuEDS11	0	0	0	0	0	0	0	0	0	0	1
QuEDS13	1	0	0	0	0	0	0	0	0	1	0
QuEDS16	0	0	1	0	0	0	0	0	0	0	0
QuEDS22	0	1	0	0	0	0	0	0	0	0	0
QuEDS23	0	1	0	0	0	0	0	0	0	0	0
QuEDS26	0	0	1	0	0	0	0	0	0	0	0
QuEDS28	0	0	0	0	0	1	1	0	0	0	0
QuEDS31	0	0	0	0	0	0	1	0	0	0	0
QuEDS35	0	0	0	0	0	1	1	1	0	0	0
QuEDS39	0	0	0	0	0	1	1	0	1	1	0

TABLE 6 | The clinical context for the Affective sub-scale of the QuEDS.

Items	A1	A2	A7	A12	A17	A19	A20	A23
QuEDS7	1	0	0	0	0	1	0	0
QuEDS8	0	0	0	0	0	1	0	0
QuEDS12	0	1	0	0	1	0	0	0
QuEDS15	1	0	0	0	1	0	0	0
QuEDS17	1	0	0	0	0	0	1	0
QuEDS18	0	1	0	0	0	0	1	0
QuEDS29	0	0	0	1	0	0	1	0
QuEDS34	1	0	0	0	0	0	0	0
QuEDS36	0	1	0	0	0	0	1	0
QuEDS37	1	0	1	0	0	0	0	0
QuEDS38	0	0	1	0	0	0	0	0
QuEDS40	0	0	0	0	0	0	0	1

the algorithm was to administer the QuEDS in an adaptive form and provide the clinical state as output.

In order to check for the accuracy and the efficiency of the procedure, a number of indexes were used. First, the average number of items asked to converge (i.e., to match the stopping criterion) was used to test the efficiency of the ATS-PD version of QuEDS in terms of reduction of the number of items administered to a patient. Second, the distance between the reconstructed state and the paper and pencil pattern observed for each specific patient was used to evaluate procedure's accuracy. In fact, the higher the distance between the reconstructed state and the response pattern, the greater is the amount of information that is inconsistent between the two modalities of administration of the test. This, in turn, may be due to the error parameters of the items, to a misspecification of the model, or to problems with the algorithm. Since the first two options are excluded given the good fit and the acceptable error estimates (described in the previous section), a strong divergence between the observed pattern and the reconstructed state could be due to some errors in the algorithm; therefore, the measured distance is expected to be as low as possible if the algorithm is accurate. In this respect some further concepts need to be introduced.

A response pattern R_i is the list of the observed answers provided by a subject i to the written version of QuEDS. K_i is the state in \mathcal{K} that is the output of the adaptive assessment when the input is the response pattern R_i . It is important to emphasize that the adaptive procedure always produces a state $K \in \mathcal{K}$ as output even if the observed response pattern $R_i \notin \mathcal{K}$. Thus, we define the distance $d(K_i, R_i)$ as the cardinality of $K_i \Delta R_i$. As a consequence, the results of the simulation for each subject can fall into one of the following mutually exclusive categories:

1. $K_i = R_i$: this happens when the response pattern $R_i \in \mathcal{K}$. In this case $d(K_i, R_i) = 0$. In the specific condition, the output of the adaptive assessment is exactly the same of the output obtained with the written version of the sub-scale;
2. $K_i \neq R_i$: This happens whenever $R_i \notin \mathcal{K}$. In this case of course, $d(K_i, R_i) > 0$. In this situation two alternatives may occur:
 - i) $d(K_i, R_i)$ is minimum, that is, there is no $K^* \in \mathcal{K}$ such that $d(K^*, R_i) < d(K_i, R_i)$;
 - ii) $d(K_i, R_i)$ is not minimum, that is, there exists $K^* \in \mathcal{K}$ such that $d(K^*, R_i) < d(K_i, R_i)$.

Of course, the occurrence of this last situation should be as rare as possible if the adaptive algorithm is accurate.

6. RESULTS

Results are presented separately for the model fitting analysis, and for the algorithm efficacy and effectiveness test.

