POSITIVE PSYCHOLOGICAL ASSESSMENTS: MODERN APPROACHES, METHODOLOGIES, MODELS AND GUIDELINES

EDITED BY: Llewellyn Ellardus Van Zyl, Peter ten Klooster, Leon T. de Beer, Mariëlle Zondervan-Zwijnenburg, Maria Vera and Arianna Costantini

PUBLISHED IN: Frontiers in Psychology, Frontiers in Public Health and Frontiers in Psychiatry

frontiers Research Topics

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-83250-448-2
DOI 10.3389/978-2-83250-448-2

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

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

What are Frontiers Research Topics?

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

POSITIVE PSYCHOLOGICAL ASSESSMENTS: MODERN APPROACHES, METHODOLOGIES, MODELS AND GUIDELINES

Topic Editors:
Llewellyn Ellardus Van Zyl, North West University, South Africa
Peter ten Klooster, University of Twente, Netherlands
Leon T. de Beer, North-West University, South Africa
Mariëlle Zondervan-Zwijnenburg, Vilans, Netherlands
Maria Vera, Universidad Pablo de Olavide, Spain
Arianna Costantini, University of Trento, Italy

Citation: Van Zyl, L. E., ten Klooster, P., de Beer, L. T., Zondervan-Zwijnenburg, M., Vera, M., Costantini, A., eds. (2022). Positive Psychological Assessments: Modern Approaches, Methodologies, Models and Guidelines. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-448-2

Table of Contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 26 September 2022
DOI 10.3389/fpsyg.2022.1020653

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Chee-Seng Tan,
Universiti Tunku Abdul Rahman, Malaysia

*CORRESPONDENCE
Llewellyn Ellardus van Zyl
llewellyn101@gmail.com

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

RECEIVED 16 August 2022
ACCEPTED 25 August 2022
PUBLISHED 26 September 2022

CITATION
Costantini A, De Beer LT, Klooster PMt, Zondervan-Zwijnenburg MAJ, Vera M and van Zyl LE (2022) Editorial: Positive Psychological Assessments: Modern Approaches, Methodologies, Models and Guidelines: Current perspectives. *Front. Psychol.* 13:1020653. doi: 10.3389/fpsyg.2022.1020653

COPYRIGHT
© 2022 Costantini, De Beer, Klooster, Zondervan-Zwijnenburg, Vera and van Zyl. 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.

Editorial: Positive Psychological Assessments: Modern Approaches, Methodologies, Models and Guidelines: Current perspectives

Arianna Costantini[1], Leon T. De Beer[2], Peter M. ten Klooster[3], Marielle A. J. Zondervan-Zwijnenburg[4], Maria Vera[5] and Llewellyn Ellardus van Zyl[6,7,8,9]*

[1]Department of Psychology and Cognitive Science, University of Trento, Trento, Italy, [2]WorkWell Research Unit, North-West University, Potchefstroom, South Africa, [3]Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands, [4]Department of Methodology and Statistics, Utrecht University, Utrecht, Netherlands, [5]Social Psychology Department, Universidad Pablo de Olavide, Sevilla, Spain, [6]Department of Industrial Engineering and Innovation Sciences, University of Eindhoven, Eindhoven, Netherlands, [7]Optentia Research Unit, North-West University, Vanderbijlpark, South Africa, [8]Department of Human Resource Management, University of Twente, Enschede, Netherlands, [9]Department of Social Psychology, Institut für Psychologie, Goethe University, Frankfurt am Main, Germany

KEYWORDS

applied positive psychology, mental health, positive psychology, positive psychological assessment, positive psychology coaching, scale development, wellbeing, wellbeing assessment

Editorial on the Research Topic
Positive Psychological Assessments: Modern Approaches, Methodologies, Models and Guidelines

Introduction

Sparked by evidence showing that positive psychological approaches and practices not only foster flourishing but also help to reduce mental illness, maintain mental health, and strengthen one's psychological resources and capacities (Waters et al., 2022), positive psychology interventions and coaching have emerged as popular approaches for practitioners interested in the development and wellbeing of people (Lomas, 2020; Moskowitz et al., 2021; Richter et al., 2021). Indeed, bourgeoning evidence for the social, behavioral, and physical health benefits of positive psychology constructs (Donaldson et al., 2021; Moskowitz et al., 2021) has led to an increasing influence of positive psychology underpinning practice (Green and Palmer, 2019), with significant growth of the use of positive psychological assessment measures (PPAMs).

Despite such popularity, research has also highlighted the shortcomings of existing PPAMs (e.g., Wong and Roy, 2017; van Zyl and Rothmann, 2022), with important implications for a valid and reliable assessment of the effectiveness of positive psychology

practices as well as the advancement of our understanding of wellbeing through the development and conceptualization of new positive psychological constructs (Gruman et al., 2018; van Zyl and Rothmann, 2022; Van Zyl and Salanova, 2022). For example, it has been argued that the broad category of wellbeing, encompassing independent and separable components (Diener, 1984; Ng et al., 2021), is inconsistently operationalized across studies, making it difficult to determine whether positive psychology interventions have stronger effects on particular aspects of wellbeing compared to others (Moskowitz et al., 2021). Similarly, across different studies, diverse positive constructs can be found grouped together, for example, combining positive emotions with outcomes (e.g., meaning, purpose, life satisfaction; Sin and Lyubomirsky, 2009; Chakhssi et al., 2018) or with other cognitive and affective appraisals of one's life as a whole (Bolier et al., 2013; Hendriks et al., 2019; Moskowitz et al., 2021).

Given the key role of PPAMs in advancing the science and practice of positive psychology (van Zyl and Rothmann, 2022), this Research Topic specifically focused on collecting evidence and informed proposals of modern approaches, methodologies, models, and guidelines for PPAMs.

Structure and contribution of the Research Topic

The contributions included in this Research Topic are summarized in Table 1 and presented below. In summary, responding to our call for more thorough evidence on the validity and reliability of both newly developed and translated popular PPAMs, nine manuscripts included in this Research Topic focused on investigating the psychometric properties of different PPAMs using a variety of modern statistical modeling techniques and across different cultural contexts. Moreover, two contributions used survey data to provide an example of best practice guidelines and investigate the properties of positive psychological constructs using modern approaches. Finally, one contribution presented a systematic review of observational PPAMs developed to assess momentary wellbeing in people living with dementia.

Scale development and validation

Bauman and Ruch presented a study on the development and validation of the Fulfilled Life Scale, capturing people's experience of a fulfilled life. Using data from three different German-speaking samples (development sample $n = 282$; replication sample $n = 406$; selected exemplar participants $n = 39$), they identified three optimal factors across cognitive and affective experiences of a fulfilled life, labeled *unfolded self and life*, the *worthwhile life*, and *positive impact and legacy*. Cognitive and affective fulfillment incrementally predicted a global rating of a fulfilled life and mental wellbeing, even after controlling for subjective and eudaimonic wellbeing.

Carmona-Halty et al. presented an adaptation of the Flourishing Scale (FS) to a Chilean high school context and provided evidence of its validity. Using a cross-sectional sample of 1,348 students from three different schools in Chile, they showed that their adapted version of the FS is invariant across genders and is positively related with study–related positive feelings (i.e., happiness, pleasure, and satisfaction) and negatively related with study–related negative feelings (i.e., sadness, displeasure, and anger).

Focusing on behaviors aimed to better align life goals, personal needs, values, and capabilities, Chen et al. proposed a conceptualization of life crafting and developed, validated, and evaluated a measure of overall life crafting, the Life Crafting Scale (LCS). Using a mixed-method, multi-study research design, in the first qualitative phase, they created a pool of items; then, in Study 1, involving 331 English-speaking employees, they found support for a three-factor structure encompassing *cognitive crafting*, *seeking social support*, and *seeking challenges*. In Study 2, involving 362 employees in the Netherlands, the factorial structure of the scale was confirmed, and the LCS showed to be a reliable tool, partially invariant across genders, and positively associated with meaning in life, mental health, and work engagement, while negatively related to job burnout.

Cromhout et al. zoomed into the construct of eudaimonic wellbeing and used various analytical models, including CFA, bifactor CFA, Exploratory Structural Equation Modeling (ESEM), and bifactor ESEM, to investigate the dimensionality of the Questionnaire for Eudaimonic WellBeing (QEWB) in four culturally diverse South African samples, including three student samples (English, $n = 326$; Afrikaans, $n = 478$; and Setswana, $n = 260$) and one multicultural adult sample ($n = 262$). Their results showed that eudaimonic wellbeing is multidimensional but, at the same time, represents an overarching higher-order construct. Moreover, they found configural invariance across the different languages in which the QEWB was administered, but also that the QEWB shows differential psychometric properties across different age groups and developmental phases.

Espejo et al. focused on the Satisfaction with Life Scale (SWLS) with five response options and investigated its psychometric properties in a Colombian sample of 1,255 participants. Their results showed that the SWLS, in its Spanish version used in Colombia, is a reliable and valid tool displaying excellent psychometric properties and invariance across genders and age groups. Also, as expected, it correlates significantly with life satisfaction, flourishing, positive and negative affect, optimism, and pessimism.

Guitard et al. provided a thorough investigation of the Temporal Satisfaction with Life Scale (TSWLS) structure and number of optimal items. Based on a large international and

TABLE 1 Characteristics of the studies included in the Research Topic.

References	Construct	PPAM	Dimensions & *N* of items	Participants	Country	Language PPAM used in the research
Scale development and validation						
Bauman and Ruch	Fulfillment in life	Fulfilled Life Scale (FLS)	Fulfilled Life Cognitive Experience (24) Fulfilled Life Affective Experience (8)	Sample 1: 282 adults (50–93 yo) Sample 2: 406 adults (40–85 yo) Sample 3: 39 adults (41–89 yo)	German-speaking countries	German
Carmona-Halty et al.	Flourishing	Flourishing Scale (FS) adapted to the educational setting	Flourishing (8)	1,348 students (13–18 yo)	Chile	Spanish
Chen et al.	Life crafting	Life Crafting Scale (LCS)	Cognitive Crafting (3) Seeking Social Support (3) Seeking Challenges (3)	Sample 1: 331 employees Sample 2: 362 employees	United Kingdom, Portugal, Poland, The Netherlands	English
Cromhout et al.	Eudaimonic wellbeing	Questionnaire for Eudaimonic WellBeing (QEWB)	Sense of Purpose (7) Purposeful Personal Expressiveness (9) Effortful Engagement (5)	Sample 1: 326 univ students Sample 2: 478 univ students Sample 3: 260 univ students Sample 4: 262 adults	South Africa	Sample 1, 4: English Sample 2: Afrikaans Sample 3: Setswana
Espejo et al.	Satisfaction with life	Satisfaction with Life Scale (SWLS)	Satisfaction with Life (5)	1,255 adults	Colombia	Spanish
Guitard et al.	Satisfaction with life	Temporal Satisfaction with Life Scale (TSWLS)	Past Life Satisfaction (4) Present Life Satisfaction (4) Future Life Satisfaction (4)	3,982 Eng-speakers (over 16 yo) 2,930 non-Eng speakers (over 16 yo)	Worldwide	English Hungarian Spanish Finnish Slovene Czech Chinese
van Zyl et al.	Strengths use	Strengths Use Scale (SUS)	Affinity for Strengths (6) Strengths Use Behaviors (8)	360 univ students	The Netherlands	Dutch
Youssef-Morgan et al.	Work gratitude	Work Gratitude Scale (WGS)	Grateful appraisals (3) Gratitude toward others (4) Intentional attitude of gratitude (3)	625 employees	USA	English
Zábó et al.	Mental health	Mental Health Test (MHT)	Wellbeing (3) Savoring (3) Creative and Executive Efficiency (5) Self-Regulation (3) Resilience (4)	Sample 1: 1,736 adults Sample 2: 1,083 adults	Hungary	Hungarian
Guidelines and Survey Design Papers						
van Zyl and ten Klooster	Mental health	Mental Health Continuum-Short Form (MHC-SF)	Emotional wellbeing (3) Social wellbeing (5) Psychological wellbeing (6)	1,804 adults	The Netherlands	Dutch
Ratchford et al.	Mindsets across domains	Implicit Theories of Morality and Intelligence Scale	Morality mindset (3) Ability mindset (3)	618 adolescents (15–19 yo)	United States	English
Systematic review						
Madsø et al.	Momentary wellbeing	Observational instruments assessing momentary wellbeing		People with dementia		

The dimensions and number of items reported are based on the results of the studies. yo, years old; univ, university; Er

10.3389/fpsyg.2022.1020653

multicultural sample ($n = 6912$), their findings showed that a 12-item version of the scale was optimal compared to the original 15-item one, and it was equivalent and valid across English speakers in different geographic regions of the world, including Oceania, North America, Europe, and Asia. Also, their results showed that six different translations of the TSWLS function in similar ways, yet some differences exist in item functioning across cultures. All three subscales of the TSWLS, that is, past, present, and future, displayed positive correlations with aspects of wellbeing (strengths use and knowledge, subjective happiness, gratitude, hope, and the presence of meaning in life) and negative ones with aspects of ill-being (search for meaning in life, rumination, depression), as expected.

Van Zyl et al. investigated the psychometric properties, longitudinal invariance, and criterion validity of the Strengths Use Scale (SUS) within 360 students in the Netherlands. Their results showed that the SUS comprises two first-order factors, namely *affinity for strengths* and *strengths use behaviors*. This factorial structure showed to be consistent across time, and longitudinal evidence showed that strengths use remained stable over time. Moreover, strengths use predicted study engagement assessed 3 months after, providing evidence of the criterion validity of the SUS.

Youssef-Morgan et al. introduced the new construct of work gratitude, defined as "the intentional choice to engage in positive appraisals and feelings of thankfulness and appreciation toward the characteristics, situations, and people currently present in the work context," and presented a new instrument, the Work Gratitude Scale (WGS), to assess it. Using cross-sectional data from 625 employees from a school district in the United States, they found support for the validity of a second-order model of work gratitude with three underlying dimensions: *grateful appraisals*, *gratitude toward others*, and *intentional attitude of gratitude*. The WGS showed to be a valid and reliable instrument, useful to spark research on how to promote grateful appraisals, gratitude toward others, and intentional attitudes of gratitude in employees.

Zabo et al. presented a new five-scale mental health test, the Mental Health Test (MHT), that operationalizes a set of indicators of a newly introduced concept of positive mental health. Based on cross-sectional self-reported data collected in Hungary, they found support for a five-factor structure with 17 items ($n = 1736$), which was confirmed in a separate sample ($n = 1083$). The MHT maps aspects of wellbeing, savoring, creative and executive efficiency, self-regulation, and resilience. Results showed that the MHT displayed a high level of internal consistency, and correlations confirmed the content validity of the subscales with established measures of psychological wellbeing. Moreover, test-retest reliability was confirmed by longitudinal data collected after 2 weeks and again after 11 months. Overall, their results showed that the MHT could be considered a new reliable, valid measurement tool to assess several aspects of mental health.

Guidelines and survey design papers

Using data collected on the Mental Health Continuum-Short Form (MHC-SF) as an illustrative example, van Zyl and ten Klooster provided a practical tutorial on how to use Exploratory Structural Equation Modeling (ESEM) with a convenient online tool for M*plus*. In their paper, they illustrated the applicability of the ESEM as an alternative to traditional CFA approaches and provided an overview of ESEM and structured guidelines on how to use and apply ESEM models, including a step-by-step guide to producing ESEM syntax to be used with M*plus*. Contributing to the literature on mental health assessment, they further showed that when measuring mental health with the MHC-SF, an approach that accounts for a bifactor ESEM model should be preferred to CFA models.

Focusing on mindsets across different domains, Ratchford et al. contributed evidence to the debate about domain specificity and generality of mindset while exploring how cohesion and divergence across moral and ability mindsets affect self-system, self-regulatory strategies, and wellbeing outcomes in a sample of 618 adolescents in metropolitan southern California. To assess congruence and discrepancies, they used response surface analysis to consider the within-person effects of domain specificity across various outcomes. Their findings showed that overall congruency between moral and ability mindsets did not relate significantly with any of the wellbeing outcomes considered, suggesting that ability and moral mindsets are distinct qualities for which congruence is not relevant for wellbeing. Hence, by showing that mindsets display high levels of domain specificity, this study offers implications for the assessment of mindsets as characteristic adaptations, suggesting that different mindsets should therefore be assessed and accounted for independently in future survey designs.

Systematic review paper

Madsø et al. presented a systematic review of 36 articles describing 22 observational instruments assessing momentary wellbeing in people with dementia. The instruments included in the review mapped three categories: observations of emotions, observations of positive behavioral expressions, and observations of engagement. Their analysis included risk of bias at the study level and assessment of measurement properties at the instrument level (content validity, construct validity, structural validity, internal consistency, measurement invariance, cross-cultural validity, measurement error, and inter-rater/intra-rater/test-retest reliability and responsiveness). Results showed that among the instruments included in the review, 11 were supported by high-quality evidence for content validity, while the presence of high-quality evidence of other

central psychometric aspects was sparse. However, several instruments have the potential to meet such quality criteria if further investigated.

Conclusion

While much research in positive psychology is primarily situated within a positivistic paradigm and adopts quantitative designs (Rich, 2017; Gruman et al., 2018; Lomas et al., 2021), there is still a need for a more robust understanding of the properties of—both widespread and newly developed—PPAMs, so as to enhance the credibility of the discipline and our knowledge and impact of positive psychology (van Zyl and Rothmann, 2022). In an effort to address these challenges, this Research Topic provided a collection of contributions on assessment tools and operationalizations of positive psychology constructs, allowing us to gauge evidence regarding different approaches and instruments needed to understand the conditions and processes that foster optimal functioning and flourishing in people, groups, and institutions.

Author contributions

AC drafted the first version of the editorial. LvZ edited the manuscript. All authors provided conceptual input and approved the final draft.

Conflict of interest

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

Publisher's note

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

References

Bolier, L., Haverman, M., Westerhof, G. J., Riper, H., Smit, F., and Bohlmeijer, E. (2013). Positive psychology interventions: a meta-analysis of randomized controlled studies. *BMC Public Health* 13, 119. doi: 10.1186/1471-2458-13-119

Chakhssi, F., Kraiss, J. T., Sommers-Spijkerman, M., and Bohlmeijer, E. T. (2018). The effect of positive psychology interventions on wellbeing and distress in clinical samples with psychiatric or somatic disorders: a systematic review and meta-analysis. *BMC Psychiatr.* 18, 211. doi: 10.1186/s12888-018-1739-2

Diener, E. (1984). Subjective wellbeing. *Psychologic. Bull.* 95, 542–575. doi: 10.1037/0033-2909.95.3.542

Donaldson, S. I., Van Zyl, L. E., and Donaldson, S. I. (2021). PERMA+4: A framework for work-related wellbeing, performance and positive organizational psychology 2.0. *Front. Psychol.* 12, 817244. doi: 10.3389/fpsyg.2021.817244

Green, S., and Palmer, S. (2019). Positive psychology coaching: Science into practice. In *Positive psychology coaching in practice* (pp. 1-15). Routledge. doi: 10.4324/9781315716169-1

Gruman, J. A., Lumley, M. N., and González-Morales, M. G. (2018). Incorporating balance: challenges and opportunities for positive psychology. *Canadian Psychol./Psychologie Canadienne* 59, 54–64. doi: 10.1037/cap0000109

Hendriks, T., Schotanus-Dijkstra, M., Hassankhan, A., de Jong, J., and Bohlmeijer, E. (2019). The efficacy of multi-component positive psychology interventions: a systematic review and meta-analysis of randomized controlled trials. *J. Happiness Stud.* 19, 82. doi: 10.1007/s10902-019-00082-1

Lomas, T. (2020). Positive coaching psychology: A case study in the hybridization of positive psychology. *Int. J. Wellbeing* 10, 37–52. doi: 10.5502/ijw.v10i2.1083

Lomas, T., Waters, L., Williams, P., Oades, L. G., and Kern, M. L. (2021). Third wave positive psychology: broadening towards complexity. *J. Positiv. Psychol.* 16, 660–674. doi: 10.1080/17439760.2020.1805501

Moskowitz, J. T., Cheung, E. O., Freedman, M., Fernando, C., Zhang, M. W., Huffman, J. C., et al. (2021). Measuring positive emotion outcomes in positive psychology interventions: a literature review. *Emot. Rev.* 13, 60–73. doi: 10.1177/1754073920950811

Ng, W., Tov, W., Veenhoven, R., Rothmann, S., Chambel, M. J., Chen, S., et al. (2021). In memory of Edward Diener: reflections on his career, contributions and the science of happiness. *Front. Psychol.* 12, 706447 doi: 10.3389/fpsyg.2021.706447

Rich, G. J. (2017). The promise of qualitative inquiry for positive psychology: diversifying methods. *J. Positiv. Psychol.* 12, 220–231. doi: 10.1080/17439760.2016.1225119

Richter, S., Van Zyl, L. E., Roll, L. C., and Stander, M. W. (2021). Positive psychological coaching tools and techniques: a systematic review and classification. *Front. Psychiatr.* 12, 667200. doi: 10.3389/fpsyt.2021.667200

Sin, N. L., and Lyubomirsky, S. (2009). Enhancing wellbeing and alleviating depressive symptoms with positive psychology intervention: a practice-friendly meta-analysis. *J. Clinic. Psychol.* 65, 467–487. doi: 10.1002/jclp.20593

van Zyl, L. E., and Rothmann, S. (2022). Grand challenges for positive psychology: future perspectives and opportunities. *Front. Psychol.* 13, 833057. doi: 10.3389/fpsyg.2022.833057

Van Zyl, L. E., and Salanova, M. (2022). Facilitating the third wave of positive psychology: perspectives on the future of the discipline. *Front. Psychol.* 13, 167. doi: 10.3389/fpsyg.2022.963167

Waters, L., Algoe, S. B., Dutton, J., Emmons, R., Fredrickson, B. L., Heaphy, E., et al. (2022). Positive psychology in a pandemic: buffering, bolstering, and building mental health. *J. Positiv. Psychol.* 17, 303–323. doi: 10.1080/17439760.2021.1871945

Wong, P. T. P., and Roy, S. (2017). "Critique of positive psychology and positive interventions," in *The Routledge International Handbook of Critical Positive Psychology*, eds N. J. L. Brown, T. Lomas, and F. J. Eiroa-Orosa (Routledge: Taylor and Francis Group) 142–160.