6.1. Model Fitting

It is important to stress that the size of the three structures was relatively small counting 124 states for the Cognitive scale, 163 for the Somatic scale, and 142 for the Affective scale. The results of the model fitting for the three structures demonstrated an adequate fit of the models to the set of collected data [Cognitive:

$\chi^2_{(32,144)} = 23,348$, bootstrap- $p = 0.07$; Somatic: $\chi^2_{(15,972)} = 7,237$, bootstrap- $p = 0.16$; Affective: $\chi^2_{(3,928)} = 8,696$, bootstrap- $p = 0.06$]. Therefore, a general adequate fit of the structures was observed.

Another fundamental information obtained by the model fitting was the estimate of both the η and β parameters for each item of the scale. With respect to the η parameters, the estimated values are in general adequately small for almost all items, ranging between 0.01 and 0.18. Only few items presented relatively high values of the estimated β . For the Cognitive scale such items are QuEDS9 with $\beta = 0.50$, and QuEDS30 with $\beta = 0.44$. This criticism may be explained by the phrasing of the items which both include two coordinate sentences. In the Affective scale of QuEDS two items reported high β estimates: namely item QuEDS7 with $\beta = 0.44$, and item QuEDS17 with $\beta = 0.45$. Interestingly both these items are related to crying, suggesting that either the subjects could intentionally fake the specific answer, or that subjects' answer could be affected by a poor introspection about "crying." Although not fully satisfying, it will be shown in the next section that these values did not affect the performance of the adaptive procedure.

Given the number of participants and the number of states, the estimate of the π_K for some states $K \in \mathcal{K}$ was 0. Therefore, we did not use such information in the following part of the simulation and we fixed the starting values of all π_K in the adaptive procedure according to the uniform distribution. This specific implementation is quite common in several applications.

In general, both the fit indexes and the parameters' estimates were satisfactory and were, then, implemented in the adaptive procedure whose performance is analyzed in the next subsection.

6.2. Clinical State Reconstruction

The results of both accuracy and effectiveness tests supported the goodness of the adaptive algorithm. First of all (and actually, as expected) the system, in all the tested versions, was able to correctly reproduce the patient's pattern whenever the pattern R_i was a state in \mathcal{K} . Moreover, for the great majority of the patterns, whenever $R_i \notin \mathcal{K}$, the algorithm mapped the pattern into the closest state in the structure, thus, $d(K_i, R_i) = \min$ in most cases. Only in a limited number of cases happened that the algorithm mapped a response pattern $R_i \notin \mathcal{K}$ to a state that was not at the minimum distance, thus, for some cases happened that there was a state $K_i^* \in \mathcal{K}$ such that $d(K_i^*, R_i) < d(K_i, R_i)$. This could depend on the sequence of questions asked by the system, which in turn is affected by the error parameters, and by the type of update used by the system. However, this situation has rarely occurred in the simulations.

Table 7 summarizes the results with respect to both accuracy and efficiency of the algorithm.

The table displays that the best performing configuration of the ATS-PD version of the QuEDS is the one with on-line Bayesian update and the parameter ζ computed as a function of η_q and β_q . In the cognitive scale, which has 15 items in total, with this configuration we had a maximum of 11 questions asked and a minimum of 7 to reach the stopping criterion; the average is 8.83 items asked ($SD = 0.47$). It means that the saving in terms

TABLE 7 | Main results of the adaptive algorithm testing.

ζ	Bayesian update	Cognitive		Somatic		Affective	
		Max items	Not Min	Max items	Not Min	Max items	Not Min
21	off-line	17	12	9	8	8	40
21	on-line	11	15	9	6	8	35
21	absent	17	11	9	8	8	39
(η_q, β_q)	off-line	20	7	11	8	23	39
(η_q, β_q)	on-line	11	10	9	5	8	29
(η_q, β_q)	absent	20	7	11	8	29	34

The first column contains the levels of the variable ζ ; The second column refers to the implementation of the Bayesian update. The remaining columns contain, for each sub-scale, the maximum number of questions asked to reach the stopping criterion, and the number of response patterns mapped to a state K_i such that there exists $d(K^*, R_i) < d(K_i, R_i)$ for the three sub-scales. In bold the best performing configuration.

of question posed is between 31 and 53%. We found 10 response patterns R in which the distance $d(K_i, R_i)$ was not minimal. In the specific case, $d(K_i, R_i) - d(K_i^*, R_i) \leq 2$. It means that the distance between the output state K_i and the state K_i^* that was the closest one to R_i was never greater than 2 (that is, no more than two answers in K_i were different from K_i^* , whose distance was minimum from R_i).

The somatic scale has 14 items in total. In the best performing configuration we observed a maximum of 9 items asked to reach the stopping criterion and a minimum of 8 item asked to achieve the output of the assessment; the average is 8.42 items asked ($SD = 0.82$). The saving in terms of questions posed is between 36 and 50%. Out of the 173 observed different response patterns, five were mapped to a state whose distance from the pattern was not minimal. Also in this case $d(K_i, R_i) - d(K_i^*, R_i) \leq 2$.

The affective scale counted a total 12 items in the written version. In the best performing configuration we had a maximum of 8 items asked and a minimum of 7; the average is 7.66 items asked ($SD = 0.47$). The saving in terms of questions posed is between 33 and 42%. In this scale 29 response patterns were mapped to a state at a non minimal distance. In the specific case $d(K_i, R_i) - d(K_i^*, R_i) \leq 3$. This last scale seemed to perform in a less accurate, although still adequate and effective, way.

It is important to stress how the procedure is carried out on-line: this means that the questioning rule, the updating rule (together with the Bayesian correction) and the stopping rule are applied in real time even on a standard machine. This indicates an adequate optimization of the computational costs of the procedure.

7. DISCUSSION

This paper aimed at presenting the adaptive version of the three sub-scales (namely, cognitive, somatic, affective) of the QuEDS questionnaire. The computerized algorithm was implemented for the new questionnaire based on an extension of an already existing algorithm for the assessment of knowledge (Doignon and Falmagne, 1999; Falmagne and Doignon, 2011). The parameters of the probabilistic model (i.e., π_K for each clinical state $K \in \mathcal{K}$, η_q and β_q for every $q \in \mathcal{Q}$) were estimated through an iterative procedure based on maximum likelihood (Dempster et al., 1977)

on data from the 383 participants. The estimated parameters were then used to calibrate the adaptive algorithm. The simulation study was carried out to test the efficiency and accuracy of the implemented adaptive procedure. Results supported that the adaptive version of the QuEDS provides clinicians with accurate information collected in an efficient way. Moreover, the information collected by means of the adaptive version of QuEDS allows the differentiation of individuals with the same score but with different symptoms (i.e., with different clinical states) and, possibly, different severity of the episode. These properties represent a relevant improvement in the amount, and quality of the collected diagnostic information, as well as in the amount of time needed for case formulation.

The parameter estimates provided the starting point for the implementation of the questioning rule and of the updating rule. This last was tested under different conditions with respect to the computation of the multiplicative parameter ζ . Finally, the opportunity to apply a Bayesian update was tested. Results showed that the most efficient and accurate implementation of the algorithm included the estimate of ζ via the η and β parameters, and the application of an on-line Bayesian updating.

It is important to highlight how the adaptive version of the questionnaire allows for a consistent reduction of the number of questions asked. In the classical written form of the QuEDS, each participant had to answer all 41 items, 15 for the cognitive sub-scale, 14 for the somatic sub-scale, 12 for the affective sub-scale. In the adaptive form of QuEDS only a percentage ranging between 50 and 70% of the items is asked.