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 June 2021
doi: 10.3389/fpsyg.2021.676153

The Strengths Use Scale: Psychometric Properties, Longitudinal Invariance and Criterion Validity

Llewellyn E. van Zyl[1,2,3,4]*, *Diane Arijs*[5], *Matthew L. Cole*[6], *Aldona Gliińska-Newes*[7], *Lara C. Roll*[2,5], *Sebastiaan Rothmann*[2], *Rebecca Shankland*[8], *Jacqueline M. Stavros*[6] and *Nicolas B. Verger*[9]

[1] Department of Industrial Engineering & Innovation Sciences, Human Performance Management, University of Eindhoven, Eindhoven, Netherlands, [2] Optentia Research Focus Area, North-West University Vaal Triangle Campus (VTC), Vanderbijlpark, South Africa, [3] Department of Human Resource Management, University of Twente, Enschede, Netherlands, [4] Department of Social Psychology, Institut für Psychologie, Goethe University, Frankfurt, Germany, [5] Department of Work and Organisation Studies, KU Leuven, Leuven, Belgium, [6] College of Business and Information Technology, Lawrence Technological University, Southfield, MI, United States, [7] Faculty of Economic Sciences and Management, Nicolaus Copernicus University, Torun, Poland, [8] DIPHE, Department of Psychology, Education and Vulnerabilities, Université Lumière Lyon 2, Lyon, France, [9] Department of Psychology, School of Life and Health Sciences, Glasgow Caledonian University, Glasgow, United Kingdom

OPEN ACCESS

Edited by:
Jeff Klibert,
Georgia Southern University,
United States

Reviewed by:
Sai-fu Fung,
City University of Hong Kong, China
Willibald Ruch,
University of Zurich, Switzerland

***Correspondence:**
Llewellyn E. van Zyl
llewellyn101@gmail.com

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

Received: *04 March 2021*
Accepted: *11 May 2021*
Published: *15 June 2021*

Citation:
van Zyl LE, Arijs D, Cole ML, Gliińska-Newes A, Roll LC, Rothmann S, Shankland R, Stavros JM and Verger NB (2021) The Strengths Use Scale: Psychometric Properties, Longitudinal Invariance and Criterion Validity. Front. Psychol. 12:676153. doi: 10.3389/fpsyg.2021.676153

Strengths use is an essential personal resource to consider when designing higher-educational programs and interventions. Strengths use is associated with positive outcomes for both the student (e.g., study engagement) and the university (e.g., academic throughput/performance). The Strengths Use Scale (SUS) has become a popular psychometric instrument to measure strengths use in educational settings, yet its use has been subjected to limited psychometric scrutiny outside of the U.S. Further, its longitudinal stability has not yet been established. Given the wide use of this instrument, the goals of this study were to investigate (a) longitudinal factorial validity and the internal consistency of the scale, (b) its equivalence over time, and (c) criterion validity through its relationship with study engagement over time. Data were gathered at two-time points, 3 months apart, from a sample of students in the Netherlands ($n = 360$). Longitudinal confirmatory factor analyses showed support for a two-factor model for overall strengths use, comprised of *Affinity for Strengths* and *Strengths Use Behaviors*. The SUS demonstrated high levels of internal consistency at both the lower- and upper bound limits at both time points. Further, strict longitudinal measurement invariance was established, which confirmed the instrument's temporal stability. Finally, criterion validity was established through relating strengths use to study engagement at different time stamps. These findings support the use of the SUS in practice to measure strengths use and to track the effectiveness of strengths use interventions within the higher education sector.

Keywords: Strengths Use Scale, strengths assessment, psychometric properties, longitudinal invariance, positive psychological assessment, psychological strengths

INTRODUCTION

University students are three times more likely to develop psychopathological complaints and common mental health problems than the general population (Blanco et al., 2008; Seligman, 2012). This stems from severe psychological distress experienced as a result of an imbalance between their study demands (e.g., workload/time pressure), their study resources (e.g., lecturer support), and personal resources (e.g., strengths use; Lesener et al., 2020). The problem is exacerbated by intensive educational programmes, poor social relationships with peers (Houghton et al., 2018; Basson and Rothmann, 2019), drastic life changes, elevated levels of social comparison, peer pressure, and an imbalance between their studies and home life (Bergin and Pakenham, 2015). This, in turn, negatively affects students' motivation, study engagement, learning potential, academic performance, and overall academic throughput (Ebert et al., 2018). Therefore, it is not surprising that universities are implementing interventions to help students either (a) find a balance between their study demands/resources or (b) develop the internal personal resources needed to offset university life's impact on their well-being and academic performance (Seligman, 2012).

An essential personal resource targeted by these interventions relates to identifying and using personal strengths during one's studies. Strengths refer to the inherent psychological traits that students are naturally good at, leading to optimal functioning or performance in desired outcomes (Govindji and Linley, 2007). These are naturally occurring capacities that are universally valued by society (Huber et al., 2017). When students can live out their strengths during their studies, it could lead to positive outcomes for the self and others. Research shows that strengths are associated with positive self-esteem, goal achievement, pro-social behaviors, happiness, and well-being (Littman-Ovadia et al., 2017). Further, when students can live out their strengths at university, it also reduces reported levels of stress, depression, and anxiety (Schutte and Malouff, 2018). When students use their strengths during their studies, they are also more likely to perform academically and less likely to fall out of or change academic programmes (Seligman, 2012).

However, despite these positive associations, intervention studies centered around strengths-based development have shown mixed results (White, 2016; Roll et al., 2019; White et al., 2019). Although some strengths-based interventions have led to mental health and well-being changes, others did not (Quinlan et al., 2012; White et al., 2019). Van Zyl et al. (2019) argued that this is primarily because of poor intervention design and -measurement, where the focus is on measuring outcomes rather than on the underlying mechanisms being targeted by the intervention. In other words, strengths interventions aim to develop strengths *use*; however, what is ultimately measured is strengths possession or strengths knowledge. In fact, several studies have shown that only knowing what one's strengths are (strengths knowledge) is not enough to facilitate sustainable changes in positive individual outcomes (Seligman et al., 2005; Wood et al., 2011; Seligman, 2012; Proyer et al., 2015a,b; Miglianico et al., 2020). Only when one can actively apply one's strengths (i.e., strengths use) would it lead to happier and healthier lives (Govindji and Linley, 2007). Therefore, strengths use has become a central tenet in recent strengths-based intervention studies.

To measure such, Govindji and Linley (2007) developed the Strengths Use Scale (SUS), a 14 item self-report scale that aims to measure active strengths use. The instrument aims to measure both opportunities to use strengths (affinity for strengths use), as well as individual strengths, use behaviors (strengths use behaviors) (Van Woerkom et al., 2016a). The SUS is the most widely used instrument to assess general strengths use and has been translated into German (Huber et al., 2017), French (Forest et al., 2012), Hebrew (Littman-Ovadia et al., 2017), Finish (Vuorinen et al., 2020), Chinese (Bu and Duan, 2020), and even adapted to work settings (Dubreuil et al., 2014). Despite its wide use, only four studies have actively attempted to investigate its validity and reliability: Govindji and Linley (2007) and Wood et al. (2011) in the US, Huber et al. (2017) in Germany, and Duan et al. (2018) in China. Although all four studies have shown that SUS was a reliable and valid tool, those outside of the U.S. required several modifications (e.g., correlating error terms or item parceling) to ensure data-model fit. This trend is also prevalent in several empirical studies where the SUS was used (e.g., Mahomed, 2019; Mahomed and Rothmann, 2020; Vuorinen et al., 2020). Any form of statistical modification of a psychometric instrument fundamentally changes the content of what is being measured, thus limiting comparisons between studies (Price, 2017). As such, a thorough investigation as to the psychometric properties of the SUS is needed.

Therefore, the purpose of this study was to investigate the psychometric properties, longitudinal invariance, and criterion validity of the SUS within a student population. Specifically, it aimed to determine the (a) longitudinal factorial validity and the internal consistency of the instrument (b) its temporal equivalence, and (c) its relationship with strengths use and study engagement over time.

LITERATURE REVIEW

Conceptualization and Measurement of Strengths Use

Positive psychology is rooted in the tenet that individuals have inherent psychological strengths, which are activated to manage hardships and promote optimal human functioning (Peterson and Seligman, 2004). Strengths develop out of adversity and are *essential* to one's definition of self, are *effortless* in their enactment and *energizing* when activated (Matsuguma and Niemiec, 2020). Therefore, psychological strengths can be seen as positive, trait-like capacities that define good character and highlight "what is right" about an individual (Richter et al., 2020). These ideas are in line with Linley and Harrington's (2006, p. 86) definition of strengths as "natural capacities for behaving, thinking or feeling in a way that allows optimal functioning and performance in the pursuit of valued outcomes." These capacities are universally valued by society as they lead to positive outcomes and benefits

for both the self (e.g., positive mental health) and others (e.g., positive community climate) (Huber et al., 2017).

Further, research suggests that strengths are also relatively stable over time (Snow, 2019), are valued across cultures and educational contexts (McGrath, 2015), buffer against the onset of psychopathology (Peterson et al., 2006), enhance mental health (Seligman, 2012; Proyer et al., 2015a), and lead to context-specific positive outcomes such as study engagement, and academic performance (Kwok and Fang, 2020). Further, despite being relatively stable over time, strengths remain malleable and can be developed through interventions to promote strengths awareness and active strengths use (Huber et al., 2017).

Govindji and Linley (2007) argued that merely possessing a strength is not an effective means to promote personal growth and development. Instead, individuals need to both become aware- and develop a deep understanding of their strengths (i.e., strengths awareness/knowledge) and exert conscious effort to apply such in different situations (Wood et al., 2011). Strengths *awareness/knowledge* refers to the ability to know the things one is naturally good at and understand what role strengths play in one's daily life (Wood et al., 2011). On the other hand, strengths *use* refers to the extent to which one is both driven to apply and opportunities to use one's strengths in different situations (Wood et al., 2011; Van Woerkom et al., 2016a). Govindji and Linley (2007) conceptualization of strengths use is built on the organismic value process (OVP). The OVP proposes that strengths are naturally occurring traits that develop from within, where individuals are inherently driven to actively use, develop, apply, and play to their strengths in daily life. Further, individuals yearn to live by their strengths and are unconsciously drawn to activities, hobbies, studies, or work aligned to their strengths (Wood et al., 2011). Therefore, individuals are naturally drawn to activities aligned to their strengths (i.e., strengths affinity) and exhibit active strengths use behaviors (Wood et al., 2011; Van Woerkom et al., 2016a).

Although strengths possession and awareness/knowledge are shown to be important within the educational environment, intervention studies have shown that its indeed the conscious use of strengths that leads to sustainable changes in mental health and well-being over time (Wood et al., 2011; Seligman, 2012; Van Zyl and Rothmann, 2019; Miglianico et al., 2020). Govindji and Linley (2007) found that active strengths use leads to higher levels of happiness, personal fulfillment, and subjective- and psychological well-being. In contrast, strengths possession/awareness were not independent predictors of happiness or well-being (Seligman et al., 2005; Govindji and Linley, 2007). Albeit strengths awareness/possession is a precursor to active strengths use (Seligman et al., 2005). Despite these findings, most academic research has focused on the awareness-, identification- or possession of strengths, rather than the actual use thereof (Wood et al., 2011; Huber et al., 2017).

This is further indicated by the vast array of propriety psychometric instruments used to identify or assess strengths (Richter et al., 2020). These include, but are not limited to, the Clifton Strengths-Finder (Rath, 2007), the VIA Signature Strengths Inventory for adults (Peterson and Seligman, 2004) and children (Ruch et al., 2014), the Signature Strengths Questionnaire-72 (Rashid et al., 2017), the Personal Strengths Inventory (Kienfie Liau et al., 2011), the Realise2 Strengths Finder (Linley et al., 2010), and the Employee Strengths At Work Scale (Bhatnagar, 2020). Each of these instruments aims to measure various forms of manifested strengths ranging from character strengths to inherent talents. In contrast, only two psychometric instruments are available that measures strengths use: the Strengths-Use and Deficit Correction Behavior Scale (SUDCO; Van Woerkom et al., 2016a) and the Strengths Use Scale (SUS: Govindji and Linley, 2007; Wood et al., 2011).

The SUDCO aims to measure (a) strengths use behaviors, (b) deficit correction behaviors and perceived organizational support for (c) strengths use, and (d) -deficit correction (Van Woerkom et al., 2016a). Although this instrument has shown to be a valid and reliable tool to measure strengths use, it was crafted to be used within organizational settings (Van Woerkom et al., 2016a). This implies that the SUDCO cannot measure strengths use in other contexts (e.g., educational settings) or assess general strengths use behaviors or opportunities. Given that the SUDCO also focuses on deficit correction, the tool is not in line with the tenets of positive psychology (i.e., moving away from a focus on "fixing what is wrong," but rather focus on developing what already works well; Seligman and Csikszentmihalyi, 2014). Further, the instrument is also not widely used within the literature (with only 71 citations on Google Scholar at the time of writing, i.e., early December 2020).

In contrast, the SUS is currently the most popular psychometric tool to measure strengths use behaviors and -opportunities within the literature with over a 1,000 citations (Govindji and Linley, 2007; Wood et al., 2011, p. 499). This 14 item self-report instrument aims to measure the extent to which individuals are drawn to activities that are aligned to their strengths and the extent to which strengths are actively used in a general way (Wood et al., 2011). The SUS has been translated into German (Huber et al., 2017), French (Forest et al., 2012), Hebrew (Littman-Ovadia et al., 2017), Finish (Vuorinen et al., 2020), Chinese (Bu and Duan, 2020), and adapted to work settings (Dubreuil et al., 2014). The instrument's popularity may be attributable to the fact that it was the first instrument developed to measure strengths use and that it is more inclined with the purest, functional principles of positive psychology.

Factorial Validity of the Strengths Use Scale

The Strengths Use Scale (SUS) was initially developed as a self-report measure to understand the extent to which individuals can apply their strengths in daily life (Govindji and Linley, 2007). The instrument was developed around the idea that "strengths are natural, they come from within, and we are urged to use them, develop them, and play to them by an inner, energizing desire. Further, when we use our strengths, we feel good about ourselves, we are better able to achieve things, and we are working toward fulfilling our potential" (Linley and Harrington, 2006, p. 41). From this conceptualization, strengths use has both an active application component

(strengths use behaviors) and encompasses opportunities to apply strengths to achieve personal goals or to facilitate personal development (opportunities to apply; Van Woerkom et al., 2016a).

Based on this conceptualization, Govindji and Linley (2007) generated 19 initial items, rated on a 7-point agreement type Likert scale, to measure strengths use from this perspective. Participants were instructed that these questions "*ask you about your strengths, that is, the things that you are able to do well or do best*" (Govindji and Linley, 2007, p. 147). A sample of 214 university students from the U.S. was requested to complete the SUS (Govindji and Linley, 2007). Principal component analysis revealed that three components with eigenvalues >1 could be extracted. However, the screen-plot showed that only a single component with 14 items could meaningfully be extracted from the data. These 14 items declared 56.2% of the total variance in a single "Strengths use" factor, with item loadings ranging from 0.52 to 0.79 (Govindji and Linley, 2007). The one-factor model showed to be significantly related to self-esteem, subjective well-being, psychological well-being, and subjective vitality, which established its concurrent validity (Govindji and Linley, 2007). However, this study only employed an exploratory approach, drawing a small sample from a single context. Therefore, the factorial validity could not formally be established nor verified. Despite showing promises, the authors argued that further validation studies on the SUS were needed.

In response, Wood et al. (2011) argued for the validation of the SUS within a general adult population ($N = 227$). This was done to increase the generalizability of the SUS within the U.S. Wood et al. (2011) employed both traditional factor analyses and parallel analyses to determine the factorial structure of the SUS. The results showed that a single strengths use factor could be extracted from the data based on eigenvalues. Items loaded between 0.66 and 0.87 on the single factor and declared 70.25% of the total variance.

Outside of the U.S., the SUS showed slightly different results. In the German validation, Huber et al. (2017) attempted to validate a translated version of the SUS within a sample of native German speakers. The authors employed both a traditional Exploratory Factor Analysis (EFA)- as well as a Confirmatory Factor Analysis (CFA) approach (through Structural Equation Modeling; SEM) to validate the instrument. The EFA showed that a single-factorial model, explaining 58.4% variance, with factor loadings ranging between 0.58 and 0.86, could be extracted from the data. The first factor had an eigenvalue of 8.60, with the remaining values clearly below the point of intersection (0.855–0.172). However, three items did not load sufficiently on the single strengths use factor (with factor loadings ranging from 0.336 to 0.410). The CFA was then conducted to determine if the hypothesized structure of the German SUS sample fitted the data well. However, the initial model fit of the German version was not satisfactory. Several modifications to the overall model needed to be implemented to enhance both model fit and measurement quality. This indicates that there may be conceptual overlap in understanding some items and that the factorial structure of the 14-items SUS may need further investigation.

Internal Consistency of the SUS

Another factor to consider when considering the SUS as a viable and reliable tool to measure strengths use is its level of internal consistency or "reliability." Reliability refers to the consistency and stability of an instrument to produce stable results (Wong and Wong, 2020). The SUS has shown to be a reliable measure across cultures; however, the level of internal consistency seems to vary within and between samples. In the original two U.S. validation studies, the SUS produced Cronbach's alpha coefficients ranging from 0.95 (Govindji and Linley, 2007) to 0.97 (Wood et al., 2011). Outside of the U.S., the SUS has shown acceptable levels of internal consistency in Germany ($\alpha = 0.84$: Huber et al., 2017), China ($\alpha = 0.94$: Bu and Duan, 2020), Finland ($\alpha = 0.88$: Vuorinen et al., 2020), and the U.K. ($\alpha = 0.90$: McTiernan et al., 2020).

Further, the test-retest reliability of the SUS was tested through intra-class correlations spanning three-time points (3 and 6 months after the first measurement). The test statistic was significant and very high ($r_{icc} = 0.85$), indicating that the SUS scores remained sufficiently stable without any specific intervention. Conversely, after a positive psychology intervention, strengths use scores have been shown to increase (e.g., Dubreuil et al., 2016), indicating that the scale is sensitive to measure changes.

However, despite the criticisms around Cronbach's alpha, only one other study employed a more restrictive and robust metric for internal consistency. Mahomed and Rothmann (2020) found that the composite reliability (i.e., upper bound level of internal consistency) of the SUS was 0.92. No other study specifically attempted to determine the upper level of internal consistency of the SUS.

Stability of the SUS Over Time: Longitudinal Measurement Invariance

The temporal stability of the SUS is another essential metric to consider. This can be assessed through longitudinal measurement invariance (LMI). LMI is concerned with testing the factorial equivalence or equality of a construct over time (rather than across groups; Wong and Wong, 2020). Specifically, LMI assesses if the SUS produces similar factorial structures (configural invariance), if items load similarly on their respective factors (metric invariance), if the SUS shows to have similar intercepts (scalar invariance), and if similar residual errors are produced over time (Wong and Wong, 2020). LMI is a desirable characteristic of a measurement instrument as it provides evidence that a construct can be both measured and interpreted the same across different time stamps; therefore making meaningful interpretations and comparisons of mean scores of strengths use over time possible (Cheung and Rensvold, 2002; Widaman et al., 2010). No study has attempted to assess the LMI of the SUS over time, and therefore no specific reference points for such can be established from the current literature.

However, both Peterson and Seligman (2004) and Govindji and Linley (2007) argued that strengths are considered trait-like factors that are relatively stable over time. Further, the extent to which one would apply or use one's strengths is also considered

stable over time, unless individuals are exposed to- or engage in strengths-based developmental initiatives (Seligman, 2012; Huber et al., 2017). Therefore, it is expected that strengths-use, without intervention, should stay relatively stable over time.

Criterion Validity: Strengths Use and Study Engagement

A final metric to consider when validating an instrument is criterion validity. Criterion validity can be measured through establishing relationships with theoretically closely related variables (concurrent validity), and through the ability to predict outcomes on these related variables over time (predictive validity; Van Zyl, 2014). An important criterion to consider associated with active strengths use is study engagement (Ouweneel et al., 2011; Seligman, 2012; Stander et al., 2015; Kwok and Fang, 2020). Study engagement is a persistent and pervasive positive, fulfilling, and study-related state of mind characterized by feelings of vigor, showing dedication to one's studies and being absorbed in one's study-related tasks (Schaufeli et al., 2002). Drawing from desire theory, Seligman (2012) argued that when students can live in accordance with their strengths (i.e., engage in learning activities congruent with their strengths), or if they engage in study-related activities that are aligned to their strengths, that they will experience more engagement in their studies. The broaden-and-build theory of positive emotions further postulates that strengths are essential personal resources individuals can activate to translate positive emotional experiences into study-related engagement (Fredrickson, 2001). Several studies have also specifically shown that higher levels of active strengths-use lead to increased study and work-related engagement (Ouweneel et al., 2011; Seligman, 2012; Stander et al., 2015; Kwok and Fang, 2020). As such, both concurrent validity and predictive validity could be established by associating SUS with study engagement at different points in time.

The Current Study

Given the importance of strengths use, and the popularity of the SUS within the literature, it is imperative to ensure that it is a valid and reliable instrument. As such, the purpose of this study was to investigate the psychometric properties, longitudinal invariance, and criterion validity of the Strengths Use Scale (SUS) within a student population. Specifically, the aim was to determine the: (a) longitudinal factorial validity and the internal consistency of the instrument, (b) its equivalence over time, and (c) criterion validity through its relationship with study engagement over time.

RESEARCH METHODS

Research Approach

A quantitative, electronic survey-based longitudinal design was employed to determine the psychometric properties, longitudinal invariance and criterion validity of the SUS. This design entailed the distribution of questionnaires at two-time points over 3 months.

TABLE 1 | Demographic characteristics of participants ($n = 360$).

Item	Category	Frequency (*f*)	Percentage (%)
Gender	Male	266	73.9
	Female	93	25.8
	Other	1	0.3
Age (years)	20–22 years	284	78.9
	23–25 years	67	18.6
	23–30 years	8	2.2
Nationality	Dutch	352	97.8
	Other	8	2.2
Level of education	Bachelor's degree	219	60.8
	Master's degree	141	39.2

Participants and Sampling Strategy

An availability-based sampling strategy was employed to draw 360 respondents from a University in The Netherlands to participate in this study. **Table 1** provides an overview of the demographic characteristics of the sample. Validity responses were established by implementing two attention check items. If participants failed to score on these items, they were excluded from the analysis. As presented in **Table 1**, the majority of the participants were Dutch (97.8%) males (73.9%) between the ages of 20 and 22 years (78.9%) with a Bachelor's Degree (60.8%).

Research Procedure

The data obtained for this paper are drawn from two large-scale cross-cultural student well-being projects. The Dutch sample consisted of two different datasets: one contained only third-year students and the other only master students. Data collection occurred during 2019-2020. The first cohort of data was collected between February to May 2019 and the second from November 2019 to January 2020 (before the COVID-19 outbreak). The period between measurements was 3 months. Online surveys were distributed at Time 1 and repeated at Time 2. A unique code was assigned to individuals to match Time 1 and Time 2 responses. Links were sent out to participants to their institutional email via Qualtrics™ (www.qualtrics.com). In each survey, the rights and responsibilities of the participants were discussed. Participants provided online written informed consent. They were informed that their anonymity would be guaranteed and that their data would be stored in password-secured systems. Participants were informed they could withdraw their participation in this study at any time, without any repercussion for them. The purpose of the study was explained alongside the risks and benefits of the study. Participants' questions were answered at any step of the study.

Measuring Instruments

The study made use of the three psychometric instruments.

A *demographic questionnaire* was used to gather basic biographic and demographic information about the participants. It aimed to capture respondents' self-identified gender identity, current age, nationality, home language, and level of education.

The *Strengths Use Scale* (SUS)[1] developed by Govindji and Linley (2007) to measure how students actively used their strengths. The 14-item self-report questionnaire measured strengths use on an agreement-type Likert scale ranging from 1 (Strongly disagree) to 7 (Strongly agree) with items such as "I achieve what I want by using my strengths" and "Most of my time is spent doing things that I am good at doing." The SUS showed acceptable levels of internal consistency at the lower bound limit with a Cronbach's alphas of 0.95 (Govindji and Linley, 2007).

The *Utrecht Work Engagement Scale* for students (UWES-9S) developed by Schaufeli et al. (2006) was used to measure study engagement. The 9-item questionnaire is rated on a six-point agreement type Likert scale ranging from 1 (Never) to 7 (Always). It measures the three components of study engagement with three items each. Example items are "When I am doing my work as a student, I feel bursting with energy" (vigor), "I am proud of my studies" (dedication), and "I get carried away when I am studying" (absorption). The UWES-9S has shown to be a valid and reliable measure in various contexts, with Cronbach Alpha's ranging from 0.72 to 0.93 (Schaufeli et al., 2006; Cadime et al., 2016).

Statistical Analyses

Data were analyzed using SPSS v26 (IBM SPSS, 2019) and Mplus v 8.4 (Muthén and Muthén, 2020). A six-phased longitudinal factor analytical strategy through structural equation modeling was employed to investigate the psychometric properties, temporal stability, and concurrent/predictive validity of the SUS over time.

First, to explore the factorial structure of the SUS, an exploratory factor analytical (EFA) strategy was employed on the baseline data. To determine factorability, the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's sphericity test was used. A KMO value larger than 0.60 and a statistically significant chi-square value on Barlett's test of sphericity would indicate that the data were factorable (Kaiser and Rice, 1974). Thereafter, an EFA was conducted through the structural equation modeling approach with the maximum likelihood estimation method and a Geomin (Oblique) rotation. Competing EFA factorial models were specified to be extracted based on Eigenvalues larger than 1 (Muthén and Muthén, 2020). Model fit statistics (c.f. **Table 2**) were used to establish data-model fit and to compare the competing EFA models. Further, items were required to load statistically significantly (Factor loading >0.40; $p < 0.01$) on their respective extracted factors and needed to declare at least 50% of the overall variance.

Second, a competing confirmatory factor analytical (CFA) measurement modeling strategy with the maximum likelihood estimation method (ML) was employed. As a baseline measure, three competing measurement models were specified and sequentially compared for each of the two-time points, separately. This approach verifies the best factorial structure and measurement quality of the instrument at each time point before evaluating temporal stability (Feldt et al., 2000). These separate and competing models were specified according to the traditional independent cluster model confirmatory factor analytical conventions where items were estimated to load onto their a priori theoretical factors and cross-loadings were constrained to zero (Wong and Wong, 2020)[2].

To determine the best fitting measurement model at each time point and to mitigate the criticism of Hu and Bentler (1999) method of establishing model fit by *solely* looking at series of "cut-off points" and "standardized values" of fit indices, a sequential process of evaluation was implemented. As an initial step, the Hu and Bentler (1999) model fit criteria (c.f. **Table 2**) was used to determine data-model fit and to discriminate between measurement models for each time point. Thereafter, measurement quality was assessed through inspecting the standardized item loadings ($\lambda > 0.40$; $p < 0.01$), standard errors, item uniqueness (range between 0.1 and 0.9; $p < 0.01$), and the presence of multiple cross-loadings to further discriminate between models (Asparouhov and Muthén, 2009; Kline, 2011). Only models that showed both excellent fit *and* -measurement quality (with no items significantly loading on multiple factors) were retained for further analyses (McNeish and Hancock, 2018; McNeish et al., 2018; Shi et al., 2019).

Third, a longitudinal CFA (L-CFA) strategy was used to determine the temporal stability of the SUS's factorial structure. Here, the three measurement models from Time 1 were regressed on their corresponding counterparts in Time 2 (Von Eye, 1990). Again, these competing longitudinal measurement models were assessed for model fit/measurement quality and then systematically compared based on the same criteria as in the previous phase. As a first step in establishing temporal stability of the factorial models, two criteria needed to be met: (a) the regressive path between the factorial models of Time 1 and Time 2 were required to be large (Standardized $\beta > 0.50$) and statistically significant ($p < 0.01$) and (b) factorial models at Time 1 needed to declare at least 50% of the variance in its corresponding counterpart at Time 2 (Von Eye, 1990). The model that fit all the criteria was then retained for a more detailed item level inspection and further analyses.

Fourth, based on the best fitting L-CFA model, *item-level descriptive statistics, standardized factor loadings,* and *internal consistency* were investigated. Item related descriptive statistics were computed to provide a descriptive overview of each item in terms of means and standard deviations, inspect the corrected item-total correlations (CITC), and determine absolute normality (Skewness and Kurtosis). Based on Kim (2013)

[1]Following the guidelines from the International Test Commission regarding the use and adaption of tests across cultures (Muñiz et al., 2013), before administration, the 14 items were piloted in a small group of master students to verify their clarity ($n = 5$). Based on feedback from the group, one item of the original instrument (STU_3 "I play to my strengths") needed to be rephrased ("I pursue goals and activities that are aligned to my strengths") in order to improve its comprehension within the Dutch context.

[2]The Strengths Use Scale was found to be comprised of two factors, Affinity for Strengths and Strengths Use Behaviors, and it was important and necessary for us to test a second-order factorial model to assess the hierarchical nature of the SUS. After first finding support for the first-order factorial model of the SUS items loading on the two first-order latent factors (Affinity for Strengths, Strengths Use Behaviors), we found support for a second-order factorial model in which the two first-order factors load onto the second-order factor, Overall Strengths Use. This therefore suggests that the SUS can be useful for measuring overall strengths use.

TABLE 2 | Model fit statistics.

Fit indices	Cut-off criterion	Sensitive to *N*	Penalty for model complexity
Absolute fit indices			
Chi-square (χ^2)	Lowest comparative value between measurement models Significant ($p > 0.01$)	Yes	Yes
χ^2/df	<3 = Excellent and <5 = Acceptable	No	No
Approximate fit indices			
Root-Means-Square Error of Approximation (RMSEA)	<0.08 but >0.01 90% CI Range doesn't include zero	Yes	Yes
Standardized Root Mean Square Residual (SRMR)	<0.08 but > 0.01	Yes	No
Incremental fit indices			
Comparative Fit Index (CFI)	>0.90 but <0.99	No	Yes
Tucker-Lewis Index (TLI)	>0.90 but <0.99	No	Yes
Akaike Information Criterion (AIC)	Lowest value in comparative measurement models	No	No
Bayes Information Criterion (BIC)	Lowest value in comparative measurement models	No	No

Adapted from Kline (2011) and Wong and Wong (2020).

suggestion, absolute values for Skewness (<2) and Kurtosis (<2) were used as indicators of normality as our sample size was smaller than 500. The CITC represents the relationship of each item to the overall SUS, where correlations of less than $r = 0.30$ indicate that an item may not represent the overall factor (Zijlmans et al., 2019). Subsequently, point-estimate reliability (upper-bound; $\rho > 0.80$; Raykov, 2009), Mc Donald's Omega ($\omega > 0.80$; Hayes and Coutts, 2020) and Cronbach's alpha (lower-bound; $\alpha > 0.70$; Nunnally and Bernstein, 1994) for the best fitting model were computed to determine the internal consistency of the SUS and its subscales. Further, the average variance extracted (AVE) acts as an indicator for the average reliability of each individual indicator (item) in a scale, where a value over 50% is acceptable (Kline, 2011).

Fifth, second-order *longitudinal measurement invariance* (LMI) was implemented to determine whether the SUS is measured similarly at Time 1 and Time 2. LMI was assessed through applying increasingly restrictive equality constraints on the best fitting (second-order) L-CFA through estimating:

1. configural invariance (similar factor structures at baseline).
2. metric invariance for the first-order factorial model (similar factor loadings over time).
3. metric invariance for the second-order factorial model.
4. scalar invariance for the first-order factorial model (similar intercepts over time).
5. scalar invariance for the second-order factorial model.
6. strict invariance for the overall model (similar residual errors over time).

Invariance was established by comparing these ever-restrictive models on predefined criteria (Chen, 2007). A chi-square difference test was first computed but not used due to its sensitivity to minor parameter changes in small samples and model complexity (Cheung and Rensvold, 2002; Chen, 2007; Widaman et al., 2010). Instead, changes in RMSEA ($\Delta < 0.015$), SRMR ($\Delta < 0.015$), CFI (< 0.01), TLI (< 0.01), and chi-square/df (<1) indicated invariance (Cheung and Rensvold, 2002; Widaman et al., 2010). For comparisons, the least restrictive model was compared to the increasingly constrained models in each sequential step of the estimation process. If invariance was established, latent mean differences between the time points could be computed. Here, the Time 1 mean score was constrained to zero and used as the reference group. Time 2 mean score was freely estimated. Should Time 2 latent mean score differs significantly from zero, it would indicate a significant difference between timestamps (Wickrama et al., 2016; Wong and Wong, 2020).

Finally, to establish *concurrent and predictive validity,* separate structural models were estimated with the best fitting L-CFA model as an exogenous factor and study engagement as the endogenous factor. For concurrent validity, Strengths Use at Time 1 was regressed on Study Engagement Time 1 and Strengths Use Time 2 regressed on Study Engagement Time 2. To establish predictive validity, Strengths Use Time 1 was regressed on Study Engagement Time 2. A significance level of $p < 0.01$ (99% confidence interval) for each regressive path.

RESULTS

The results of the exploratory factor analyses, baseline competing measurement models, longitudinal factor analyses, item-level descriptive (and internal consistency), longitudinal measurement invariance, and concurrent/predictive validity are reported separately in this section. The results are presented in a tabulated format with brief subsequent interpretations.

Exploratory Factor Analysis

To explore the factorial structure of the SUS an EFA approach was employed on the baseline data. First, factorizability was established through the KMO measure and Bartlett's test for sphericity. The results showed that the KMO value was larger than 0.60 (KMO = 0.94) and produced a significant chi-square

TABLE 3 | EFA: Geomin rotated factor loadings and declared variance.

	Item	λ_1	λ_2	R^2 (%)
Affinity				7.48
STU_1	I am regularly able to do what I do best	0.80	−0.01	
STU_2	I pursue goals and activities that are aligned to my strengths	0.83	−0.01	
STU_3	[Redacted]	0.48	0.21	
STU_4	[Redacted]	0.53	0.25	
STU_7	[Redacted]	0.58	0.07	
STU_12	Most of my time is spent doing the things that I am good at doing	0.41	0.26	
Active use				63.94
STU_5	I use my strengths everyday	0.19	0.59	
STU_6	I use my strengths to get what I want out of life	0.27	0.53	
STU_8	[Redacted]	0.21	0.54	
STU_9	[Redacted]	−0.09	0.81	
STU_10	I find it easy to use my strengths in the things I do	−0.02	0.81	
STU_11	[Redacted]	0.05	0.78	
STU_13	Using my strengths is something I am familiar with	0.13	0.65	
STU_14	[Redacted]	−0.00	0.81	

($p < 0.01$). Meaningful factors could therefore be extracted, and we proceeded to estimate the EFA models.

As an initial measure, one to five factorial models were specified to be extracted. The results showed that two factors could be extracted with eigenvalues larger than 1. Further, only two models converged: A *single first order factorial model* ($\chi^2_{(360)} = 391.48$; $df = 77$; $\chi^2/df = 5.08$; CFI = 0.89; TLI = 0.87; RMSEA = 0.11 [0.097,0.118]; SRMR = 0.05; AIC= 12588.99; BIC = 12751.73; Eigenvalue: 7.55; $R^2 = 53.94\%$) and a *two first order factorial model* ($\chi^2_{(360)} = 228.96$; $df = 64$; $\chi^2/df = 3.58$; CFI = 0.94; TLI = 0.92; RMSEA = 0.08 [.073,0.097]; SRMR = 0.03; AIC = 12452.47; BIC = 12665.59; Eigenvalue Factor 1 = 7.55; $R^2 = 53.94\%$; Eigenvalue Factor 2 = 1.05; $R^2 = 7.48\%$). Only the two first-order factorial model fitted the data. This model showed significantly better fit than the single first-order factorial model. The item loadings and declared variance for this model are presented in **Table 3**[3]. All items loaded larger than 0.40 onto their respective factors. The first factor was labeled *Affinity for Strengths* ("Affinity") and the second factor as *Strengths Use Behaviors* ("Active Use"). The Geomin factorial correlation showed that Affinity and Active Use were strongly correlated ($r = 0.73$; $p < 0.01$).

Cross-Sectional Factorial Validity: Competing Measurement Models for Time 1 and Time 2

A competing measurement modeling strategy was employed to establish the factorial validity of the SUS on each of the "cross-sectional" data points. Here, observed items were used as indicators of latent factors. No items were removed and error terms were permitted to correlate.

[3]For copyright purposes, several items were redacted. Items are, however, numbered and presented in the same order as in Govindji and Linley (2007) and Wood et al. (2011).

The following models were estimated separately at both Time 1 and Time 2:

- **Model 1 & Model 4**: A one-factor first-order factorial model was estimated where all 14 items loaded directly on to a single factor called "Overall Strengths Use."
- **Model 2 & Model 5**: Two correlated first-order factor models were estimated for a factor labeled "Strengths Affinity" (comprised of items 1, 2, 3, 4, 7, 12) and "Active Use" (comprised out of items 5, 6, 8, 9, 10, 11, 13, 14).
- **Model 3 & 6**: A second-order factorial model comprised out of the two first-order factors specified in the previous model was specified to directly load onto overall Strengths Use.

Table 4 presents the model fit indices for each of the estimated models. At ***Time 1***, the results showed that only Models 2 and 3 fitted the data ($\chi^2_{(360)} = 267.48$; $df = 76$; $\chi^2/df = 3.52$; CFI = 0.93; TLI = 0.92; RMSEA = 0.08 [0.073,0.095]; SRMR = 0.04). Both models further fitted the data significantly better than Model 1 ($\Delta\chi^2 = -124.00$; $\Delta df = -1$; $\chi^2/df = -1.56$; ΔCFI = 0.04; ΔTLI = 0.05; ΔRMSEA = −0.03; ΔSRMR = −0.01; ΔAIC: −122.00; ΔBIC: −118.13).

The result showed a similar pattern at ***Time 2***, only Models 5 and 6 fitted the data ($\chi^2_{(360)} = 328.40$; $df = 76$; $\chi^2/df = 4.32$; CFI = 0.92; TLI = 0.91; RMSEA = 0.10 [0.087,0.108]; SRMR = 0.04). Both models fitted the data significantly better than Model 4 ($\Delta\chi^2 = -34.10$; $\Delta df = -1$; $\chi^2/df = -0.56$; ΔCFI = 0.03; ΔTLI = 0.02; ΔRMSEA = −0.00; ΔSRMR = 0.00; ΔAIC: −31.41; ΔBIC: −28.23).

In respect of measurement quality, all models at both Time 1 and Time 2 showed acceptable levels with standardized factor loadings ($\lambda > 0.40$; $p < 0.01$), standard errors, and item uniqueness ($\delta < 0.10$ but > 0.90; $p < 0.01$) meeting the classification criteria (Asparouhov and Muthén, 2009; Kline, 2011).

TABLE 4 | Cross-sectional confirmatory factor analyses: measurement model fit statistics for time 1 and time 2.

Model	Description	χ^2	*df*	χ^2 / *df*	CFI	TLI	RMSEA	SRMR	AIC	BIC	aBIC	Meets criteria
Time 1												
Model 1	One-factor model	391.48	77	5.08	0.89	0.87	0.11 [0.097, 0.118]	0.05	12588.99	12751.74	12618.50	No
Model 2	Two-factor model	267.48	76	3.52	0.93	0.92	0.08 [0.073, 0.095]	0.04	12466.99	12633.61	12497.20	Yes
Model 3	Second-order factor model	267.48	76	3.52	0.93	0.92	0.08 [0.073, 0.095]	0.04	12466.99	12633.61	12497.20	Yes
Time 2												
Model 4	One-factor model	362.50	77	4.71	0.91	0.89	0.10 [0.092, 0.114]	0.04	11526.02	11688.17	11554.93	Partially
Model 5	Two-factor model	328.41	76	4.32	0.92	0.91	0.10 [0.087, 0.108]	0.04	11493.92	11659.94	11523.52	Yes
Model 6	Second-order factor model	328.41	76	4.32	0.92	0.91	0.10 [0.087, 0.108]	0.04	11493.92	11659.94	11523.52	Yes

χ^2, Chi-square; df, degrees of freedom; TLI, Tucker-Lewis Index; CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation [90%CI]; SRMR, Standardised Root Mean Square Residual; AIC, Akaike Information Criterion; BIC, Bayes Information Criterion; aBIC, Adjusted Bayes Information Criterion.

Longitudinal Factor Analyses: Longitudinal Factorial Validity and Temporal Stability

The next step in the process was to determine the stability of the SUS over time using L-CFA. In each L-CFA model, the corresponding measurement model specified in Time 1 was regressed on Time 2. The following models were tested:

- **Model 7:** The single, first-order factor (with all 14 items loading directly on to such) of Time 1 was regressed on the single first-order factor of Time 2.
- **Model 8:** The two first-order factor models of "Strengths Affinity" (comprised of items 1, 2, 3, 4, 7, 12) and "Active Use" (comprised out of items 5, 6, 8, 9, 10, 11, 13, 14) at Time 1 was regressed onto their corresponding Strengths Affinity and Proactive Use factorial counterparts. Covariances between the factors at each time point was permitted.
- **Model 9:** The second-order factorial model of Time 1 was regressed on that of Time 2 was regressed. Both models comprised out of the two first-order factors specified in the previous model. Covariances between the factors at each time point was not permitted. Error terms on Item 14 and 11 were permitted to covary at Time 2.

The results summarized in **Table 5** indicated that only Model 9, the second-order longitudinal factorial model, fitted the data ($\chi^2_{(360)} = 974.93$; $df = 344$; $\chi^2/df = 2.83$; CFI = 0.90; TLI = 0.90; RMSEA = 0.07 [0.066,0.077]; SRMR = 0.04). Model 9 also fitted the data significantly better than Model 7 ($\Delta\chi^2 = -224.31$; $\Delta df = -5$; $\chi^2/df = -0.60$; ΔCFI = 0.03; ΔTLI = 0.04; ΔRMSEA = −0.01; ΔSRMR = 0.00; ΔAIC: −214.20; ΔBIC: −194.77) and Model 8 ($\Delta\chi^2 = -53.58$; $\Delta df = -2$; $\chi^2/df = -0.14$; ΔCFI = 0.00; ΔTLI = 0.01; ΔRMSEA = 0.00; ΔSRMR = 0.00; ΔAIC: −49.58; ΔBIC: −41.81). All longitudinal models showed acceptable levels of measurement quality with standardized factor loadings ($\lambda > 0.40$; $p < 0.01$), standard errors, and item uniqueness ($\delta > 0.10$ but < 0.9; $p < 0.01$) exceeding the specified thresholds (Asparouhov and Muthén, 2009; Kline, 2011).

Further, to assess the final two assumptions for L-CFA, the regressive paths and covariances, as well as the variance declared by factorial models of Time 1 in Time 2, were estimated and summarized in **Table 6**. Although all the factors at Time 1 statistically significantly predicted the factors in Time 2, the results showed that only Model 9 met both the significance and variance criteria. The second-order factorial Strengths Use factor at Time 1 statistically significantly predicted 51% of the variance of Strengths Use in Time 2 with a large effect ($\beta = 0.71$; S.E = 0.03; $p < 0.01$). Therefore, only Model 9 was retained for further analyses.

Longitudinal Factor Loadings, Item Level Descriptive and Internal Consistency

Next, item-level descriptive statistics (means, standard deviations, skewness, kurtosis, CICT), standardized factor loadings, the Average Value Explained (AVE) and the level of internal consistency was computed for the second-order longitudinal factor

TABLE 5 | Longitudinal confirmatory factor analyses: measurement model fit statistics for time 1 and time 2.

Model	Description	χ^2	*df*	χ^2 */df*	CFI	TLI	RMSEA	SRMR	AIC	BIC	aBIC	Meets criteria
Model 7	One-factor models	1199.14	349	3.44	0.87	0.86	0.08 [0.077, 0.0.87]	0.05	23941.81	24272.13	24002.47	No
Model 8	Two-factor models	1028.51	346	2.97	0.90	0.89	0.07 [0.069, 0.0.79]	0.05	23777.19	24119.17	23839.98	Partially
Model 9	Second-order factor models	974.93	344	2.83	0.90	0.90	0.07 [0.066, 0.077]	0.05	23727.61	24077.36	23791.83	Yes

χ^2, Chi-square; df, degrees of freedom; TLI, Tucker-Lewis Index; CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation [90%CI]; SRMR, Standardised Root Mean Square Residual; AIC, Akaike Information Criterion; BIC, Bayes Information Criterion; aBIC, Adjusted Bayes Information Criterion.

model (Model 9). **Table 7** provided a summary for the results.

The results of the item level descriptive statistics show that all items were normally distributed (Skewness and Kurtosis $< +2; -2$: Kim, 2013), that each item was clearly associated with the overall factor being assessed (CITC $r > 0.30$: Zijlmans et al., 2019) and that each sub-factor and overall strengths-use scale showed to be reliable at both the upper- ($\rho > 0.80$; $\omega > 0.80$) and lower-bound level of internal consistency ($\alpha > 0.70$) at both time points.