The present study and the adaptive version of the QuEDS presents some limitations which should be addressed in future research. Although the sample size is adequate to obtain reliable estimates of the error parameters of the items and of the model fit, it is too small to achieve reliable estimates of the clinical states' probabilities. In fact, given the size of the three structures (respectively 124, 163, and 142 states) and the obtained error parameters of the items, a reliable estimate of the π_K parameters would need a sample of approximately 1,000 individuals. Notice that this limitation is not crucial since, in general, with large structures the *a priori* probability of each state is very low, thus in the adaptive form, the possibility of starting from a uniform distribution on the states is not that strong and generally

accepted. The second limitation of the present study is the relatively low number of patients in the clinical sub-sample. This limitation, which could appear critical in the perspective of classical methodologies in psychological measurement, within the framework of FPA is not that crucial, since the estimate of the parameters and the fit involve the sample in the whole rather than taking into account different sub-samples. Nonetheless, the recruitment of a greater number of patients to refine the estimates and the efficiency-efficacy of the adaptive version of the QuEDS will be the subject matter of future research. One final limitation deserves mention: The present version of the QuEDS does not contain any control scale for social desirability. The inclusion of social desirability scale into self-report tools for the assessment of depression is an important and debated issue (e.g., Langevin and Stancer, 1979; Pichot, 1986; Tanaka-Matsumi and Kameoka, 1986; Cappeliez, 1990; Balsamo and Saggino, 2007) and it will receive further attention in future research in order to provide users of the QuEDS with complete information about this specific issue.

To conclude, this new form of the QuEDS allows a clinician to differentiate the individual's depressive symptoms beyond the score and to administer only the items related to its symptomatology following the logical flow of question-answer. Thus, two patients who obtain the same score to the test can be treated differently according to their symptoms, since answering the same number of items does not mean having the same symptoms configuration.

The future directions of the development of the adaptive version of the QuEDS questionnaire are twofold: on the one hand it is necessary to improve the user interface in order to

achieve a simple graphical output able to provide the clinician with a helpful and accessible way to interact with the system. On the other hand, the formal definition of the suggestions for further investigation on the patient have to be formalized and implemented. This last issue is in continuity with the operational approach adopted by the Cognitive Behavioral Assessment 2.0 (CBA 2.0; Bertolotti et al., 1990; Sanavio et al., 2008), and represents the fundamental philosophical approach implemented by FPA methodology. Furthermore, several refinements of ATS-PD system can be implemented, for example the possibility of simplifying the updating rule for real-time application of QuEDS as implemented by Augustin et al. (2013). Another important future direction will be the extension of this approach to the case of polytomous items. The implementation of these extensions would allow FPA to be used with Likert scales, promoting its wider application in both psychological measurement and clinical practice.

AUTHOR CONTRIBUTIONS

AS defined the framework of the paper, conducted the parameter estimates analysis and prepared the final version of the manuscript. FS created the items of the questionnaire, supervised the definition of the formal contexts and contributed to the introduction and discussion of the paper. ID implemented the adaptive algorithm and tested the accuracy and efficiency of the procedure. UG prepared the description of the main deterministic and probabilistic issues of Formal Psychological Assessment, and prepared part of the introduction. GV supervised the whole research project.