All items on Affinity and Active Use loaded statistically significantly on their respective factors at both time points with standardized factor loadings ranging from 0.56 to 0.81 ($p < 0.01$). The AVE for Affinity was acceptable, with 0.50 reported at Time 1 and 0.54 at Time 2. Similarly, the AVE for Active Use at Time 1 (AVE = 0.58) and Time 2 (AVE = 0.58) exceeded the 0.50 threshold.

Further, both the first-order Affinity (Time 1: $\lambda = 0.90$, SE = 0.03 $p < 0.01$; Time 2 = $\lambda = 0.98$ SE = 0.02; $p < 0.01$) and Active Use factors (Time 1: $\lambda = 0.94$, SE = 0.03, $p < 0.01$; Time 2: $\lambda = 0.97$, SE = 0.02, $p < 0.01$) loaded statistically significantly onto the second-order Strengths Use factor. The second-order longitudinal factorial model therefore showed to have an excellent level of measurement quality and can therefore be subjected to more robust assessments of longitudinal stability over time.

Longitudinal Measurement Invariance and Mean Comparisons

Next, longitudinal measurement invariance (LMI) was tested to determine the factorial equivalence of the SUS over time. The results, summarized in **Table 8**, showed that all invariance models fitted the data based on the criteria mentioned in **Table 2** and that longitudinal measurement invariance of the SUS could be established between the different time points. No significant differences in terms of RMSEA ($\Delta < 0.015$), SRMR ($\Delta < 0.015$), CFI (< 0.010), TLI (< 0.010), and χ^2/df (<1) between the configural, metric, scalar, and strict invariance models were found (Cheung and Rensvold, 2002; Widaman et al., 2010; Wong and Wong, 2020). Therefore, the SUS showed to be a consistent measure over time and that meaningful mean comparisons between Time 1 and Time 2 can be made.

Further, to compare latent means on the first- and second-order factors of the SUS, all mean scores at Time 1 were constrained to zero within the strict invariance model. Affinity, Active Use, and Overall Strengths Use at Time 2 were then freely estimated. For the first-order factors, the results showed that Affinity ($\Delta\ \bar{x} = -0.7$; SE = 0.04; $p = 0.10$) and Active Strengths Use ($\Delta\ \bar{x} = 0.7$; SE = 0.05; $p = 0.11$) at Time 2 did not meaningfully differ from Time 1. Similarly, at a second-order factorial level, Overall Strengths Use at Time 2 ($\Delta\ \bar{x} = 0.0$; SE = 0.04; $p = 0.908$) did also not meaningfully differ from Time 1.

Concurrent and Predictive Validity

To establish concurrent and predictive validity, separate structural models were estimated with the second-order Strengths Use models specified as an exogenous factor and Study Engagement (as a second-order factor made up of three

TABLE 6 | Longitudinal confirmatory factor analyses: regressive paths and covariances between time 1 and time 2.

Model	Relational path			Standardized						Meets criteria
				β	*r*	S.E	*t*-value	*p*	R^2	
Model 7	Strengths use time 1	➔	Strengths Use Time 2	0.68	-	0.03	20.63	0.00	0.46	Partially
Model 8	Affinity time 1	➔	Affinity Time 2	0.66	-	0.04	17.22	0.00	0.44	Partially
	Active use time 1	➔	Active Use Time 2	0.63	-	0.04	16.71	0.00	0.40	
	Affinity time 1	⟷	Active Use Time 1	-	0.88	0.02	46.61	0.00	-	
	Affinity time 2	⟷	Active Use Time 2	-	0.95	0.02	38.77	0.00	-	
Model 9	Strengths use time 1	➔	Strengths Use Time 2	0.71	-	0.03	20.84	0.00	0.51	Yes

➔*Regression,* ⟷*Covariance.*

first-order factors: Vigor, Dedication, and Absorption) specified as endogenous factors. The results for both concurrent and predictive validity are summarized in **Table 9**.

For *concurrent validity*, Strengths Use at Time 1 was first was regressed on Study Engagement at Time 1. The model showed adequate fit ($\chi^2_{(360)} = 595.83$; $df = 225$; $\chi^2/df = 2.65$; CFI = 0.92; TLI = 0.91; RMSEA = 0.06 [0.061,0.075]; SRMR = 0.06; AIC = 20707.47; BIC = 20994.22). Strengths Use at Time 1 was directly associated with Study Engagement at Time 1 ($\beta = 0.49$; S.E = 0.05; $p < 0.01$; $R^2 = 0.24$). Similarly, Strengths Use at Time 2 was also directly associated with Study Engagement at Time 2 ($\beta = 0.58$; S.E = 0.04; $p < 0.01$; $R^2 = 0.33$). This model also showed adequate fit ($\chi^2_{(360)} = 689.80$; $df = 225$; $\chi^2/df = 3.07$; CFI = 0.91; TLI = 0.90; RMSEA = 0.08 [0.070,0.083]; SRMR = 0.06; AIC = 19403.58; BIC = 19689.28).

For *predictive validity*, Strengths Use at Time 1 was regressed on Study Engagement at Time 2. This model showed adequate fit ($\chi^2_{(360)} = 576.37$; $df = 225$; $\chi^2/df = 3.07$; CFI = 0.93; TLI = 0.91; RMSEA = 0.07 [0.069,0.073]; SRMR = 0.06; AIC = 20423.66; BIC = 20711.24). Here, Strengths Use at Time 1 predicted 22% of the variance in Study Engagement at Time 2 ($\beta = 0.47$; S.E = 0.05; $p < 0.01$; R^2: 0.24). Both concurrent and predictive validity of the SUS could therefore be established.

DISCUSSION

This study aimed to investigate the psychometric properties, longitudinal invariance, and criterion validity of the SUS within a Dutch student population. Longitudinal confirmatory factor analysis showed that a second-order factorial model, comprised of two first-order factors (*Affinity for Strengths* and *Strengths Use Behaviors*), fitted the data best. Further, this model showed support for strict longitudinal measurement invariance over 3 months with similar factorial structures, -factor loadings, item intercepts, and item uniqueness. Further, the SUS produced high levels of internal consistency at both the lower- and upper bound limits at both time stamps. Mean comparisons showed that neither overall strengths use, nor its two components, differed between Time 1 and Time 2. This confirmed the stability of the SUS over time. Finally, strengths use was related to study engagement at both time points. Strengths use at Time 1 also predicted study engagement at Time 2. Therefore, supporting the assumptions of criterion validity.

The Psychometric Properties of the Strengths Use Scale

Longitudinal factor analyses showed that a second-order factorial model of overall strengths use, comprising two first-order factors called *Affinity for Strengths* and *Strengths Use Behaviors*, fitted the data. *Affinity for Strengths* comprised six items related to opportunities where individuals could live out or apply their strengths. These opportunities related to activities that individuals are drawn to and that are naturally aligned to their strengths (Wood et al., 2011; Van Woerkom et al., 2016a). Individuals seek out activities where they can both live out- and pursue goals aligned to their strengths. They further show a natural affinity for mastering new skills/hobbies where these strengths are required (Govindji and Linley, 2007).

On the other hand, active *Strengths Use Behaviors* was measured by eight items related to the behaviors' individuals exhibit when applying strengths in everyday life. These behaviors related to actions employed by individuals to actively develop and apply their strengths to achieve life goals. Here, individuals can actively deploy their strengths to get what they want out of life (Govindji and Linley, 2007).

This two-factorial permutation of the SUS contrasts with Govindji and Linley (2007) and Wood et al. (2011), who reported strengths use as a single, first-order factor. Although our findings contrast with these authors' empirical results, it is in line with the original theoretical tenet on which the instrument was built. Govindji and Linley (2007) argued that strengths use is a function of the organismic value process and the self-concordant goal theory (from which items of the SUS was generated). According to Joseph and Linley (2005), the organismic value process suggests that strengths are psychological traits that individuals are inherently driven to use, develop, and apply (i.e., behaviors). Further, individuals express an inherent desire to live by their strengths and are unconsciously attracted to and show an affinity for activities/hobbies, studies, or work that are aligned to their strengths (i.e., affinity) (Wood et al., 2011; Huber et al., 2017). Therefore, our results are more closely aligned to the original theoretical ideas underpinning strengths use as proposed by Govindji and Linley (2007), rather than their empirical results.

TABLE 7 | Item level descriptive statistics, standardized factor loadings, average value explained, and internal consistency for the longitudinal model 9.

Factor	Item	Time 1–Second-order strengths use model (model 9)												Time 2–Second-order strengths use model (model 9)									
		$\bar{x}$	σ	Skewness	Kurtosis	CITC	λ	S.E.	δ	AVE	ρ	α	ω	$\bar{x}$	σ	Skewness	Kurtosis	CITC	λ	S.E.	δ	AVE	ρ	α	ω
Affinity										0.50	0.86	0.85	0.85									0.54	0.87	0.87	0.87
	STU_1	5.20	1.05	−0.98	1.82	0.67	0.77	0.03	0.41	-	-	-		5.13	1.11	−0.73	0.23	0.73	0.78	0.02	0.40	-	-	-	
	STU_2	5.24	1.02	−0.70	0.85	0.70	0.79	0.02	0.38	-	-	-		5.20	0.90	−0.55	0.19	0.72	0.79	0.02	0.39	-	-	-	
	STU_3	5.59	0.81	−0.65	0.90	0.61	0.67	0.03	0.56	-	-	-		5.44	0.99	−0.73	0.53	0.71	0.76	0.03	0.43	-	-	-	
	STU_4	5.38	0.93	−0.66	0.79	0.70	0.75	0.03	0.43	-	-	-		5.26	0.88	−0.63	0.45	0.74	0.78	0.02	0.39	-	-	-	
	STU_7	4.77	1.62	−0.43	−0.15	0.56	0.63	0.04	0.61	-	-	-		4.92	1.28	−0.63	0.51	0.62	0.65	0.03	0.57	-	-	-	
	STU_12	4.68	1.53	−0.36	−0.36	0.60	0.63	0.04	0.60	-	-	-		4.89	1.13	−0.53	0.36	0.64	0.65	0.03	0.58	-	-	-	
Active Use										0.58	0.92	0.92	0.92									0.58	0.92	0.92	0.92
	STU_5	4.94	1.35	−0.56	0.36	0.71	0.75	0.03	0.44	-	-	-		5.04	1.28	−0.51	0.19	0.75	0.78	0.02	0.40	-	-	-	
	STU_6	5.22	1.29	−0.69	0.71	0.73	0.76	0.03	0.42	-	-	-		5.31	1.18	−0.61	0.36	0.74	0.77	0.02	0.40	-	-	-	
	STU_8	5.05	1.23	−0.48	−0.04	0.68	0.72	0.03	0.49	-	-	-		5.14	1.10	−0.30	−0.03	0.74	0.77	0.02	0.40	-	-	-	
	STU_9	5.17	1.41	−0.82	0.87	0.66	0.72	0.03	0.49	-	-	-		5.31	1.19	−0.63	0.30	0.70	0.74	0.03	0.45	-	-	-	
	STU_10	5.10	1.41	−0.71	0.29	0.72	0.78	0.02	0.40	-	-	-		5.18	1.03	−0.42	0.13	0.74	0.78	0.02	0.40	-	-	-	
	STU_11	5.10	1.20	−0.65	0.65	0.75	0.81	0.02	0.35	-	-	-		5.11	1.22	−0.51	0.10	0.74	0.76	0.03	0.42	-	-	-	
	STU_13	5.22	1.19	−0.63	0.43	0.71	0.74	0.03	0.45	-	-	-		5.28	1.10	−0.58	0.37	0.72	0.75	0.03	0.44	-	-	-	
	STU_14	5.03	1.19	−0.45	0.04	0.74	0.80	0.02	0.35	-	-	-		5.18	1.03	−0.40	0.08	0.77	0.78	0.02	0.39	-	-	-	
Strength-Use										0.85	0.92	0.93	0.93									0.95	0.97	0.94	0.94
	Affinity	-	-	-	-	-	0.90	0.03	0.19	-	-	-		-	-	-	-	-	0.98	0.02	0.04	-	-	-	
	Active Use	-	-	-	-	-	0.94	0.03	0.12	-	-	-		-	-	-	-	-	0.97	0.02	0.06	-	-	-	

$\bar{x}$, Mean; σ, Standard deviation; CICT, Corrected item total correlation; λ, Standardized factor loadings; S.E., Standard Error; δ, Item Uniqueness; ρ, Composite Reliability; α, Cronbach's Alph; ω:McDonald's Omega.

TABLE 8 | Longitudinal measurement invariance for the second-order strengths use model (model 9).

	Model	χ^2	*df*	χ^2/df	CFI	TLI	RMSEA	SRMR	Model comparison	$\Delta\chi^2$	$\Delta\ \chi^2/df$	ΔCFI	ΔTLI	ΔRMSEA	ΔSRMR	Meets criteria
M1	Configural invariance	850.49	330	2.58	0.921	0.909	0.066 [0.061, 0.072]	0.043	-	-	-	-	-	-	-	Yes
M2	Metric invariance: first-order	868.17	342	2.54	0.919	0.911	0.065 [0.060, 0.071]	0.052	M2 vs. M1	17.68	−0.039	−0.002	0.002	0.001	0.009	Yes
M3	Metric invariance: second-order	879.19	343	2.56	0.918	0.909	0.066 [0.061, 0.071]	0.053	M3 vs. M2	11.02	0.025	−0.001	−0.002	0.001	0.001	Yes
M4	Scalar invariance: first-order	917.75	356	2.58	0.914	0.909	0.066 [0.061, 0.071]	0.057	M4 vs. M3	38.56	0.015	−0.004	0.000	0.000	0.004	Yes
M5	Scalar invariance: second-order	927.51	355	2.61	0.912	0.907	0.067 [0.062, 0.072]	0.057	M5 vs. M4	9.76	0.035	−0.002	−0.002	0.001	0.000	Yes
M6	Strict invariance	942.64	358	2.63	0.910	0.905	0.067 [0.062, 0.073]	0.060	M6 vs. M5	15.13	0.020	−0.002	−0.002	0.000	0.003	Yes

χ^2, Chi-square; df, degrees of freedom; TLI, Tucker-Lewis Index; CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation [90%CI]; SRMR, Standardised Root Mean Square Residual.

On the factorial level, the results showed that all items loaded significantly and sufficiently on their respective factors at both time points. All standardized factor loadings loaded significantly on their respective factors and ranged from 0.63 to 0.81 at Time 1 and 0.65 and 0.78 at Time 2. This exceeds the suggested cut-off criteria of 0.40, as Asparouhov and Muthén (2009) and Kline (2011) suggested. Further, no cross-loadings were present, item uniqueness was acceptable (>0.10 but <0.90; $p < 0.01$), and the average variance extracted was more than 50% for both factors at both time points (Asparouhov and Muthén, 2009; Kline, 2011). Further, all items showed a corrected item-total correlation coefficient larger than 0.3 (ranging from 0.56 to 0.77), implying that all items belong to their respective factors. This contrasts with other studies where a single factor of strengths use was reported. In the majority of international studies, several modifications to the SUS scale (such as correlating error terms, and item parceling) were required to enhance model fit and to increase measurement quality (c.f. Wood et al., 2011; Huber et al., 2017; Bu and Duan, 2020; Vuorinen et al., 2020). Enhancing model fit through statistical modification artificially inflates data-model fit but does not address the theoretical reasoning why the instrument did not perform as intended (McNeish et al., 2018). These modifications to the instrument also change the theoretical foundation on which the instrument is built, making comparisons to other studies improbable. Given that no modifications were made to artificially inflate model fit or measurement quality within the current sample, it would seem as though the two-factor model shows more promise.

Finally, the level of internal consistency at both the lower- and upper bound levels for all constructs at both time points suggest that the SUS was a reliable measure of strengths use. This is inline with other findings that showed high levels of internal consistency for the overall strengths use factor in the USA (Govindji and Linley, 2007; Wood et al., 2011), Germany (Huber et al., 2017), China (Bu and Duan, 2020), Finland (Vuorinen et al., 2020), South Africa (Mahomed and Rothmann, 2020), and the U.K. (McTiernan et al., 2020). The second-order factorial model could therefore be used as a reliable measure for Affinity for Strengths and Strengths Use Behaviors within the current context.

Longitudinal Measurement Invariance and Factor Mean Comparisons

The results further showed that strict longitudinal measurement invariance of the SUS could be established over 3 months. Both the components (Affinity for Strengths and Strengths Use Behaviors) and overall strengths use factorial model was therefore measured (and interpreted) equally across time. This implies that the SUS showed similar factor structures, factor loadings, intercepts, and residual errors over time. Therefore, the data provide support for the stability of the SUS over time. When strengths use is assessed at two different time points, the mean difference indicates actual changes over time (Wong and Wong, 2020), rather than changes in the meaning of the constructs (Duncan et al., 2013). Meaningful comparisons between means

TABLE 9 | Concurrent and predictive validity of the second-order strengths use model on study engagement.

Model	Regressive path			Standardized					Validity established
				β	S.E	*t*-value	*p*	R^2	
Concurrent validity	Strengths use time 1	→	Study engagement time 1	0.49	0.05	9.58	0.00	0.24	Yes
	Strengths use time 2	→	Study engagement time 2	0.58	0.04	13.54	0.00	0.33	Yes
Predictive validity	Strengths use time 1	→	Study engagement time 2	0.47	0.05	9.18	0.00	0.22	Yes

→*Regression.*

and growth trajectories can, therefore, be made over time (Duncan et al., 2013).

No mean differences in neither strengths use, nor its components were reported within the current study. This shows that strengths use remained relatively stable over time (Duncan et al., 2013). This is in line with the assumption proposed by Peterson and Seligman (2004) and Govindji and Linley (2007) that strengths are considered psychological traits and that both the trait and its active use remain relatively stable over time. The stability in both the affinity for and active use of strengths would remain unchanged unless individuals are exposed to- or are engaging in strengths-based developmental initiatives (Seligman, 2012; Huber et al., 2017).

These findings are also relevant for long-term studies on strengths use like within intervention research. When employing longitudinal analytical strategies such as Latent Growth Modeling, where there are multiple measurement occasions, the input matrix of factors is large (Widaman et al., 2010). This leads to convergent problems and/or results in various statistical artifacts, which affects the interpretation of the results (Duncan et al., 2013; Wong and Wong, 2020). To reduce the complexity of these models, researchers would either parcel items or create mean scores to simplify the measurement models at the different time points within the study (Widaman et al., 2010). However, item parceling affects measurement invariance assessments at an item level, producing biased results (Meade and Kroustalis, 2006). Item parceling in longitudinal research should only be considered if there is a strong theoretical argument for such or when strict longitudinal measurement invariance has previously been established (Widaman et al., 2010; Duncan et al., 2013). Therefore, establishing strict longitudinal measurement invariance in the current study supports other researchers to parcel items on the scale when used in similar populations. However, these findings would need to be replicated in other populations to establish firmer conclusions.

The Relationship Between Strengths Use and Study Engagement

The final objective of the paper was to establish criterion validity through relating Strengths Use to Study Engagement. First, concurrent validity was established by showing that Strengths Use at both Time 1 and Time 2 was positively related to engagement at the same time stamps. Further, predictive validity was established by showing that Strengths Use at Time 1 predicted Study Engagement at Time 2. The results imply that when a student can activate his/her strengths during their studies, it would lead to higher levels of study-related engagement. According to Van Woerkom et al. (2016b) this is because when individuals use their strengths, it aids them to live more authentically and therefore acts as an energizing mechanism. When students use their strengths during their studies, it leads to more inspiration, enthusiasm, excitement, and dedication to their study-related content (Seligman, 2012). Active strengths use therefore, has an invigorating effect (Huber et al., 2017). The results are aligned to several studies showing that higher levels of active strengths use lead to increased study and work-related engagement (Ouweneel et al., 2011; Seligman, 2012; Stander et al., 2015; Kwok and Fang, 2020). The SUS can therefore, be used as a measure to predict study engagement.

Limitations and Recommendations

Although the study provides some unique insights, it is not without its limitations. First, the sample is relatively small and drawn from a single Dutch student population from a single Dutch University. This implies that the results may not be generalizable to other contexts or even institutions. It is suggested that the study be replicated in other educational contexts to further investigate the viability of the SUS as a measure of strengths use. Second, the interpretation of what is considered a strength was left to the participant. Therefore it is questionable if the SUS in deed measures "natural capacities coming from within that we yearn to use, that enable authentic expression, energize us and belong to positive traits and/or psychological capacities/talents refined with knowledge and skills" as articulated by Govindji and Linley (2007, p. 147). Although considered a strength of the instrument to measure strengths use in a general way, without providing a clear definition of what a strength is, could possibly lead to statistical artifacts within the data. This is because participants could understand strengths use as character strengths, talents, skills, abilities, or any other behavior pertaining to doing someone really well. It is suggested that the definition of a strength, as articulated by Govindji and Linley (2007), be included in the instructions to participants in the future. Further, it is suggested that a qualitative, open-ended question be added to the SUS requesting participants to both describe their definition of a strength, and provide 3 practical examples of their own strengths. This would aid in standardization in interpretation between participants.

Third, only student engagement was used as a metric to investigate criterion validity. Given that student engagement is a single (self-report) factor, future research should consider including "hard" or "objective" criterions such as academic performance or academic throughput. Fourth, the sample

consisted out of predominantly males. Future studies should aim to include a more even distribution in terms of gender. Fifth, future research should investigate the convergent and discriminant validity of the SUS. Evidence for convergent validity could be tested by comparing the SUS with a measure of personal resilience, the 10-item Connor–Davidson Resilience Scale (CD-RISC) (Campbell-Sills and Stein, 2007). Evidence for discriminant validity could be tested by differentiating the SUS from a measure of personal emotional intelligence (e.g., the 16-item Wong and Law Emotional Intelligence Scale; Wong and Law, 2002). Additionally, future research should also investigate the correlation between the quantitative responses on the SUS with qualitative perceptions about the connotation they gave to strengths use when they responded to the items. This relates to what Alexandrova (2017) refers to as tracking the measurement of a construct as it is understood and endorsed by the respondent.

Sixth, it is suggested that more diverse population groups be considered for future validation studies. The SUS would benefit from a large scale cross-cultural validation study to determine if strengths use is seen and measured the same between cultures. Finally, future research should investigate the psychometric properties of a short-form SUS for rapid use by researchers and practitioners.

CONCLUSION

Strengths use is a crucial factor to consider when designing both educational programmes and positive psychological interventions at universities. The current study shows support for the use of the SUS as a practical means to assess strengths use and to track the effectiveness of strengths use interventions within higher education environments.

DATA AVAILABILITY STATEMENT

The data employed in this study is available upon reasonable request from suitably qualified individuals and will be provided without undue reservations. Data management is aligned to the requirements of the General Data Protection Regulation.

ETHICS 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. Secondary data were employed in this study and the APA guidelines on Ethical Research practices were adhered to. All procedures performed in this study were in accordance with the ethical standards of the institutional requirements and in line with the Declaration of Helsinki. Written, informed consent was obtained from all participants before being permitted to complete the questionnaires. Participation in the study was entirely voluntary, participants were informed of their rights and responsibilities, and that they had the right to withdraw at any time. Given the nature of the study, no ethical clearance was required by the institution.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

We would like to extend our appreciation for the Associate Editor (Prof. Jeff Klibert) and the two reviewers that aided us in enhancing the quality of this manuscript. We particularly appreciate their empowering and developmental approach toward their feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.676153/full#supplementary-material

REFERENCES

Alexandrova, A. (2017). *A Philosophy for the Science of Wellbeing*. New York, NY: Oxford University Press. doi: 10.1093/oso/9780199300518.001.0001

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

Basson, M. J., and Rothmann, S. (2019). Pathways to flourishing among pharmacy students: the role of study demands and lecturer support. *J. Psychol. Afr.* 29, 338–345. doi: 10.1080/14330237.2019.1647953

Bergin, A., and Pakenham, K. (2015). Law student stress: relationships between academic demands, social isolation, career pressure, study/life imbalance and adjustment outcomes in law students. *Psychiatry Psychol. Law* 22, 388–406. doi: 10.1080/13218719.2014.960026

Bhatnagar, V. R. (2020). Conceptualizing employee strengths at work and scale development. *Manag. Res. Rev.* 43, 1273–1288. doi: 10.1108/MRR-08-2019-0367

Blanco, C., Okuda, M., Wright, C., Hasin, D. S., Grant, B. F., Liu, S. M., et al. (2008). Mental health of college students and their non–college-attending peers: results from the national epidemiologic study on alcohol and related conditions. *Arch. Gen. Psychiatry* 65, 1429–1437. doi: 10.1001/archpsyc.65.12.1429

Bu, H., and Duan, W. (2020). Strength-based flourishing intervention to promote resilience in individuals with physical disabilities in disadvantaged communities: a randomized controlled trial. *Res. Soc. Work Pract.* 31, 53–64. doi: 10.1177/1049731520959445

Cadime, I., Lima, S., Pinto, A. M., and Ribeiro, I. (2016). Measurement invariance of the Utrecht work engagement scale for students: a study across secondary school pupils and university students. *Eur. J. Dev. Psychol.* 13, 254–264. doi: 10.1080/17405629.2016.1148595

Campbell-Sills, L., and Stein, M. B. (2007). Psychometric analysis and refinement of the connor–davidson resilience scale (CD-RISC): validation of a 10-item measure of resilience. *J. Trauma. Stress* 20, 1019–1028. doi: 10.1002/jts.20271

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Eq. Model.* 14, 464–504. doi: 10.1080/10705510701301834

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

Duan, W., Li, J., and Mu, W. (2018). Psychometric characteristics of strengths knowledge scale and strengths use scale among adolescents. *J. Psychoeduc. Assess.* 36, 756–760. doi: 10.1177/0734282917705593

Dubreuil, P., Forest, J., and Courcy, F. (2014). From strengths use to work performance: the role of harmonious passion, subjective vitality, and concentration. *J. Posit. Psychol.* 9, 335–349. doi: 10.1080/17439760.2014.898318

Dubreuil, P., Forest, J., Gillet, N., Fernet, C., Thibault-Landry, A., Crevier-Braud, L., et al. (2016). Facilitating wellbeing and performance through the development of strengths at work: results from an intervention program. *Int. J. Appl. Posit. Psychol.* 1, 1–19. doi: 10.1007/s41042-016-0001-8

Duncan, T. E., Duncan, S. C., and Strycker, L. A. (2013). *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Application*. New York, NY: Routledge Academic. doi: 10.4324/9780203879962

Ebert, D. D., Van Daele, T., Nordgreen, T., Karekla, M., Compare, A., Zarbo, C., et al. (2018). Internet-and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health. *Eur. Psychol.* 23, 167–187. doi: 10.1027/1016-9040/a000318

Feldt, T., Leskinen, E., Kinnunen, U., and Mauno, S. (2000). Longitudinal factor analysis models in the assessment of the stability of sense of coherence. *Pers. Individ. Dif.* 28, 239–257. doi: 10.1016/S0191-8869(99)00094-X

Forest, J., Mageau, G. A., Crevier-Braud, L., Bergeron, É., Dubreuil, P., and Lavigne, G. L. (2012). Harmonious passion as an explanation of the relation between signature strengths' use and wellbeing at work: test of an intervention program. *Hum. Relat.* 65, 1233–1252. doi: 10.1177/0018726711433134

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: the broaden-and-build theory of positive emotions. *Am. Psychol.* 56:218. doi: 10.1037/0003-066X.56.3.218

Govindji, R., and Linley, P. A. (2007). Strengths use, self-concordance and wellbeing: implications for strengths coaching and coaching psychologists. *Int. Coach. Psychol. Rev.* 2, 143–153. doi: 10.1037/t54221-000

Hayes, A. F., and Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But… *Commun. Methods Measures* 14, 1–24. doi: 10.1080/19312458.2020.1718629

Houghton, S., Lawrence, D., Hunter, S. C., Rosenberg, M., Zadow, C., Wood, L., et al. (2018). Reciprocal relationships between trajectories of depressive symptoms and screen media use during adolescence. *J. Youth Adolesc.* 47, 2453–2467. doi: 10.1007/s10964-018-0901-y

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

Huber, A., Webb, D., and Höfer, S. (2017). The German version of the strengths use scale: the relation of using individual strengths and wellbeing. *Front. Psychol.* 8:637. doi: 10.3389/fpsyg.2017.00637

IBM SPSS (2019). *IBM SPSS Software Version 26*. IBM Press. doi: 10.4135/9781483385655.n2

Joseph, S., and Linley, P. A. (2005). Positive adjustment to threatening events: An organismic valuing theory of growth through adversity. *Rev. Gen. Psychol.* 9, 262–280. doi: 10.1037/1089-2680.9.3.262

Kaiser, H. F., and Rice, J. (1974). Little jiffy, mark IV. *Educ. Psychol. Meas.* 34, 111–117. doi: 10.1177/001316447403400115

Kienfie Liau, A., Chow, D., Tan, T. K., and Senf, K. (2011). Development and validation of the personal strengths inventory using exploratory and confirmatory factor analyses. *J. Psychoeduc. Assess.* 29, 14–26. doi: 10.1177/0734282910365648

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor. Dentist. Endod.* 38, 52–54. doi: 10.5395/rde.2013.38.1.52

Kline, R. B. (2011). *Principles and Practices of Structural Equation Modelling, 3rd Edn.* New York, NY: Gilford Press.

Kwok, S. Y., and Fang, S. (2020). A cross-lagged panel study examining the reciprocal relationships between positive emotions, meaning, strengths use and study engagement in primary school students. *J. Happiness Stud.* 22, 1033–1053. doi: 10.1007/s10902-020-00262-4

Lesener, T., Pleiss, L. S., Gusy, B., and Wolter, C. (2020). The study demands-resources framework: an empirical introduction. *Int. J. Environ. Res. Public Health* 17:5183. doi: 10.3390/ijerph17145183

Linley, A., Willars, J., and Biswas-Diener, R. (2010). *The Strengths Book: Be Confident, be Successful, and Enjoy Better Relationships by Realizing the Best of You*. Coventry: CAPP.

Linley, P. A., and Harrington, S. (2006). Strengths coaching: a potential-guided approach to coaching psychology. *Int. Coach. Psychol. Rev.* 1, 37–46.

Littman-Ovadia, H., Lavy, S., and Boiman-Meshita, M. (2017). When theory and research collide: examining correlates of signature strengths use at work. *J. Happiness Stud.* 18, 527–548. doi: 10.1007/s10902-016-9739-8

Mahomed, F. E. (2019). *Thriving of academics in higher education institutions: a strengths-based approach* (Doctoral dissertation). Vanderbijlpark: North-West University.

Mahomed, F. E., and Rothmann, S. (2020). Strength use, training and development, thriving, and intention to leave: the mediating effects of psychological need satisfaction. *South Afr. J. Psychol.* 50, 24–38. doi: 10.1177/0081246319849030

Matsuguma, S., and Niemiec, R. M. (2020). Hikikomori from the perspective of overuse, underuse, and optimal use of character strengths. *Int. J. Appl. Posit. Psychol.* 1, 1–13. doi: 10.1007/s41042-020-00047-3

McGrath, R. E. (2015). Character strengths in 75 nations: an update. *J. Posit. Psychol.* 10, 41–52. doi: 10.1080/17439760.2014.888580

McNeish, D., An, J., and Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cut-offs in latent variable models. *J. Pers. Assess.* 100, 43–52. doi: 10.1080/00223891.2017.1281286

McNeish, D., and Hancock, G. R. (2018). The effect of measurement quality on targeted structural model fit indices: a comment on Lance, Beck, Fan, and Carter 2016. *Psychol. Methods* 23, 184–190. doi: 10.1037/met0000157

McTiernan, K., Gullon-Scott, F., and Dudley, R. (2020). An exploration of strength use and its relationship with life satisfaction, positive self-beliefs and paranoid ideation. *Int. J. Wellbeing* 10, 53–70. doi: 10.5502/ijw.v10i2.883

Meade, A. W., and Kroustalis, C. M. (2006). Problems with item parceling for confirmatory factor analytic tests of measurement invariance. *Organ. Res. Methods* 9, 369–403. doi: 10.1177/1094428105283384

Miglianico, M., Dubreuil, P., Miquelon, P., Bakker, A. B., and Martin-Krumm, C. (2020). Strength use in the workplace: a literature review. *J. Happiness Stud.* 21, 737–764. doi: 10.1007/s10902-019-00095-w

Muñiz, J., Elosua, P., and Hambleton, R. K. (2013). International test commission guidelines for test translation and adaption: second edition. *Psicothema* 25, 151–157. doi: 10.7334/psicothema2013.24

Muthén, L. K., and Muthén, B. O. (2020). *Mplus (Version 8.4). [Statistical software]*. Los Angeles, CA: Muthén and Muthén.

Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory, 3rd Edn.* New York, NY: McGraw-Hill.

Ouweneel, E., Le Blanc, P. M., and Schaufeli, W. B. (2011). Flourishing students: a longitudinal study on positive emotions, personal resources, and study engagement. *J. Posit. Psychol.* 6, 142–153. doi: 10.1080/17439760.2011.558847

Peterson, C., Park, N., and Seligman, M. E. (2006). Greater strengths of character and recovery from illness. *J. Posit. Psychol.* 1, 17–26. doi: 10.1080/17439760500372739

Peterson, C., and Seligman, M. E. P. (2004). *Character Strengths and Virtues: A Handbook and Classification*. New York, NY: Oxford University Press.

Price, L. (2017). *Psychometric Methods. Theory into Practice*. New York, NY: Guilford Press.

Proyer, R. T., Gander, F., Wellenzohn, S., and Ruch, W. (2015a). Strengths-based positive psychology interventions: a randomized placebo-controlled online trial on long-term effects for a signature strengths- vs. a lesser strengths-intervention. *Front. Psychol.* 6:456. doi: 10.3389/fpsyg.2015.00456

Proyer, R. T., Wellenzohn, S., Gander, F., and Ruch, W. (2015b). Toward a better understanding of what makes positive psychology interventions work: predicting happiness and depression from the person × intervention fit in a follow-up after 3.5 years. *Appl. Psychol. Health Wellbeing* 7, 108–128. doi: 10.1111/aphw.12039

Quinlan, D., Swain, N., and Vella-Brodrick, D. A. (2012). Character strengths interventions: building on what we know for improved outcomes. *J. Happiness Stud.* 13, 1145–1163. doi: 10.1007/s10902-011-9311-5

Rashid, T., Louden, R., Wright, L., Chu, R., Maharaj, A., Hakim, I., et al. (2017). "Flourish: a strengths-based approach to building student resilience," in *Positive Psychology Interventions in Practice*, ed C. Proctor (Cham: Springer), 29–45. doi: 10.1007/978-3-319-51787-2_3

Rath, T. (2007). *StrengthsFinder 2.0*. New York, NY: Simon and Schuster.

Raykov, T. (2009). Evaluation of scale reliability for unidimensional measures using latent variable modeling. *Meas. Eval. Counsel. Dev.* 42, 223–232. doi: 10.1177/0748175609344096

Richter, S., Van Zyl, L.E., Roll, L.C, and Stander, M. W. (2020). Positive psychological coaching tools: a systematic literature review. *Front. Psychol.* 11:793. doi: 10.3389/fpsyg.2020.00793

Roll, L. C., Van Zyl, L. E., and Griep, Y. (2019). "Brief positive psychological interventions within multi-cultural organizational contexts: a systematic literature review," in *Theoretical Approaches to Multi-Cultural Positive Psychological Interventions*, eds L. E. van Zyl, and S. Rothmann (Cham: Springer), 523–544. doi: 10.1007/978-3-030-20583-6_24

Ruch, W., Weber, M., and Park, N. (2014). Character strengths in children and adolescents. *Eur. J. Psychol. Assess.* 30, 57–64. doi: 10.1027/1015-5759/a000169

Schaufeli, W. B., Bakker, A. B., and Salanova, M. (2006). The measurement of work engagement with a short questionnaire: a cross-national study. *Educ. Psychol. Meas.* 66, 701–716. doi: 10.1177/0013164405282471

Schaufeli, W. B., Salanova, M., González-Romá, V., and Bakker, A. B. (2002). The measurement of engagement and burnout: a two sample confirmatory factor analytic approach. *J. Happiness Stud.* 3, 71–92. doi: 10.1023/A:1015630930326

Schutte, N. S., and Malouff, J. M. (2018). Mindfulness and connectedness to nature: a meta-analytic investigation. *Pers. Individ. Dif.* 127, 10–14. doi: 10.1016/j.paid.2018.01.034

Seligman, M. E. (2012). *Flourish: A Visionary New Understanding of Happiness and Wellbeing*. New York, NY: Simon and Schuster.

Seligman, M. E., and Csikszentmihalyi, M. (2014). "Positive psychology: an introduction," in *Flow and the Foundations of Positive Psychology*, eds M. Csikszentmihalyi (Dordrecht: Springer), 279–298. doi: 10.1007/978-94-017-9088-8_18

Seligman, M. E. P., Steen, T. A., Park, N., and Peterson, C. (2005). Positive psychology progress: empirical validation of interventions. *Am. Psychol.* 60, 410–421. doi: 10.1037/0003-066X.60.5.410

Shi, D., Lee, T., and Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educ. Psychol. Meas.* 79, 310–334. doi: 10.1177/0013164418783530

Snow, N. E. (2019). Positive psychology, the classification of character strengths and virtues, and issues of measurement. *J. Posit. Psychol.* 14, 20–31. doi: 10.1080/17439760.2018.1528376

Stander, F. W., Diedericks, E., Mostert, K., and De Beer, L. T. (2015). Proactive behaviour towards strength use and deficit improvement, hope and efficacy as predictors of life satisfaction amongst first-year university students. *S.A. J. Indus. Psychol.* 41, 1–10. doi: 10.4102/sajip.v41i1.1248

Van Woerkom, M., Mostert, K., Els, C., Bakker, A. B., De Beer, L. T., and Rothmann, S. (2016a). Strengths use and deficit correction in organizations: development and validation of a questionnaire. *Eur. J. Work Organ. Psychol.* 25, 1–16. doi: 10.1080/1359432X.2016.1193010

Van Woerkom, M., Oerlemans, W., and Bakker, A. B. (2016b). Strengths use and work engagement: a weekly diary study. *Eur. J. Work Organ. Psychol.* 25, 384–397. doi: 10.1080/1359432X.2015.1089862

Van Zyl, L. E. (2014). *Research Methodology for the Economic and Management Sciences*. London: Pearson.

Van Zyl, L. E., Efendic, E., Rothmann, S., and Shankland, R. (2019). "Best-practice guidelines for positive psychological intervention research design," in *Positive Psychological Intervention Design and Protocols for Multi-Cultural Contexts*, eds L. E. van Zyl, and S. Rothmann Sr. (Cham: Springer), 1–32. doi: 10.1007/978-3-030-20020-6_1

Van Zyl, L. E., and Rothmann, S. (2019). *Positive Psychological Intervention Design and Protocols for Multi-Cultural Contexts*. Cham: Springer International Publishing. doi: 10.1007/978-3-030-20020-6

Von Eye, A. (Ed.). (1990). *Statistical Methods in Longitudinal Research: Principles and Structuring Change (Vol. 1)*. New York, NY: Elsevier Press.

Vuorinen, K., Hietajärvi, L., and Uusitalo, L. (2020). Students' usage of strengths and general happiness are connected via school-related factors. *Scand. J. Educ. Res.* 1–13. doi: 10.1080/00313831.2020.1755361

White, C. A. (2016). *Meta-analyses of positive psychology interventions on wellbeing and depression: reanalyses and replication* (Doctoral dissertation). Kelowna, BC: University of British Columbia.

White, C. A., Uttl, B., and Holder, M. D. (2019). Meta-analyses of positive psychology interventions: the effects are much smaller than previously reported. *PLoS ONE* 14:e0216588. doi: 10.1371/journal.pone.0216588

Wickrama, K. K., Lee, T. K., O'Neal, C. W., and Lorenz, F. O. (2016). *Higher-Order Growth Curves and Mixture Modeling With Mplus: A Practical Guide.* London. Routledge Publications. doi: 10.4324/9781315642741

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. Persp.* 4, 10–18. doi: 10.1111/j.1750-8606.2009.00110.x

Wong, C., and Law, K. S. (2002). Wong and law emotional intelligence scale. *Psyctests.* doi: 10.1037/t07398-000

Wong, J., and Wong, X. (2020). *Structural Equation Modelling: Applications Using Mplus (2nd Edn.)*. Chichester: Wiley and Sons.

Wood, A. M., Linley, P. A., Maltby, J., Kashdan, T. B., and Hurling, R. (2011). Using personal and psychological strengths leads to increases in wellbeing over time: a longitudinal study and the development of the strengths use questionnaire. *Pers. Ind. Diff.* 50, 15–19. doi: 10.1016/j.paid.2010.08.004

Zijlmans, E. A., Tijmstra, J., van der Ark, L. A., and Sijtsma, K. (2019). Item-score reliability as a selection tool in test construction. *Front. Psychol.* 9:2298. doi: 10.3389/fpsyg.2018.02298

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

Copyright © 2021 van Zyl, Arijs, Cole, Gliíska-Newes, Roll, Rothmann, Shankland, Stavros and Verger. 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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 July 2021
doi: 10.3389/fpsyg.2021.701510

Mindset as Characteristic Adaptations: Using Response Surface Analysis to Assess Mindset in the Personality System

Juliette L. Ratchford[1]*, *Emily G. Williams*[1], *Leanne Bishara*[2], *Benjamin J. Houltberg*[3] and *Sarah A. Schnitker*[1]

[1] Science of Virtues Laboratory, Department of Psychology and Neuroscience, Baylor University, Waco, TX, United States, [2] Fuller Theological Seminary, School of Psychology, Pasadena, CA, United States, [3] Search Institute, Minneapolis, MN, United States

This study aimed to assess the congruencies and discrepancies between mindset domains in relation to well-being and sought to demonstrate that mindset falls into the characteristic adaptation level of personality. Data ($N = 618$, $M_{age} = 16.07$, $SD_{age} = 0.99$) from Wave 1 of a longitudinal study on primarily ethnic-minority adolescents were used in response surface analyses to examine the effects of (in)congruence on well-being. The response surface analyses suggested no overall congruence effect between moral and ability mindsets. However, two-thirds of the participants demonstrated differing levels of mindsets, highlighting the domain specificity of mindsets. Results suggest that mindsets are contextual, domain-specific constructs, suiting the characteristic adaptation level of personality. Congruence for moral and ability mindset does not affect adolescent well-being.

Keywords: mindset, personality, characteristic adaptations, well-being, response surface analysis

OPEN ACCESS

Edited by:
Arianna Costantini, University of Verona, Italy

Reviewed by:
Chris Martin, Georgia Institute of Technology, United States
Daoyang Wang, Hangzhou Normal University, China

***Correspondence:**
Juliette L. Ratchford
julietteratchford@gmail.com

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

Received: *28 April 2021*
Accepted: *22 June 2021*
Published: *20 July 2021*

Citation:
Ratchford JL, Williams EG, Bishara L, Houltberg BJ and Schnitker SA (2021) Mindset as Characteristic Adaptations: Using Response Surface Analysis to Assess Mindset in the Personality System. Front. Psychol. 12:701510. doi: 10.3389/fpsyg.2021.701510

INTRODUCTION

In the three decades since Dweck and Leggett (1988) first defined mindsets, or the tendencies that people hold in viewing their capabilities and attributes as more or less malleable, research on the topic has proliferated. Dweck and Leggett (1988) compared two types of mindsets: growth mindset, wherein the person believes that certain attributes (e.g., intelligence) are inherently malleable and can be changed, and fixed mindset, in which the person believes that a certain attribute is an unchangeable, static trait. Individuals vacillate across a spectrum with growth mindset at one pole and fixed mindset at the other (Dweck and Leggett, 1988). Research has largely placed an emphasis on developing interventions to foster growth mindsets in children and adolescents (Dweck and Yeager, 2019). Yet, even with the increasing interest in mindsets, little research has investigated mindset variability across distinct domains.

People hold mindsets across many unique domains including intelligence (e.g., Dweck and Leggett, 1988), morality (e.g., Chiu et al., 1997), ability (e.g., Wang et al., 2018), and emotion (e.g., Tamir et al., 2007). Moral mindsets consider beliefs regarding the nature of the moral self (Chiu et al., 1997), whereas ability mindsets may develop in any performance-based domain including music, art, and sports (Wang et al., 2018). The way people organize their goals within the self-system, regulate their behavior and emotions in the pursuit of goals, and flourish in a given domain

is largely dependent on their mindsets, with a more growth mindset typically predicting higher goal achievement (e.g., Dweck and Leggett, 1988; Dupeyrat and Mariné, 2005).