REFERENCES

- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM-5, 5th Edn*. Washington, DC: American Psychiatric Association.
- Andersen, P. A. (1999). *Nonverbal Communication: Forms and Functions*. Mountain View, CA: Mayfield.
- Annen, S., Roser, P., and Brüne, M. (2012). Nonverbal behavior during clinical interviews: similarities and dissimilarities among schizophrenia, mania, and depression. *J. Nerv. Ment. Dis.* 200, 26–32. doi: 10.1097/NMD.0b013e31823e653b
- Augustin, T., Hockemeyer, C., Kickmeier-Rust, M. D., Podbregar, P., Suck, R., and Albert, D. (2013). The simplified updating rule in the formalization of digital educational games. *J. Comput. Sci.* 4, 293–303. doi: 10.1016/j.jocs.2012.08.020
- Baek, S.-G. (1997). Computerized adaptive testing using the partial credit model for attitude measurement. *Object. Meas.* 4:37.
- Balsamo, M., Giampaglia, G., and Saggino, A. (2014). Building a new rasch-based self-report inventory of depression. *Neuropsychiatr. Dis. Treat.* 10:153. doi: 10.2147/NDT.S53425
- Balsamo, M., Innamorati, M., Van Dam, N. T., Carlucci, L., and Saggino, A. (2015a). Measuring anxiety in the elderly: psychometric properties of the state trait inventory of cognitive and somatic anxiety (sticsa) in an elderly italian sample. *Int. Psychogeriatr.* 27, 999–1008. doi: 10.1017/S1041610214002634
- Balsamo, M., Macchia, A., Carlucci, L., Picconi, L., Tommasi, M., Gilbert, P., et al. (2015b). Measurement of external shame: an inside view. *J. Pers. Assess.* 97, 81–89. doi: 10.1080/00223891.2014.947650
- Balsamo, M., and Saggino, A. (2007). Test per l'assessment della depressione nel contesto italiano: un'analisi critica [tests for depression assessment in italian context: a critical review]. *Psicoter. Cogn. Comput.* 13:167.
- Benazzi, F., Helmi, S., and Bland, L. (2002). Agitated depression: unipolar? bipolar? or both? *Ann. Clin. Psychiatry* 14, 97–104. doi: 10.3109/10401230209149096
- Bertolotti, G., Zotti, A. M., Michielin, P., Vidotto, G., and Sanavio, E. (1990). A computerized approach to cognitive behavioural assessment: an introduction to cba-2.0 primary scales. *J. Behav. Ther. Exp. Psychiatry* 21, 21–27. doi: 10.1016/0005-7916(90)90045-M
- Birkhoff, G. (1937). Rings of sets. *Duke Math. J.* 3, 443–454. doi: 10.1215/S0012-7094-37-00334-X
- Birkhoff, G. (1940). *Lattice Theory*. Providence, RI: American Mathematical Society.
- Bottesi, G., Spoto, A., Freeston, M. H., Sanavio, E., and Vidotto, G. (2015). Beyond the score: clinical evaluation through formal psychological assessment. *J. Pers. Assess.* 97, 252–260. doi: 10.1080/00223891.2014.958846
- Cappeliez, P. (1990). Social desirability response set and self-report depression inventories in the elderly. *Clin. Gerontol.* 9, 45–52. doi: 10.1300/J018v09n02_06
- Davey, B. A., and Priestley, H. A. (2002). *Introduction to Lattices and Order*. Cambridge: Cambridge University Press.
- Dell'Osso, L., Armani, A., Rucci, P., Frank, E., Fagiolini, A., Corretti, G., et al. (2002). Measuring mood spectrum: comparison of interview (SCI-MOODS) and self-report (MOODS-SR) instruments. *Comprehens. Psychiatry* 43, 69–73. doi: 10.1053/comp.2002.29852
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B (Methodol.)* 39, 1–38.
- Doignon, J.-P., and Falmagne, J.-C. (1985). Spaces for the assessment of knowledge. *Int. J. Man Mach. Stud.* 23, 175–196. doi: 10.1016/S0020-7373(85)80031-6
- Doignon, J.-P., and Falmagne, J.-C. (1999). *Knowledge Spaces*. New York, NY: Springer.