Researchers studying mindsets typically choose a specific domain to investigate that best fits the context of their study, following Dweck (1996) suggestion to adapt items to specific context of the participants (e.g., studying athletic ability in sporting contexts, studying intelligence in academic contexts). Researchers have yet to investigate differences and similarities in mindsets across domains and the way congruencies and discrepancies might impact outcomes. Moreover, little theory or research examines how congruence and incongruence across mindset domains might impact the larger personality system.

Mindset Within the Personality System

Current research typically investigates specific domains of mindset in relation to expected, contextual outcomes. Such a focus suggests an underlying assumption that mindset domains diverge from one another, but this assumption has yet to be investigated empirically. Most psychological measures do not attend to contextual influences, instead measuring global, wholistic constructs that assess general tendencies across all situations and domains (Furr, 2009). Global measures typically invoke dispositional traits, the broad consistencies across contexts and time (e.g., agreeableness; Big Five). The extant measures of mindsets diverge from this approach, as each is focused on a specific domain (e.g., ability; morality). This approach is best conceptualized at the characteristic adaptation level of personality, which involves contextualized, specific motivational, social-cognitive, and developmental variables (McAdams and Pals, 2006). If empirical research does not support the domain specificity of mindsets, then researchers might consider whether a trait-based conceptualization is better suited than a characteristic adaptation conceptualization.

Mindset is not typically assessed alongside dispositional traits, and the scant existing research suggests that mindsets are typically uncorrelated with the Big Five traits (e.g., Spinath et al., 2003; Burnette and Pollack, 2013). Higher growth intelligence mindset has been related to higher conscientiousness and lower neuroticism (Satchell et al., 2017). However, there is little consensus on the relation between mindsets across multiple domains and dispositional traits. The present study seeks to investigate mindsets for ability and morality in relation to the Big Five traits, which will further clarify relations across the different personality levels and situate mindset within the broader personality system.

The two domains of mindsets of particular interest in the present study are ability and morality. These two domains were selected because of the context of our sample, following Dweck (1996) suggestion to contextualize the mindsets. Our sample consists of adolescents who are mostly engaged in an athletic context (71.9% of sample). In pairing these two domains together, we attempt to tap into how the congruencies between them affect well-being outcomes.

These two domains are often linked with distinct outcomes. Ability mindset is typically associated with outcomes related to self-regulation, achievement, and performance (Burnette and Pollack, 2013), whereas morality mindset is linked to outcomes related to prosociality (Han et al., 2018). These diverging outcomes likely reflect the distinct foci of the two domains. Whereas, ability mindset is focused on people's beliefs about their propensity to be skilled and talented, morality mindset reflects people's beliefs about their innate goodness and moral character of the self and others.

Do Congruencies and Discrepancies Across Mindset Domains Matter?

Despite numerous studies demonstrating the importance of growth mindset in bolstering and buffering well-being, no current studies explore the effects of (in)congruence across different mindset domains on well-being. However, researchers might expect that being highly growth-minded in one domain while being highly fixed-minded in another could lead to detrimental outcomes. Consider the example of Denise, a high school volleyball athlete who has a growth mindset in morality but a fixed mindset in ability. Denise believes that people are capable of changing their moral character but are unable to change their abilities and talents. In situations where only ability or morality are relevant, these distinct mindsets would not affect well-being, but incoherence may occur in scenarios where moral and ability domains overlap. Suppose Denise was struggling in her abilities to perform well in the previous match, so she cheated to make up for her deficits. Would Denise believe that she can change her behavior in the future, which would preference her growth-minded approach to morality, or will she think she is stuck in this pattern of failure, which would preference her fixed-minded approach to her athletic ability? This resulting dissonance among her mindsets could cause conflicts in the self-system that affect her self-regulation and well-being. Moreover, the quantity or quality of the dissonance Denise experiences might differ from her teammates who have other patterns of mindset convergence and divergence. For example, Rachel has a fixed mindset in morality and a growth mindset in ability, and Alexis, Lauren, and LaTonya have matching ability and moral mindsets—but at low, medium, and high levels of growth, respectively. We endeavor to understand whether and how these combinations of mindsets affect self-regulation and well-being.

Because of the dearth in the literature on (in)congruence across mindset domains, research on related characteristic adaptations is necessary to form preliminary hypotheses. Meta-analytic results demonstrate that growth ability mindset predicts distinctive self-regulatory, goal-oriented processes (e.g., goal setting, goal operating, and goal monitoring), which ultimately lead to goal achievement (Burnette and Pollack, 2013). Thus, examination of (in)congruence among personal goals and strivings, may provide the closest parallels with mindsets to form our hypotheses. Harmony among goals is often associated with positive effects, whereas conflict is correlated with maladaptive outcomes (e.g., Sheldon and Elliot, 1999; Gray et al., 2017). On a theoretical level, we expect that congruencies across mindset domains would enhance well-being because the self-concordance model suggests that self-consistency positively affects well-being (Sheldon and Elliot, 1999). Gray et al. (2017) meta-analysis

suggests that adolescents are much more likely than any other age demographic to experience goal conflict, which highlights the importance of assessing these effects during a period when the self-system is developing.

To date, there is very little research regarding the overlap between mindset domains within people. Meta-analytic findings do suggest small to moderate moderation effects for mindset domain on the associations between mindset and well-being, but these analyses compare domains across studies (Burnette et al., 2020). Moreover, most mindset analyses are typically conducted on a linear level, which may mask non-linear effects (e.g., Dawson, 2014). As such, one line of empirical inquiry in the present study is determining whether congruence is of import to different mindset domains and whether such effects might be interactive or non-linear.

Based on these studies, congruence of mindsets across domains would theoretically lead to positive outcomes, and the relevance of domain-specific mindsets to an individual's identity may influence the impact that any discrepancies or congruencies in mindsets have on the self (Sheldon and Elliot, 1999). Additionally, the literature on mindsets would suggest that congruencies among growth mindsets would be the most beneficial for well-being as growth mindsets are associated with flourishing (Howell et al., 2016; Zhao et al., 2021).

Whereas, most studies to date have examined conflict and unity regarding the *content* of the developing self-system (e.g., conflict between morality and ability goals within the self-system), we propose to examine the effects of incongruencies and congruencies in mindset *processes* across domains (e.g., diverging mindset for morality and ability). Such consideration will inform how harmony and conflict in the processes of the self-system affect well-being. An analytic approach that allows for such investigation is response surface analysis (RSA). Much of the personality literature to date using within-person RSA analyzes congruence and incongruence between desired and actual states in relation to well-being outcomes (e.g., Brandstätter et al., 2016; Verhagen et al., 2018). Such studies typically find that incongruence between desired and actual states is associated with worse well-being, whereas congruence is associated with higher well-being (Verhagen et al., 2018). This type of dissonance between desired and actual states may carry over to dissonance in mindset domains as well. Moreover, RSA allows for the assessment of non-linear effects, which is useful for understanding mindset. It is possible that growth mindsets are not incrementally beneficial. Instead, it might be possible that there are diminishing returns on growth mindset after reaching a certain level, or mid-levels of mindsets might serve as resources because they provide flexibility or as vulnerabilities because they are less coherent. Although exploratory, inspection for such non-linear effects in addition to interactions is overdue in the mindset literature.

Outcomes of Interest

We selected several domains of outcomes with our mindsets: self-system, personality, self-regulation, and well-being. These domains were pertinent to our interests given the rich literature regarding their associations with individual, specific mindset domains.

Self-Systems and Trait Personality

Regarding the self-system, self-efficacy, contingencies of self-worth, and trait personality were assessed. Much of the existing literature points to positive associations between higher growth mindsets and adaptive self-system organization like increased self-efficacy (Komarraju and Nadler, 2013; Diseth et al., 2014). Although mindset is reliably related to self-efficacy, Niiya et al. (2010) found that academic contingent self-worth suppressed the influence of mindsets, such that adolescents with scores across the spectrum of fixed and growth intelligence mindsets engaged in self-handicapping behaviors if their self-worth was contingent on academics. This suggests that contingency of self-worth might suppress or interact with mindset, as both deal with specific domains (Crocker et al., 2003). The present study assessed trait-level personality to compare with mindset, given that we expect mindset to fall into the characteristic adaptation level of personality.

Self-Regulation

Mindset is linked with the ability to self-regulate, such that those with higher growth mindsets engage self-regulation more effectively (Burnette and Pollack, 2013). Higher growth mindsets of willpower buffered adverse effects of ego-depletion and emotional dissonance (Konze et al., 2019) and increased self-control in resisting temptations and controlling emotions (Bernecker and Job, 2017). Self-regulation is a central component in people's ability to pursue goals (Sheldon and Kasser, 1998). Because goal consistency leads to better outcomes (Sheldon and Elliot, 1999), it seems likely that congruence across multiple mindset domains may have a positive effect on self-regulatory behaviors and strategies. Emmons et al. (1993) found that goal conflict across domains typically undermines self-regulation, though this has not been explored regarding mindset domains. Thus, this study investigates how congruence or incongruence across ability and moral mindsets affects self-regulatory processes including general regulatory behaviors and self-control.

Well-Being

The literature on mindset strongly establishes a link between growth mindset and well-being outcomes like subjective well-being (Zhao et al., 2021), psychological well-being (Zeng et al., 2016), and hedonic and eudaimonic well-being (Howell et al., 2016). In much the same way, more fixed mindsets are associated with worse well-being such as greater negative affect (King, 2016), anxiety, and depression (Schroder et al., 2019). Fear of failure is also associated with higher fixed mindset (Lewis et al., 2020) and can undermine self-image and sense of achievement (Sagar et al., 2007). Growth mindset buffers against negative well-being and increase positive outcomes. In a single-session growth mindset intervention with adolescents, researchers found that adolescents in the intervention group had significant improvements in both self- and parent-reported depression and anxiety compared to the control group (Schleider and Weisz, 2018). As such,

anxiety, depression, and fear of failure were assessed in relation to mindsets.

Current Study

We seek to contribute to the broader discussion of domain specificity and domain generality of mindset while exploring how cohesion and divergence across mindset affect self-system, regulatory, and well-being outcomes. Given that mindset theoretically falls into the characteristic adaptation level of personality, we expect differences across specific domains of mindset. Such exploration will contribute to the literature on personality systems.

We chose to investigate these domains in adolescent participants because such participants are most likely still developing their self-system, which consequently shapes self-regulatory strategies and well-being outcomes. The development of the self-system is characterized by differentiation and hierarchical integration during adolescence (Werner, 1957). Increasing cognitive capacities result in adolescents possessing the ability to differentiate themselves across multiple contexts and domains. At the same time, such differentiation across contexts is pieced into a coherent whole by way of hierarchical integration. Indeed, adolescents with less coherence across domains typically experience self-regulatory deficits, as evinced in behaviors like greater impulsivity (Goth et al., 2012) and tendencies to engage in risky behaviors (Schwartz et al., 2015). In contrast, those with greater coherence and integration usually have a stronger sense of purpose and experience greater subjective well-being (Sumner et al., 2015). Additionally, most mindset intervention studies are directed toward adolescents because this may be a critical period to develop growth mindset (Dweck and Yeager, 2019).

Hypotheses

Based on the literature showing that mindsets are associated with goal pursuit processes (including goal setting, pursuit, operating, and monitoring) coupled with the robust literature showing that goal conflict is detrimental for the self and well-being (e.g., Sheldon and Elliot, 1999; Gray et al., 2017), we hypothesize that (*H1*) congruence among the mindset domains will lead to better well-being outcomes. Additionally, based on our conceptualization of mindsets as characteristic adaptations, we hypothesize that (*H2*) mindset domains will demonstrate discrepancies for a majority of participants, indicating domain specificity in line with the characteristic adaptation level of personality, as opposed to dispositional traits. Based on the literature on fixed and growth mindsets, we hypothesize that (*H3*) congruence among the mindset domains with lower fixed mindsets (and thus higher growth mindsets) will be associated with better well-being outcomes.

Analytic Plan

To assess congruence and discrepancies, we used response surface analysis (RSA). RSA assesses the congruence and incongruence of two variables (X and Y) in relation to a given outcome (Z) using polynomial regression. This places data into a three-dimensional space as X and Y are used to predict Z (see **Figure 1** for example RSA plot). This approach is of particular use in the present study, as it allows us to consider the within-person effects of domain specificity across various outcomes. Within the present study, X is moral mindset, and Y is ability mindset; both are assessed in relation to Z, which is represented by different self-system, personality, regulatory, and well-being outcomes in each analysis.

To assess *H1* and *H3*, RSAs were estimated for each pair of moral and ability fixed mindsets in *RStudio* (*R* version 3.6.1) with the RSA Package (Schönbrodt and Humberg, 2020)[1]. This package allowed us to consider how mindset congruence and incongruence were associated with personality traits, self-system indicators, self-regulatory behaviors, and well-being outcomes. RSA produces polynomial regression coefficients (b_1–b_5), which are used to compute the surface coefficients (a_1–a_4), which are then used to compute the effects of (in)congruence on the outcome. The slope and curvature of the line of congruence (LOC) and line of incongruence (LOIC) are reported. The LOC is analyzed with two tests: a_1 and a_2. A_1 ($b_1 + b_2$) assesses the slope of the LOC with whether an outcome is higher when moral and ability fixed mindset match at higher levels or lower levels. A_2 ($b_3 + b_4 + b_5$) assesses the curvilinear effect of the LOC with whether an outcome is higher when moral and ability fixed mindset converge at extreme levels or midrange levels. The LOIC is considered with two tests: a_3 and a_4. The slope of LOIC is tested with a_3 (b_3–b_4), which examines whether the outcome is higher when moral fixed mindset is higher than ability fixed mindset or vice versa. A_4 (b_3–$b_4 + b_5$) tests the curvilinear effect of the LOIC with whether congruence is better or worse than discrepancies in moral and ability fixed mindset. Additionally, following suggestions from Humberg et al. (2019), to determine an overall congruency effect, p_{10} must be non-significant, the confidence interval of p_{11} must include 1, a_4 must be significantly negative, and a_1, a_2, and a_3 must be non-significant. Specific coefficients of the LOC or LOIC should only be interpreted if there is an overall congruency effect.

In addition to RSAs, to address *H2*, we conducted correlational and confirmatory factor analyses on the two mindset domains to determine the overlap and uniqueness of moral and ability mindset domains.

MATERIALS AND METHODS

Participants

Data were taken from Wave 1 of a 4-wave longitudinal study[2]. Participants ($N = 618$) were recruited via athletic and extracurricular programs at partnering schools in metropolitan southern California. The majority of the sample was female (58.3%). Most participants were older adolescents ($M_{age} = 16.07$, $SD_{age} = 0.99$) ranging from 15 to 19 years old (35% were 15 years old, 30.7% were 16 years old, 26% were 17 years old, 7.1% were 18 years old, and 0.7% were 19 years old). The sample was ethnically diverse: 42.2% identified as Asian/Asian-American,

[1] Analyses were not pre-registered.

[2] Other waves of data were not considered as they were administered after experimental manipulations that could affect mindset.

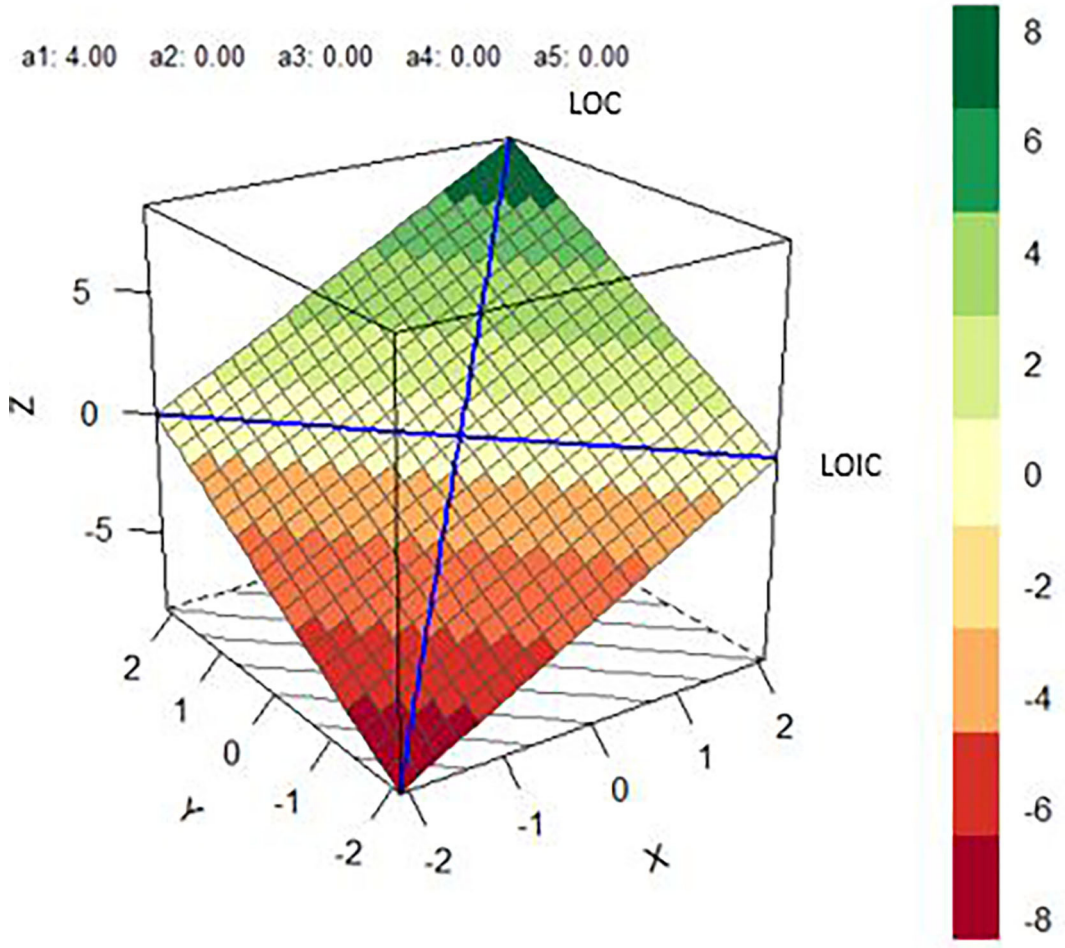

FIGURE 1 | Example response surface analysis.

29.3% Hispanic, 12.7% Caucasian, 4.8% African American, 0.2% Native American, and 2.4% other. Self-reported socioeconomic status also differed: 10.4% identified as "very poor or poor," 32.2% "lower middle-class," 43.8% "middle-class," and 13.5% "upper middle-class or rich." Of the participants, 72.5% identified as recently engaging in a sport or competitive activity whereas 27.5% had not. Further demographic information collected as part of the study is available in the **Supplementary Table 1**.

Procedure

After receiving approval from the Institutional Review Board, we obtained informed consent from each participant. If the participant was below 18 years of age, we obtained informed consent and assent from their parent or guardian as well, taking into consideration different languages that were spoken by participants' families. Consent and assent forms were available in English, Spanish, Mandarin, and Vietnamese, the primary languages spoken within our sample's population. Participants were emailed a link to the survey via Qualtrics, an online survey platform. The survey took about 50–75 min to complete. Upon completion, participants were thanked and compensated with $14.00 for their time. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study in the **Supplementary Material**. The data that support the findings of this study are openly available in OSF at https://osf.io/3pyuh/.

Measures

All measures have been previously validated in adolescent and young adult samples.

Mindset

Morality and ability mindsets were measured using the six-item Implicit Theories of Morality and Intelligence scale (Dweck et al., 1995) which included two subscales: morality (e.g., "A person's moral character is something very basic about them and it can't be changed much.") and ability (e.g., "You can learn new things, but you can't really change your basic talent."). The ability subscale was adapted from the original intelligence subscale; researchers substituted the term "intelligence" with "talent" following Dweck (1996) suggestion to adapt items to

specific context. Item responses ranged from 1 (*strongly disagree*) to 6 (*strongly agree*). Most research anchors these items in the opposite direction with 1 representing *strongly agree* and 6 representing *strongly disagree*. However, we chose to do the inverse to keep scale anchoring consistent across all scales administered in the study (i.e., higher scores representing greater agreement with item) because previous surveys with similar participants in the area suggested our adolescent did not always notice the change in anchors. Higher scores indicated greater fixed mindset.

Self-System and Personality

Personality

Personality was assessed with the Ten-Item Personality Inventory (TIPI; Gosling et al., 2003), which consisted of five subscales: extraversion (e.g., "extraverted, enthusiastic"), agreeableness (e.g., "sympathetic, warm"), conscientiousness (e.g., "dependable, self-disciplined"), emotional stability (e.g., "calm, emotionally stable"), and openness to experiences (e.g., "open to new experiences, complex"). Item responses ranged from 1 (*disagree strongly*) to 7 (*agree strongly*).

Contingencies of Self-Worth

Self-systems contingencies were assessed with 10 items of the Contingencies of Self-Worth Scale (Crocker et al., 2003) which included 2 subscales: competition (e.g., "I feel worthwhile when I perform better than others on a task or skill.") and virtue (e.g., "Doing something I know is wrong makes me lose my self-respect."). Responses ranged from 1 (*strongly disagree*) to 7 (*strongly agree*). Higher scores on each subscale suggested higher levels of contingent self-worth.

Self-Efficacy

Self-efficacy was assessed using 10 items from the General Self-Efficacy Scale (Schwarzer and Jerusalem, 1995), which included items like, "It is easy for me to stick to my aims and accomplish my goals." Responses ranged from 1 (*not at all true*) to 4 (*exactly true*), with higher scores indicating higher self-efficacy.

Self-Regulation

Self-Control

Participants were administered the Brief Self-Control Scale (Tangney et al., 2004), a 13-item measure consisting of items such as "I have a hard time breaking bad habits." Responses ranged from 1 (*not at all*) to 5 (*very much*), so that higher scores indicated lower levels of self-control.

Regulatory Behaviors

Regulatory behaviors were measured using the General Regulatory Behavior Questionnaire (Oaten and Cheng, 2006), a 16-item scale that included items like, "how often did you … use social media when you were not supposed to." Item responses ranged from 1 (*not at all*) to 5 (*almost always*), so higher scores indicated higher dysregulation of behaviors.

Well-Being

Anxiety

Anxiety was assessed with the seven-item Generalized Anxiety Disorder scale (Spitzer et al., 2006), which consisted of items asking people the extent to which they felt certain symptoms, such as "Feeling nervous, anxious, or on edge." Item responses ranged from 0 (*not at all*) to 3 (*nearly every day*), with higher scores indicating higher feelings of general anxiety.