- Donadello, I., Spoto, A., Sambo, F., Badaloni, S., Granzio, U., and Vidotto, G. (2017). Ats-pd: an adaptive testing system for psychological disorders. *Educ. Psychol. Meas.* 77, 792–815. doi: 10.1177/0013164416652188
- Eggen, T., and Straetmans, G. (2000). Computerized adaptive testing for classifying examinees into three categories. *Educ. Psychol. Meas.* 60, 713–734. doi: 10.1177/00131640021970862
- Embretson, S. E., and Reise, S. P. (2013). *Item Response Theory*. Psychology Press.
- Falmagne, J.-C., Albert, D., Doble, C., Eppstein, D., and Hu, X. (2013). *Knowledge Spaces: Applications in Education*. Berlin: Springer.
- Falmagne, J.-C., and Doignon, J.-P. (1988). A class of stochastic procedures for the assessment of knowledge. *Brit. J. Math. Stat. Psychol.* 41, 1–23. doi: 10.1111/j.2044-8317.1988.tb00884.x
- Falmagne, J.-C., and Doignon, J.-P. (2011). *Learning Spaces*. New York, NY: Springer.
- Fechner, G. T. (1860). *Elemente der Psychophysik, Vol. 2*. Leipzig: Breitkopf und Härtel.
- Finkelman, M. D., Smits, N., Kim, W., and Riley, B. (2012). Curtailment and stochastic curtailment to shorten the CES-D. *Appl. Psychol. Meas.* 36, 632–658. doi: 10.1177/0146621612451647
- Fiquer, J. T., Boggio, P. S., and Gorenstein, C. (2013). Talking bodies: nonverbal behavior in the assessment of depression severity. *J. Affect. Disord.* 150, 1114–1119. doi: 10.1016/j.jad.2013.05.002
- First, M., Gibbon, M., Spitzer, R. L., and Williams, J. (1996). *User's Guide for the Structured Clinical Interview for DSM-IV Axis I Disorders' Research Version*. New York, NY: Biometrics Research Department, New York State Psychiatric Institute.
- Fliege, H., Becker, J., Walter, O. B., Bjorner, J. B., Klapp, B. F., and Rose, M. (2005). Development of a computer-adaptive test for depression (D-CAT). *Qual. Life Res.* 14:2277. doi: 10.1007/s11136-005-6651-9
- Ganter, B., and Wille, R. (1999). *Formal Concept Analysis: Mathematical Foundations*. Berlin; Heidelberg: Springer Verlag.
- Geerts, E., and Brüne, M. (2009). Ethological approaches to psychiatric disorders: focus on depression and schizophrenia. *Aust. N.Z. J. Psychiatry* 43, 1007–1015. doi: 10.1080/00048670903270498
- Gibbons, R. D., Clark, D. C., Cavanaugh, S. V., and Davis, J. M. (1985). Application of modern psychometric theory in psychiatric research. *J. Psychiatr. Res.* 19, 43–55. doi: 10.1016/0022-3956(85)90067-6
- Gibbons, R. D., Weiss, D. J., Kupfer, D. J., Frank, E., Fagioli, A., Grochocinski, V. J., et al. (2008). Using computerized adaptive testing to reduce the burden of mental health assessment. *Psychiatr. Serv.* 59, 361–368. doi: 10.1176/ps.2008.59.4.361
- Girard, J. M., Cohn, J. F., Mahoor, M. H., Mavadati, S. M., Hammal, Z., and Rosenwald, D. P. (2014). Nonverbal social withdrawal in depression: evidence from manual and automatic analyses. *Image Vis. Comput.* 32, 641–647. doi: 10.1016/j.imavis.2013.12.007
- Goodwin, F. K., and Jamison, K. R. (2007). *Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression, Vol. 1*. Oxford: Oxford University Press.
- Granzio, U., Spoto, A., and Vidotto, G. (2018). The assessment of nonverbal behavior in schizophrenia through the formal psychological assessment. *Int. J. Methods Psychiatr. Res.* 27:e1595. doi: 10.1002/mpr.1595
- Grayce, C. J. (2013). “A commercial implementation of knowledge space theory in college general chemistry,” in *Knowledge Spaces: Applications in Education*, eds J.-C. Falmagne, D. Albert, C. Doble, D. Eppstein, and X. Hu (New York, NY: Springer), 93–113.
- Groth-Marnat, G. (2009). *Handbook of Psychological Assessment*. New York, NY: John Wiley & Sons.
- Gulliksen, H. (2013). *Theory of Mental Tests*. New York, NY: Routledge.
- Innamorati, M., Tamburello, S., Contardi, A., Imperatori, C., Tamburello, A., Saggino, A., et al. (2013). Psychometric properties of the attitudes toward self-revised in Italian young adults. *Depress. Res. Treat.* 2013:209216. doi: 10.1155/2013/209216
- Langevin, R., and Stancer, H. (1979). Evidence that depression rating scales primarily measure a social undesirability response set. *Acta Psychiatr. Scand.* 59, 70–79. doi: 10.1111/j.1600-0447.1979.tb06948.x
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. New York, NY: Routledge.
- Marinagi, C. C., Kaburlasos, V. G., and Tsoukalas, V. T. (2007). “An architecture for an adaptive assessment tool,” in *Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual, T3D-11*. (Milwaukee, WI: IEEE).
- Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L., et al. (2018). An introduction to network psychometrics: relating ising network models to item response theory models. *Multivar. Behav. Res.* 53, 15–35. doi: 10.1080/00273171.2017.1379379
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: a review of evidence and issues. *Amer. Psychol.* 56:128. doi: 10.1037/0003-066X.56.2.128
- Narens, L., and Luce, R. D. (1993). Further comments on the nonrevolution arising from axiomatic measurement theory. *Psychol. Sci.* 4, 127–130. doi: 10.1111/j.1467-9280.1993.tb00475.x
- Nokelainen, P., Silander, T., Tirri, H., Nevgi, A., and Tirri, K. (2001). “Modeling students' views on the advantages of web-based learning with bayesian networks,” in *Proceedings of the Tenth International PEG Conference: Intelligent Computer and Communications Technology-Learning in On-Line Communities (Tampere)*, 101–108.
- Nordgaard, J., Sass, L. A., and Parnas, J. (2013). The psychiatric interview: validity, structure, and subjectivity. *Eur. Arch. Psychiatry Clin. Neurosci.* 263, 353–364. doi: 10.1007/s00406-012-0366-z
- Novick, M. R. (1965). The axioms and principal results of classical test theory. *ETS Res. Rep. Ser.* 1965, 1–18. doi: 10.1002/j.2333-8504.1965.tb00132.x
- Petersen, M. A., Groenvold, M., Aaronson, N., Fayers, P., Sprangers, M., Bjorner, J. B., et al. (2006). Multidimensional computerized adaptive testing of the EORTC QLQ-C30: basic developments and evaluations. *Qual. Life Res.* 15, 315–329. doi: 10.1007/s11136-005-3214-z
- Philippot, P., Schaefer, A., and Herbette, G. (2003). Consequences of specific processing of emotional information: impact of general versus specific autobiographical memory priming on emotion elicitation. *Emotion* 3:270. doi: 10.1037/1528-3542.3.3.270
- Pichot, P. (1986). “A self-report inventory on depressive symptomatology (qd2) and its abridged form (qd2a),” in *Assessment of Depression*, eds N. Sartorius and T. A. Ban (Berlin; Heidelberg: Springer), 108–122.
- Rasch, G. (1960). *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests*. Oxford: Nielsen & Lydiche.
- Reise, S. P., and Waller, N. G. (1990). Fitting the two-parameter model to personality data. *Appl. Psychol. Meas.* 14, 45–58. doi: 10.1177/014662169001400105
- Sanavio, E., Bertolotti, G., Michielin, P., Vidotto, G., and Zotti, A. (2008). *Cba-2.0 Scale Primarie: Manuale. Una Batteria ad Ampio Spettro per l'Assessment Psicologico*. Firenze: Organizzazioni Speciali.
- Serra, F., Spoto, A., Ghisi, M., and Vidotto, G. (2015). Formal psychological assessment in evaluating depression: a new methodology to build exhaustive and irredundant adaptive questionnaires. *PLoS ONE* 10:e0122131. doi: 10.1371/journal.pone.0122131
- Serra, F., Spoto, A., Ghisi, M., and Vidotto, G. (2017). Improving major depressive episode assessment: a new tool developed by formal psychological assessment. *Front. Psychol.* 8:214. doi: 10.3389/fpsyg.2017.00214
- Simms, L. J., Goldberg, L. R., Roberts, J. E., Watson, D., Welte, J., and Rotterman, J. H. (2011). Computerized adaptive assessment of personality disorder: introducing the cat-pd project. *J. Pers. Assess.* 93, 380–389. doi: 10.1080/00223891.2011.577475
- Smits, N., Finkelman, M. D., and Kelderman, H. (2016). Stochastic curtailment of questionnaires for three-level classification: shortening the ces-d for assessing low, moderate, and high risk of depression. *Appl. Psychol. Meas.* 40, 22–36. doi: 10.1177/0146621615592294
- Spearman, C. (1904). “General Intelligence”, objectively determined and measured. *Am. J. Psychol.* 15, 201–292.
- Spiegel, R., and Nenih, Y. (2004). An expert system supporting diagnosis in clinical psychology. *WIT Trans. Inform. Commun. Technol.* 31, 145–154. doi: 10.2495/CI040151
- Spoto, A. (2011). *Formal Psychological Assessment Theoretical and Mathematical Foundations*. Doctoral Dissertation, University of Padua. Available online at: <http://paduaresearch.cab.unipd.it/3477>