Fear of Failure

Fear of failure was assessed with the five-item short form of the Performance Failure Appraisal Inventory (Conroy, 2001), which included items like "When I am not succeeding, people are less interested in me." Item responses ranged from −2 (*do not believe at all*) to +2 (*believe 100% of the time*).

Depression

Depression was assessed with 10 items from the Center for Epidemiologic Studies Short Depression Scale (Björgvinsson et al., 2013), which consisted of items like, "I felt depressed." Response options ranged from 0 (*rarely or none of the time*) to 3 (*most or all of the time*).

RESULTS

Correlations and measure reliabilities are displayed in **Table 1**.

Power

For response surface analysis (RSA), sample size is determined by having at least 2–3 times as many participants as would be needed to detect linear main effects (Aiken and West, 1991; Humberg et al., 2019). A priori calculations from G*Power indicated that to test linear multiple regression with two predictors and a single outcome with a small effect ($f^2 = 0.02$), a total sample of 311 participants was required. We chose a small effect as there are no extant studies or established effect sizes for the analyses being completed in the present research. Upon doubling this linear main effect for RSA, a total number of 622 participants was required. We were a few participants shy of this number.

Mindset Discrepancies and Functioning

Moral and ability fixed mindsets were congruent for 34.6% ($|\Delta z < 0.5|$) of the adolescent participants; moral fixed mindsets were higher than ability fixed mindsets for 28.8% of participants; and ability fixed mindsets were higher than moral fixed mindsets for 36.6% of participants.

Mindset Confirmatory Factor Analyses

To further support that these mindset domains are distinct but associated constructs, we ran confirmatory factor analyses (CFA) in *Mplus 8.4* (Muthén and Muthén, 1998–2017) to examine the structure of mindset in our sample. Model fit for mindset was good, $\chi^2\,(8) = 24.89$, $p < 0.01$, CFI = 0.99, TLI = 0.98, RMSEA = 0.06, and SRMR = 0.03. All items loaded highly and significantly onto their respective latent factors and can be viewed in **Figure 2**. Moral and ability mindsets were moderately and significantly associated (standardized $r = 0.43$), lending support to these domains being related but distinct constructs.

TABLE 1 | Omegas, means, standard deviations, and correlations among study variables.

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
1. Moral fixed mindset															
2. Ability fixed mindset	0.35***														
3. Extraversion	0.12**	0.08													
4. Agreeableness	0.11*	0.15***	0.00												
5. Conscientiousness	0.04	−0.00	0.17***	0.15***											
6. Emotional stability	0.05	0.07	0.17***	0.19***	0.29***										
7. Openness to experiences	0.11*	0.06	0.31***	0.14**	0.27***	0.10*									
8. CSW competition	0.01	0.02	0.00	−0.19***	−0.08	−0.05	−0.02								
9. CSW virtue	0.04	0.05	0.09*	0.10*	0.15***	0.05	0.13**	0.54***							
10. Self–efficacy	0.12**	0.04	0.26***	0.10*	0.48***	0.39***	0.39***	0.09*	0.25***						
11. Self–control	−0.06	−0.02	−0.10*	−0.20***	−0.60***	−0.36***	−0.16***	0.16***	−0.11*	−0.43***					
12. Regulatory behavior	−0.10*	−0.05	−0.05	−0.19***	−0.46***	−0.33***	−0.14***	0.15***	−0.08	−0.28***	0.65***				
13. Anxiety	−0.10*	−0.11*	−0.12**	−0.12**	−0.28***	−0.57***	0.00	0.09	0.03	−0.29***	0.38***	0.34***			
14. Fear of failure	−0.10*	−0.08	−0.12**	−0.06	−0.17***	−0.30***	−0.10*	0.33***	0.25***	−0.24***	0.28***	0.26***	0.26***		
15. Depression	−0.04	−0.10*	−0.23***	−0.10*	−0.31***	−0.54***	−0.11**	0.08	−0.07	−0.30***	0.39***	0.39***	0.38***	0.37***	
Mean	3.70	3.88	4.03	4.54	4.66	4.34	5.07	5.07	4.81	2.98	3.03	2.56	1.99	3.40	1.88
Standard deviation	1.06	1.35	1.43	0.99	1.28	1.37	1.17	1.08	0.96	0.44	0.67	0.51	0.67	0.91	0.54
Omega reliability	0.81	0.87	–	–	–	–	–	0.81	0.96	0.85	0.83	0.74	0.84	0.80	0.81

*Extraversion, Agreeableness, Emotional Stability, and Openness are measured using two-item scales, and therefore omega reliability could not be calculated. CSW = contingencies of self-worth. *p < 0.05, **p < 0.01, ***p < 0.001.*

Congruency Effects Analysis

To determine the role of moral and ability mindsets in predicting the outcomes of interest (Z), hierarchical regressions were conducted before full response surface analyses. Doing so identified if there are any additive or multiplicative effects prior to congruency analyses. As can be viewed in the first two rows for each outcome variable in **Table 2** there were only a few significant effects of moral and ability mindset. For extraversion, openness to experiences, and self-efficacy, moral fixed mindset alone was a positive additive predictor such that when moral fixed mindset increased, so too did extraversion, openness to experiences, and self-efficacy. Ability fixed mindset did not predict any outcomes in the additive model. There were no significant effects for moral mindset multiplied with ability mindset or for ability mindset squared. Moral fixed mindset squared negatively predicted competition self-worth, such that when moral fixed mindset squared increased, competition self-worth decreased.

Although it is common to stop RSA analyses when b_3-b_5 are non-significant, we still calculated full RSA parameters for all variables, which are displayed in the third row for each outcome variable. None of the RSAs for the outcomes met the requirements to evince a congruence effect. However, coefficients a_1, a_2, and a_4 suggest some potential lines of inquiry for investigators, so we report them. Although interpreted in congruence language, it is important for readers to recognize that there were not overall effects for congruence supported. For both extraversion and agreeableness, the significance of a_1 indicated an effect on the slope of the LOC, suggesting that extraversion and agreeableness were higher when moral and ability fixed mindsets matched at higher levels. For competition contingent self-worth, the significance of a_4 suggested a curvilinear effect of the LOIC. That is, competition contingent self-worth was higher when moral and ability fixed mindsets matched than when they mismatched. For self-control and general regulatory behaviors, the significance of indicated a curvilinear effect for LOC. That is, both self-dysregulation and general dysregulation were higher when moral and ability fixed mindsets were more congruent at midrange levels than when they matched at extreme levels. For both anxiety and fear of failure, the significance of a_1 indicated an effect on the slope of the LOC. Both anxiety and fear of failure were higher when moral and ability fixed mindsets matched at lower levels. However, for all these effects, it must be reiterated that an overall congruency effect was not supported.

DISCUSSION

The present study contributes to the larger literature in several compelling ways. Interestingly, *H1* was not supported; there were no overall congruency effects for any of the models (Humberg et al., 2019). This indicates that overall congruency between moral and ability mindsets did not correlate with better or worse outcomes. Additionally, the data suggest mindsets are characteristic adaptations because of their high levels of domain specificity; confirmatory factor analysis supported distinctions between moral and ability mindset, and two-thirds of the participants demonstrated discrepancies in mindset domains, which supports *H2*. Although we could not fully examine

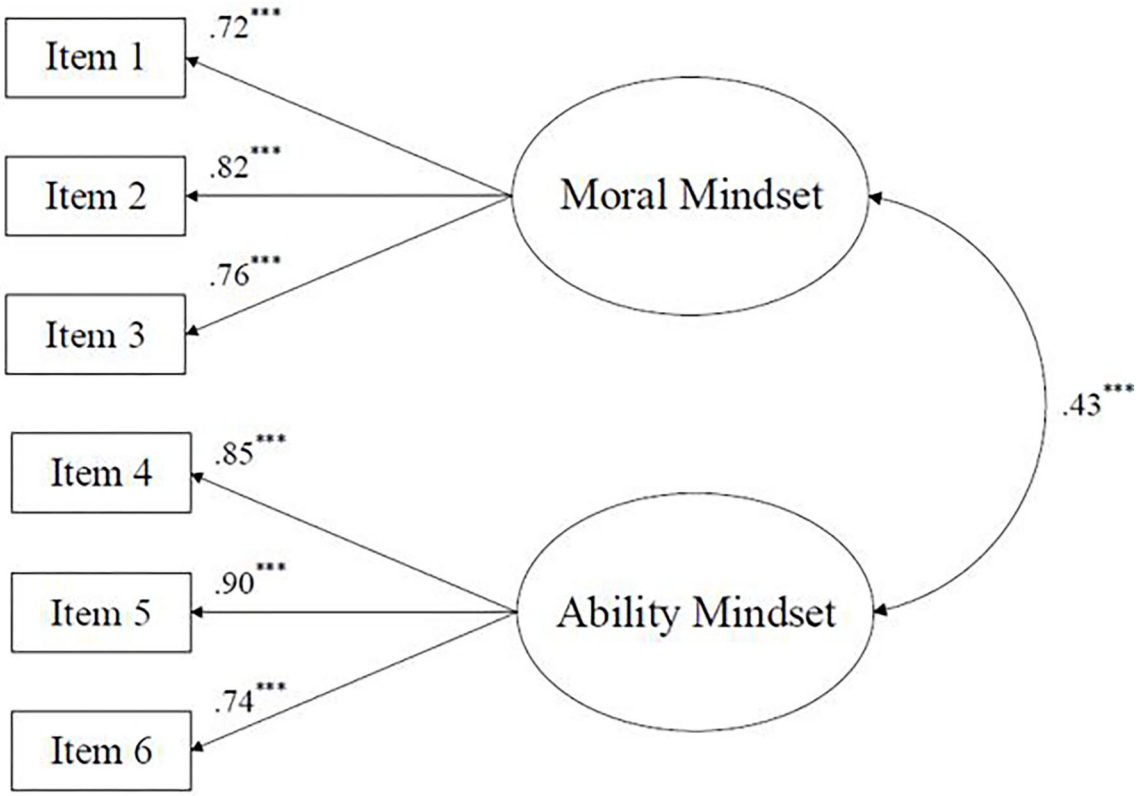

FIGURE 2 | Confirmatory factor analysis of mindset. $N = 599$. Standardized values used. Model fit, χ^2 (8) = 24.89, $p < 0.01$, CFI = 0.99, TLI = 0.98, RMSEA = 0.06, and SRMR = 0.03. ***$p < 0.001$.

H3 because of a lack of congruency effects, examinations of coefficients related to the line of congruence suggest important potential avenues of future research.

In our sample relations among mindset domains and outcomes of interest were lower than expected or found previously. For example, most literature finds a positive association between depression and fixed mindset (e.g., Schleider and Weisz, 2018); however, in our study depression was negatively associated with ability fixed mindset ($r = -0.10$) and not significantly associated with moral fixed mindset at all. Given that linear effects can mask non-linear effects, we still ran the analyses proposed in our hypotheses (Dawson, 2014).

Hypothesis 1: Congruency Between Mindset Domains Bears No Effect on Well-Being Outcomes

Contrary to *H1*, congruency between moral and ability mindsets did not significantly affect any of the well-being outcomes. The null results for congruence contrast with other literature related to characteristic adaptation constructs suggesting that (in)congruence between goals matters (Sheldon and Elliot, 1999; Gray et al., 2017). Indeed, much of extant literature on goals suggest that conflict between domains inhibits well-being and congruence enhances well-being. This finding was not reflected in the present study. Instead, it is most likely that ability and moral mindsets are distinct qualities for which congruence is not of import for well-being.

Beyond the simple explanation that congruence among domain-specific mindsets is not relevant for well-being, there are several possible explanations for the lack of an overall congruence effect. It may be possible that congruence is not the best way to model effects across mindset domains as it does not resemble congruency found between implicit and explicit expressions of the same constructs such as self-esteem (Lupien et al., 2010) or motives (Thielgen et al., 2015). In such cases, both the explicit and implicit construct reflect an underlying dimension, which does not seem to be the same case for mindset. Additionally, recent literature on other characteristic adaptation constructs (e.g., character strengths) has shifted away from attempting to assess consistency across contexts and instead focused toward assessing coherence, wherein a person demonstrates appropriate character mechanisms in varying amounts and in varying contexts (Lerner, 2019; Nucci, 2019). Perhaps mindset, like character, should be considered from a coherence framing rather than consistency framing.

It is also possible that the adolescents in the study may still be engaged in the differentiation stage of self-system development and may not have moved into the integration stage (Werner, 1957). Werner (1957) orthogenetic principle suggests that as development occurs, there is a process from a state of globality

TABLE 2 | Moral vs. ability mindsets across outcome variables.

	Estimated regression model						Position of 1st Principal Axis (PA)			LOC		LOIC	
Outcome	b_0 **Intercept**	b_1 **Moral**	b_2 **Ability**	b_3 **Moral^2**	b_4 **Moral*Ability**	b_5 **Ability^2**	p_{10} **Intercept**	p_{11} **Slope of 1 PA**	R^2	***a1*** **Slope of LOC**	***a2*** **Curvilinear Effect of LOC**	***a3*** **Slope of LOC**	***a4*** **Curvilinear Effect of LOIC**
Extraversion	3.34***	0.10*	0.05						0.02*				
	3.28***	0.09	0.10	−0.00	0.01	−0.05			0.02				
	4.00***	0.13*	0.06	0.00	0.00	−0.01	4.23	−0.01	0.02	0.19**	−0.01	0.07	−0.01
Agreeableness	3.96***	0.06	0.13						0.03***				
	3.95***	0.10	0.10	0.02	−0.05	0.03			0.03*				
	4.50***	0.06	0.10*	−0.01	0.00	0.00	−29.14	7.63	0.03	0.15**	−0.00	−0.04	−0.01
Conscientiousness	4.51***	0.05	−0.02						0.00				
	5.19***	−0.26	−0.01	0.43	0.07	−0.29			0.01				
	4.67***	0.03	−0.03	0.01	0.07	−0.04	−0.16	0.55	0.01	0.03	0.05	0.03	−0.10
Emotional Stability	3.95***	0.03	0.06						0.01				
	4.42***	−0.43	0.37	−0.04	0.48	−0.28			0.01				
	4.28***	0.01	0.09	0.08	−0.01	−0.04	1.21	−0.03	0.01	0.10	0.04	−0.08	0.06
Openness to Experiences	4.58***	0.10*	0.03						0.01*				
	5.27***	−0.03	−0.30	−0.01	0.14	0.34			0.02				
	4.95***	0.11	0.01	0.02	−0.00	0.04	−41.81	−16.36	0.02	0.11	0.06	0.10	0.06
Competition CSW	5.01***	0.01	0.01						0.00				
	5.06***	−0.02	−0.07	−0.62**	0.38	0.49			0.02				
	4.96***	0.03	−0.01	0.05	−0.09*	0.05	−1.31	−1.00	0.02	0.02	0.01	0.04	0.19**
Virtue CSW	4.60***	0.03	0.05						0.00				
	4.69***	0.14	−0.22	−0.18	−0.02	0.39			0.01				
	4.75***	0.04	0.01	−0.00	−0.02	0.04	14.12	−3.48	0.01	0.05	0.01	0.03	0.06
Self–Efficacy	2.79***	0.12*	0.00						0.01*				
	3.10***	−0.06	−0.33	0.07	0.15	0.29			0.02				
	2.93***	0.05*	−0.01	0.01	0.00	0.01	7.90	2.34	0.02	0.04	0.03	0.05	0.02
Self–Control	3.18***	−0.06	−0.00						0.00				
	2.57***	0.22	0.35	−0.40	−0.06	−0.10			0.02				
	3.07***	−0.03	0.01	−0.01	−0.04	−0.01	−0.46	−0.97	0.02	−0.02	−0.05*	−0.03	0.02
Regulatory Behavior	2.75***	−0.09	−0.02						0.01				
	2.26***	0.23	0.31	−0.41	−0.11	−0.07			0.02*				
	2.60***	−0.04	−0.00	−0.01	−0.03	−0.00	−1.11	−1.13	0.02	−0.04	−0.04*	−0.03	0.02
Anxiety	2.32***	−0.08	−0.08						0.02*				
	2.35***	−0.17	−0.01	−0.15	0.18	0.03			0.02				
	2.00***	−0.05	−0.04	0.02	−0.01	0.00	−30.01	−0.40	0.02	−0.09**	0.00	−0.01	0.03
Fear of Failure	3.79***	−0.08	−0.05						0.01*				
	3.59***	−0.21	0.29	−0.07	0.16	−0.30			0.02				
	3.45***	−0.08	−0.02	0.02	−0.01	−0.03	−0.47	−0.09	0.02	−0.10*	−0.02	−0.06	0.00
Depression	2.05***	−0.01	−0.09						0.01				
	1.88***	0.20	−0.14	−0.33	−0.02	0.27			0.01				
	1.89***	0.01	−0.04	−0.00	−0.02	0.01	−2.17	−1.83	0.01	−0.04	−0.01	0.05	0.04

*CSW, Contingencies of Self-Worth. *$p < 0.05$ **$p < 0.01$ *** $p < 0.001$.*

and lack of differentiation to a state of increasing differentiation, articulation, and integration. Returning again to the example of character, researchers have found that global character becomes more differentiated and specific as children become adolescents (Shubert et al., 2019). Mindset may follow a similar trajectory, the domains becoming more differentiated with age before later integration. Domain incongruence could be further explored through comparing the centrality of the various domains to the specific person as well as the hierarchic integration of the domains (Werner, 1957). Much of the recent literature exploring integration focuses on emerging adults rather than adolescents, perhaps, because it is more normative for integration to take place in emerging adulthood (Schwartz et al., 2015; Sumner et al., 2015). Future studies should examine samples spanning from early adolescence into adulthood to examine age differences in congruence effects.

Another potential explanation is that the measures of each mindset domain tap into other constructs besides pure mindset. The items involved in assessing moral mindset (e.g., "A person's moral character is something very basic about them and it can't be changed much") could tap into the broader construct of moral identity. Moral identity serves to pair moral motivation and reasoning with actual behavior (Hardy and Carlo, 2011). Given that growth mindset has consistently been linked with increased motivation (Yeager and Dweck, 2012; Han et al., 2020), moral mindset may assess components of moral identity. It is apparent that the moral and ability mindsets tap into specific domains.

Hypothesis 2: Mindsets Within the Personality System Are Largely Domain Specific, Suggesting They Are Characteristic Adaptations

Lending support to *H2*, 65.4% of participants demonstrated discrepancies in ability and moral mindsets, suggesting a high level of domain specificity. It is important to note that there was a modest correlation between ability and moral mindsets ($r = 0.35$ for basic correlation; $r = 0.43$ for CFA); however, despite their association, these domains were distinct from each other for nearly two-thirds of the participants. Returning to McAdams and Pals's (2006) conceptualization of characteristic adaptations, a key aspect of these personality units is that they are "contextualized in time, place, and/or social role" (p. 208). This contextualization is the central separation between characteristic adaptations and dispositional traits, in that dispositional traits are typically broad and decontextualized.

Situating mindsets as characteristic adaptations is further supported by their relations with the Big Five traits in this study. The mindsets were clearly distinct from the Big Five traits, echoing previous literature (Burnette and Pollack, 2013). Both moral and ability fixed mindsets were positively correlated with extraversion, agreeableness, and openness to experiences, but effects were small indicating distinctiveness. Moreover, the lack of congruence effect for personality traits further supports their differentiation at a different level of personality.

Null Results Related to Hypothesis 3, but Interesting Exploratory Findings

Given that there was no overall congruence effect, we cannot make any actual conclusions about how higher or lower fixed mindsets accompanying congruence relate to well-being to address *H3*. However, significant effects related to the line of congruence suggest interesting points of discussion that may be useful for spurring future research. These findings should be interpreted with extreme caution given their exploratory nature.

Exploratory Findings Suggest Congruencies With Lower Fixed Mindset Are Sometimes Associated With Lower Well-Being

Although there were no overall congruence effects, results suggest both anxiety and fear of failure, indicators of well-being, are higher when participants' mindsets are congruent in the direction of a higher growth mindset. This finding means that possessing a growth mindset across multiple mindset domains is associated with *higher* arousal and reactivity. This may be because people with more fixed mindsets are more likely to make external attributions regarding ability and failure (Bodill and Roberts, 2013), which could relieve anxiety and fear of failure. Attributions regarding potential failure would be external and, thus, would not be the responsibility of the participant. In essence, those with fixed mindsets may believe their performance, good or poor, is no longer a result of their own innate ability, which alleviates anxiety. In a similar manner, *H3* was not supported with the personality indicators. Higher levels of extraversion and agreeableness were associated with more fixed mindsets than growth mindsets. Given the unexpected nature of this finding, further exploration and consideration are necessary.

Exploratory Findings Suggest Moderate Mindsets May Be Detrimental

Results from the self-dysregulation variable suggest that congruency in mindset at either extreme—either fixed or growth—rather than midrange levels predicted greater capacity to self-regulate. In other words, participants who were congruent at more extreme ends of the poles experienced better outcomes than those who were congruent at mid-range levels for regulation[3]. This finding suggesting a potential cost of a "neutral" mindset may have serious implications for intervention research. Interventions aimed toward shifting a participant from a more fixed mindset to a more growth mindset might leave that participant in a sort of middle ground. In light of the findings in the present study, such a move could potentially be more harmful than staying at a fixed mindset. Our exploratory findings, if confirmed, would suggest future intervention work should carefully ensure that participants will not be left in a potentially worse state than the beginning of the intervention, perhaps by

[3]Indeed, there seems an implicit recognition of the problems of a neutral mindset in mindset measurement, which tends to assess mindset using a 6-point Likert system from 1 (*strongly agree*) to 6 (*strongly disagree*). Several researchers suggest that this 6-point anchoring system is used to avoid a "neutral" mindset; participants will always lean at least slightly toward growth or fixed mindsets (Burnette and Pollack, 2013; Cainëls et al., 2017). This measurement strategy may have some bearing, given the findings with self-dysregulation.

ensuring that the intervention fully moves participants into a growth mindset state. However, these findings are purely exploratory in nature because there was no overall congruence effect, but they merit discussion to spur future inquiry.