- Spoto, A., Bottesi, G., Sanavio, E., and Vidotto, G. (2013a). Theoretical foundations and clinical implications of formal psychological assessment. *Psychother. Psychosomat.* 82, 197–199. doi: 10.1159/000345317
- Spoto, A., Stefanutti, L., and Vidotto, G. (2010). Knowledge space theory, formal concept analysis, and computerized psychological assessment. *Behav. Res. Methods* 42, 342–350. doi: 10.3758/BRM.42.1.342
- Spoto, A., Stefanutti, L., and Vidotto, G. (2012). On the unidentifiability of a certain class of skill multi map based probabilistic knowledge structures. *J. Math. Psychol.* 56, 248–255. doi: 10.1016/j.jmp.2012.05.001
- Spoto, A., Stefanutti, L., and Vidotto, G. (2013b). Considerations about the identification of forward-and backward-graded knowledge structures. *J. Math. Psychol.* 57, 249–254. doi: 10.1016/j.jmp.2013.09.002
- Spoto, A., Stefanutti, L., and Vidotto, G. (2016). An iterative procedure for extracting skill maps from data. *Behav. Res. Methods* 48, 729–741. doi: 10.3758/s13428-015-0609-9
- Stefanutti, L., Spoto, A., and Vidotto, G. (2018). Detecting and explaining blims unidentifiability: forward and backward parameter transformation groups. *J. Math. Psychol.* 82, 38–51. doi: 10.1016/j.jmp.2017.11.001
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science* 103, 677–680. doi: 10.1126/science.103.2684.677
- Stevens, S. S. (1951). *Mathematics, Measurement, and Psychophysics*. Oxford: Wiley.
- Stevens, S. S. (1957). On the psychophysical law. *Psychol. Rev.* 64:153. doi: 10.1037/h0046162
- Strull, W. M., Lo, B., and Charles, G. (1984). Do patients want to participate in medical decision making? *J. Am. Med. Assoc.* 252, 2990–2994. doi: 10.1001/jama.1984.03350210038026
- Suppes, P., Krantz, D. H., Luce, R. D., and Tversky, A. (1989). *Foundations of Measurement: Geometrical, Threshold, and Probabilistic Representations, Vol. 2*. New York, NY: Academic Press.
- Suppes, P., and Zinnes, J. (1963). “Basic measurement theory,” In *Handbook of Mathematical Psychology*, eds R. D. Luce and E. Galanter (Wiley, Oxford), 1–76.
- Tanaka-Matsumi, J., and Kameoka, V. A. (1986). Reliabilities and concurrent validities of popular self-report measures of depression, anxiety, and social desirability. *J. Consult. Clin. Psychol.* 54:328. doi: 10.1037/0022-006X.54.3.328
- Uhlmann, E. L., Leavitt, K., Menges, J. I., Koopman, J., Howe, M., and Johnson, R. E. (2012). Getting explicit about the implicit: a taxonomy of implicit measures and guide for their use in organizational research. *Organ. Res. Methods* 15, 553–601. doi: 10.1177/1094428112442750
- Wainer, H. (2000). *CATs: Whither and Whence*. Princeton, NJ: ETS Research Report Series.
- Wille, R. (1982). “Restructuring lattice theory: an approach based on hierarchies of concepts,” in *Ordered Sets*, ed I. Rival (Dordrecht: Reidel), 445–470.
- Yong, S., Awang Rambli, D. R., and Anh, N. (2007). “Depression consultant expert system,” in *6th Annual Seminar on Science and Technology* (Tawau).

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

Copyright © 2018 Spoto, Serra, Donadello, Granzol and Vidotto. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read for greatest visibility and readership



FAST PUBLICATION

Around 90 days from submission to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative, and constructive peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers acknowledged by name on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data and methods to enhance research reproducibility



DIGITAL PUBLISHING

Articles designed for optimal readership across devices



FOLLOW US

[@frontiersin](https://www.instagram.com/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