Constraints on Generality

The present study involved a demographically diverse subject pool recruited from high schools in metropolitan southern California, meaning replicability of the findings will likely depend on similarity to the sample. A high proportion of participants were Asian and Latinx (70.6% of sample), which are typically culturally interdependent. Some research suggests that mindsets act universally across cultures (e.g., Church et al., 2003); however, some findings indicate that individualistic groups are more susceptible to fixed mindsets (e.g., Church et al., 2012), and collectivistic groups engage higher growth mindsets (e.g., Tang et al., 2016). There is no research on the cultural or ethnic differences of mindset congruence across multiple domains, which makes it difficult to make any predictions about generalizability of findings in a different sample. Most participants were student athletes (71.9%). Because characteristic adaptations depend largely on the centrality and relevance to one's personal identity (Sheldon and Elliot, 1999), the variables related to performance (e.g., competition contingency self-worth) may not be intrinsically important to people outside competitive contexts. Analyses were run (available in **Supplementary Material**) on differences in mindset based on race, culture, and athletic status, which were non-significant. The analysis comparing typically collectivistic cultures with individualistic cultures approached significance ($p = 0.052$) for ability mindset, such that participants from typically collectivistic cultures displayed lower fixed mindsets, in line with previous literature (Church et al., 2012; Tang et al., 2016). As previously discussed, there is good reason to expect we might find different results in adult samples. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

Limitations and Future Directions

The present findings should be interpreted in consideration of several limitations. First, the study was slightly underpowered. Future research utilizing response surface analyses on mindset domain congruency should utilize a larger sample size in order to best detect effects that exist in the data. Secondly, the present findings consider the relations among constructs only at a single time point. Future research should consider these associations across time to investigate the bidirectional relations of characteristic adaptations and dispositional traits (McAdams and Pals, 2006; McCrae et al., 2008).

Finally, congruencies and discrepancies between mindsets were limited to two specific domains: morality and ability. Ability mindset could arguably cover a broader range of specific ability-akin domains like intelligence or athletic ability. Moreover, there is little research on moral mindset, especially in relation to our outcomes of interest. Because mindsets exist within and across a vast range of domains, further research is needed to investigate congruency patterns among other specific domains of mindset (e.g., intelligence, personality; Dweck and Yeager, 2019).

CONCLUSION

Consistent with Dweck and Yeager (2019) conceptualization of mindset, these results provide empirical support that mindsets are domain specific and should be categorized at the characteristic adaptation level of personality. Contrary to hypotheses based on self-concordance theory (Sheldon and Elliot, 1999) and research on goal conflict (Gray et al., 2017), we did not find that congruence in mindsets across domains affects well-being in our sample. These findings suggests that congruence matters for the content of the developing self-system across domains but not necessarily for mindset processes of the self-system among adolescents. Within the purview of the present study, we failed to reject the null hypothesis. Future research is needed to test whether the null effects for mindset congruence hold in another sample of adolescents. Should the non-significant findings hold true, one could explore whether this is true only for those still navigating differentiation and integration of the self-system (Werner, 1957) or whether it would generalize into adulthood.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Open Science Framework https://osf.io/3pyuh/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Fuller Theological Seminary's Institutional Review Board. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

JR contributed to conceptualization, data curation, data analyses, and writing the manuscript. EW contributed to writing the manuscript. LB contributed to conceptualization, data analyses, and data curation. BH contributed to conceptualization and funding acquisition. SS contributed to conceptualization, editing the manuscript, funding acquisition, and overseeing the project. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the John Templeton Foundation [51914, 2015-17 and 61221, 2018-21].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.701510/full#supplementary-material

REFERENCES

Aiken, L. A., and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Sage.

Bernecker, K., and Job, V. (2017). "Implicit theories about willpower and their implications for health and well-being," in *Routledge International Handbook of Self-Control in Health and Well-Being*, eds D. de Ridder, M. Adriaanse, and K. Fujita (London: Routledge), 143–155. doi: 10.4324/9781315648576-12

Björgvinsson, T., Kertz, S. J., Bigda-Peyton, J. S., McCoy, K. L., and Aderka, I. M. (2013). Psychometric properties of the CES-D-10 in a psychiatric sample. *Assessment* 20, 429–436. doi: 10.1177/1073191113481998

Bodill, K., and Roberts, L. D. (2013). Implicit theories of intelligence and academic locus of control as predictors of studying behaviour. *Learn. Individ. Differ.* 27, 163–166. doi: 10.1016/j.lindif.2013.08.001

Brandstätter, V., Job, V., and Schulze, B. (2016). Motivational incongruence and well-being at the workplace: person-job fit, job burnout, and physical symptoms. *Front. Psychol.* 7:1153. doi: 10.3389/fpsyg.2016.01153

Burnette, J. L., Knouse, L. E., Vavra, D. T., O'Boyle, E., and Brooks, M. A. (2020). Growth mindsets and psychological distress: a meta-analysis. *Clin. Psychol. Rev.* 77:101816. doi: 10.1016/j.cpr.2020.101816

Burnette, J. L., and Pollack, J. M. (2013). Implicit theories of work and job fit: implications for job and life satisfaction. *Basic Appl. Soc. Psych.* 35, 360–372. doi: 10.1080/01973533.2013.803964

Cainëls, M. C. J., Semeijn, J. H., and Renders, I. H. M. (2017). Mind the mindset! The interaction of proactive personality, transformational leadership and growth mindset for engagement at work. *Career Dev. Int.* 23, 48–66. doi: 10.1108/CDI-11-2016-0194

Chiu, C.-y., Dweck, C. S., Tong, J. Y.-y., and Fu, J. H.-y. (1997). Implicit theories and conceptions of morality. *J. Pers. Soc. Psychol.* 73, 923–940. doi: 10.1037/0022-3514.73.5.923

Church, A. T., Ortiz, F. A., Katigbak, M. S., Avdeyeva, T. V., Emerson, A. M., Vargas Flores, J., et al. (2003). Measuring individual and cultural differences in implicit trait theories. *J. Pers. Soc. Psychol.* 85, 332–347. doi: 10.1037/0022-3514.85.2.332

Church, A. T., Willmore, S. L., Anderson, A. T., Ochiai, M., Porter, N., Mateo, N. J., et al. (2012). Cultural differences in implicit theories and self-perceptions of traitedness: replication and extension with alternative measurement formats and cultural dimensions. *J. Cross Cult. Psychol.* 43, 1268–1296. doi: 10.1177/0022022111428514

Conroy, D. E. (2001). Progress in the development of a multidimensional measure of fear of failure: the Performance Failure Appraisal Inventory (PFAI). *Anxiety Stress Coping* 14, 431–452. doi: 10.1080/10615800108248365

Crocker, J., Luhtanen, R., Cooper, M., Bouvrette, A., and Crocker, J. (2003). Contingencies of self-worth in college students: theory and measurement. *J. Pers. Soc. Psychol.* 85, 894–908. doi: 10.1037/0022-3514.85.5.894

Dawson, J. F. (2014). Moderation in management research: what, why, when, and how. *J. Bus. Psychol.* 29, 1–19. doi: 10.1007/s10869-013-9308-7

Diseth, Å., Meland, E., and Breidablik, H. J. (2014). Self-beliefs among students: grade level and gender differences in self-esteem, self-efficacy and implicit theories of intelligence. *Learn. Individ. Differ.* 35, 1–8. doi: 10.1016/j.lindif.2014.06.003

Dupeyrat, C., and Mariné, C. (2005). Implicit theories of intelligence, goal orientation, cognitive engagement, and achievement: a test of Dweck's model with returning to school adults. *Contemp. Educ. Psychol.* 30, 43–59. doi: 10.1016/j.cedpsych.2004.01.007

Dweck, C. S. (1996). Capturing the dynamic nature of personality. *J. Res. Pers.* 30, 348–362. doi: 10.1006/jrpe.1996.0024

Dweck, C. S., Chiu, C., and Hong, Y. (1995). Implicit theories and their role in judgments and reactions: a word from two Perspectives. *Psychol. Inq.* 6:267. doi: 10.1207/s15327965pli0604_1

Dweck, C. S., and Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychol. Rev.* 95, 256–273. doi: 10.1037/0033-295X.95.2.256

Dweck, C. S., and Yeager, D. S. (2019). Mindsets: a view from two eras. *Perspect. Psychol. Sci.* 14, 481–496. doi: 10.1177/1745691618804166

Emmons, R. A., King, L. A., and Sheldon, K. (1993). "Goal conflict and the self-regulation of action," in *Century Psychology Series. Handbook of Mental Control*, eds D. M. Wegner and J. W. Pennebakerp (Upper Saddle River, NJ: Prentice-Hall, Inc), 528–551.

Furr, R. M. (2009). Personality psychology as a truly behavioural science. *Eur. J. Pers.* 23, 369–401. doi: 10.1002/per.724

Gosling, S. D., Rentfrow, P. J., and Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *J. Res. Pers.* 37, 504–528. doi: 10.1016/S0092-6566(03)00046-1

Goth, K., Foelsch, P., Schlüter-Müller, S., Birkhölzer, M., Jung, E., Pick, O., et al. (2012). Assessment of identity development and identity diffusion in adolescence-Theoretical basis and psychometric properties of the self-report questionnaire AIDA. *Child Adolesc. Psychiatry Ment. Health* 6:27. doi: 10.1186/1753-2000-6-27

Gray, J. S., Ozer, D. J., and Rosenthal, R. (2017). Goal conflict and psychological well-being: a meta-analysis. *J. Res. Pers.* 66, 27–37. doi: 10.1016/j.jrp.2016.12.003

Han, H., Choi, Y., Dawson, K., and Jeong, C. (2018). Moral growth mindset is associated with change in voluntary service engagement. *PLoS ONE* 13:e0202327. doi: 10.1371/journal.pone.0202327

Han, H., Dawson, K. J., Choi, Y. R., Choi, Y. J., and Glenn, A. L. (2020). Development and validation of the English version of the Moral Growth Mindset measure. *F1000Research* 9:256. doi: 10.12688/f1000research.23160.2

Hardy, S. A., and Carlo, G. (2011). "Moral identity," in *Handbook of Identity Theory and Research*, eds S. J. Schwartz, K. Luyckx, and V. L. Vignoles (Springer, New York, NY), 495–513. doi: 10.1007/978-1-4419-7988-9_19

Howell, A. J., Passmore, H. A., and Holder, M. D. (2016). Implicit theories of well-being predict well-being and the endorsement of therapeutic lifestyle changes. *J. Happiness Stud.* 17, 2347–2363. doi: 10.1007/s10902-015-9697-6

Humberg, S., Nestler, S., and Back, M. D. (2019). Response surface analysis in personality and social psychology: checklist and clarifications for the case of congruence hypotheses. *Soc. Psychol. Personal. Sci.* 10, 409–419. doi: 10.1177/1948550618757600

King, R. B. (2016). A fixed mindset leads to negative affect. *Zeitschrift für Psychol.* 225, 137–145. doi: 10.1027/2151-2604/a000290

Komarraju, M., and Nadler, D. (2013). Self-efficacy and academic achievement: why do implicit beliefs, goals, and effort regulation matter? *Learn. Individ. Differ.* 25:67. doi: 10.1016/j.lindif.2013.01.005

Konze, A. K., Rivkin, W., and Schmidt, K. H. (2019). Can faith move mountains? How implicit theories about willpower moderate the adverse effect of daily emotional dissonance on ego-depletion at work and its spillover to the home-domain. *Eur. J. Work Org. Psychol.* 28, 137–149. doi: 10.1080/1359432X.2018.1560269

Lerner, R. M. (2019). Character development: four facets of virtues. *Child Dev. Perspect.* 13, 79–84. doi: 10.1111/cdep.12315

Lewis, L. S., Williams, C. A., and Dawson, S. D. (2020). Growth mindset training and effective learning strategies in community college registered nursing students. *Teach. Learn. Nurs.* 15, 123–127. doi: 10.1016/j.teln.2020.01.006

Lupien, S. P., Seery, M. D., and Almonte, J. L. (2010). Discrepant and congruent high self-esteem: behavioral self-handicapping as a preemptive defensive strategy. *J. Exp. Soc. Psychol.* 46, 1105–1108.

McAdams, D. P., and Pals, J. L. (2006). A new Big Five: fundamental principles for an integrative science of personality. *Am. Psychol.* 61, 204–217. doi: 10.1037/0003-066X.61.3.204

McCrae, R. R., and Costa, P. T., Jr. (2008). "The five-factor theory of personality," in *Handbook of Personality: Theory and Research*, eds O. P. John, R. W. Robins, and L. A. Pervin (New York, NY: The Guilford Press), 159–181.

Muthén, L. K., and Muthén, B. O. (1998–2017). *Mplus User's Guide, 8th Edn*. Los Angeles, CA: Muthén and Muthén.

Niiya, Y., Brook, A. T., and Crocker, J. (2010). Contingent self-worth and self-handicapping: do incremental theorists protect self-esteem? *Self Identity* 9, 276–297. doi: 10.1080/15298860903054233

Nucci, L. (2019). Character: a developmental system. *Child Dev. Perspect.* 13, 73–78. doi: 10.1111/cdep.12313

Oaten, M., and Cheng, K. (2006). Improved self-control: the benefits of a regular program of academic study. *Basic Appl. Soc. Psychol.* 28, 1–16. doi: 10.1207/s15324834basp2801_1

Sagar, S. S., Lavallee, D., and Spray, C. M. (2007). Why young elite athletes fear failure: consequences of failure. *J. Sports Sci.* 25, 1171–1184. doi: 10.1080/02640410601040093

Satchell, L., Hoskins, S., Corr, P., and Moore, R. (2017). Ruminating on the nature of intelligence: personality predicts implicit theories and educational persistence. *Pers. Individ. Dif.* 113, 109–114. doi: 10.1016/j.paid.2017.03.025

Schleider, J., and Weisz, J. (2018). A single-session growth mindset intervention for adolescent anxiety and depression: 9-month outcomes of a randomized trial. *J. Child Psychol. Psychiatry* 59, 160–170. doi: 10.1111/jcpp.12811

Schönbrodt, F. D., and Humberg, S. (2020). *RSA: An R Package for Response Surface Analysis (version 0.10.1)*. Available online at: https://cran.r-project.org/package=RSA

Schroder, H. S., Callahan, C. P., Gornik, A. E., and Moser, J. S. (2019). The fixed mindset of anxiety predicts future distress: a longitudinal study. *Behav. Ther.* 50, 710–717. doi: 10.1016/j.beth.2018.11.001

Schwartz, S. J., Hardy, S. A., Zamboanga, B. L., Meca, A., Waterman, A. S., Picariello, S., et al. (2015). Identity in young adulthood: links with mental health and risky behavior. *J. Appl. Dev. Psychol.* 36, 39–52. doi: 10.1016/j.appdev.2014.10.001

Schwarzer, R., and Jerusalem, M. (1995). "Generalized self-efficacy scale," in *Measures in Health Psychology: A User's Portfolio,* eds J. Weinman, S. Wright, and M. Johnston (NFER—NELSON), 35–37.

Sheldon, K. M., and Elliot, A. J. (1999). Goal striving, need satisfaction, and longitudinal well-being: the self-concordance model. *J. Pers. Soc. Psychol.* 76, 482–497. doi: 10.1037/0022-3514.76.3.482

Sheldon, K. M., and Kasser, T. (1998). Pursuing personal goals: skills enable progress, but not all progress is beneficial. *Person. Soc. Psychol. Bull.* 24, 1319–1331. doi: 10.1177/01461672982412006

Shubert, J., Wray-Lake, L., Syvertsen, A. K., and Metzger, A. (2019). Examining character structure and function across childhood and adolescence. *Child Dev.* 90, e505–e524. doi: 10.1111/cdev.13035

Spinath, B., Spinath, F. M., Riemann, R., and Angleitner, A. (2003). Implicit theories about personality and intelligence and their relationship to actual personality and intelligence. *Pers. Individ. Dif.* 35, 939–951. doi: 10.1016/S0191-8869(02)00310-0

Spitzer, R. L., Kroenke, K., Williams, J. B. W., and Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Internal Med.* 166, 1092–1097. doi: 10.1001/archinte.166.10.1092

Sumner, R., Burrow, A. L., and Hill, P. L. (2015). Identity and purpose as predictors of subjective well-being in emerging adulthood. *Emerg. Adulthood* 3, 46–54. doi: 10.1177/2167696814532796

Tamir, M., John, O. P., Srivastava, S., and Gross, J. J. (2007). Implicit theories of emotion: affective and social outcomes across a major life transition. *J. Pers. Soc. Psychol.* 92, 731–744. doi: 10.1037/0022-3514.92.4.731

Tang, M., Werner, C., and Karwowski, M. (2016). Differences in creative mindset between Germany and Poland: the mediating effect of individualism and collectivism. *Thinking Skills Creativity* 21, 31–40. doi: 10.1016/j.tsc.2016.05.004

Tangney, J. P., Baumeister, R. F., and Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *J. Pers.* 72, 271–324. doi: 10.1111/j.0022-3506.2004.00263.x

Thielgen, M. M., Krumm, S., Rauschenbach, C., and Hertel, G. (2015). Older but wiser: age moderates congruency effects between implicit and explicit motives on job satisfaction. *Motiv. Emotion* 39, 182–200.

Verhagen, M., Lodder, G. M., and Baumeister, R. F. (2018). Unmet belongingness needs but not high belongingness needs alone predict adverse well-being: a response surface modeling approach. *J. Pers.* 86, 498–507. doi: 10.1111/jopy.12331

Wang, C. K. J., Tan, L., and Dairianathan, E. I. (2018). Achievement goals, implicit theories, and intrinsic motivation: a test of domain specificity across music, visual art, and sports. *J. Res. Music Educ.* 66, 320–337. doi: 10.1177/0022429418784563

Werner, H. (1957). "The concept of development from a comparative organismic view," in *The Concept of Development*, ed D. Harris (Minneapolis, MN: University of Minnesota Press), 125–148.

Yeager, D. S., and Dweck, C. S. (2012). Mindsets that promote resilience: when students believe that personal characteristics can be developed. *Educ. Psychol.* 47, 302–314. doi: 10.1080/00461520.2012.722805

Zeng, G., Hou, H., and Peng, K. (2016). Effect of growth mindset on school engagement and psychological well-being of Chinese primary and middle school students: the mediating role of resilience. *Front. Psychol.* 7:1873. doi: 10.3389/fpsyg.2016.01873

Zhao, S., Du, H., Li, Q., Wu, Q., and Chi, P. (2021). Growth mindset of socioeconomic status boosts subjective well-being: a longitudinal study. *Pers. Individ. Dif.* 168:110301. doi: 10.1016/j.paid.2020.110301

Disclaimer: The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

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

Copyright © 2021 Ratchford, Williams, Bishara, Houltberg and Schnitker. 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.

frontiers in Psychology

SYSTEMATIC REVIEW
published: 23 November 2021
doi: 10.3389/fpsyg.2021.742510

Assessing Momentary Well-Being in People Living With Dementia: A Systematic Review of Observational Instruments

Kristine Gustavsen Madsø[1,2], Elisabeth Flo-Groeneboom[1], Nancy A. Pachana[3] and Inger Hilde Nordhus[1,4]*

[1] Department of Clinical Psychology, Faculty of Psychology, University of Bergen, Bergen, Norway, [2] NKS Olaviken Gerontopsychiatric Hospital, Bergen, Norway, [3] School of Psychology, The University of Queensland, Brisbane, QLD, Australia, [4] Department of Behavioral Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

Optimizing the possibility to lead good lives is at the core of treatment and care for people with dementia. This may be monitored by assessing well-being and quality of life. However, cognitive impairment following dementia may complicate recall-based assessment with questionnaires, and proxy-ratings from family-caregivers do not correspond well to self-reports. Thus, using observational measures represents a potentially advanced option. Systematic reviews evaluating measurement properties, interpretability and feasibility of observational instruments assessing well-being in people living with dementia are lacking. Thus, this review performed systematic searches to find peer reviewed validated instruments of relevance in the databases MEDLINE, EMBASE, PsycINFO, Web of Science, CINAHL and ProQuest. Twenty-two instruments assessing well-being were included for evaluation of measurement properties based on the systematic approach of the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN). The evaluation included risk of bias on study level, and assessment of measurement properties on instrument level including content validity, construct validity, structural validity, internal consistency, measurement invariance, cross-cultural validity, measurement error and inter-rater/intra-rater/test–retest reliability and responsiveness. Additionally, the feasibility and interpretability of the measures were evaluated. No single instrument could be recommended based on existing publications. Thus, we provide general recommendations about further assessment and development of these instruments. Finally, we describe the most promising instruments and offer guidance with respect to their implementation and use in clinical and research contexts.

Keywords: well-being, dementia, observation, emotion, systematic review, psychometric properties, engagement

OPEN ACCESS

Edited by:
Peter ten Klooster, University of Twente, Netherlands

Reviewed by:
Jannis Kraiss, University of Twente, Netherlands
Katherine Algar-Skaife, Norwegian University of Science and Technology, Norway

****Correspondence:***
Kristine G. Madsø
kristine.madso@uib.no

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

Received: *16 July 2021*
Accepted: *26 October 2021*
Published: *23 November 2021*

Citation:
Madsø KG, Flo-Groeneboom E, Pachana NA and Nordhus IH (2021) Assessing Momentary Well-Being in People Living With Dementia: A Systematic Review of Observational Instruments. Front. Psychol. 12:742510. doi: 10.3389/fpsyg.2021.742510

INTRODUCTION

Well-being and quality of life (QoL) are identified as core outcomes for psychosocial interventions by people living with dementia (Øksnebjerg et al., 2018; Reilly et al., 2020), in public health initiatives (World Health Organization [WHO], 2017), national guidelines (National Institute for Health and Care Excellence [NICE], 2018), and research recommendations (Dröes et al., 2016).

World-wide, dementia is estimated to affect 50 million people (Livingston et al., 2017). Dementia is defined as a public health priority, causing disability and increasing dependency on help from others in the people affected (World Health Organization [WHO], 2017). However, increasing evidence highlights how people with dementia may live good lives in environments adapted to their physical, social, emotional, and psychological needs (Livingston et al., 2017).

To be able to ascertain whether the dementia care and interventions implemented actually promote individual well-being, valid measurement approaches reflecting well-being as it is described by the target group are needed (Madsø and Nordhus, 2021). In a recent scoping review, relevant well-being domains close to the experiences of people living with dementia were defined. These domains include positive emotions, experiencing meaning, a positive sense of self and a sense of agency, having rewarding relationships with significant others, and experiencing life satisfaction (Clarke et al., 2020). Well-being and QoL originate from separate research fields (Skevington and Böhnke, 2018), but have also been used synonymously in the dementia literature (Bowling et al., 2015). In this review, the term well-being is used when the domains are in line with Clarke et al. (2020).

In other populations, well-being is often measured by self-report (Ferring and Boll, 2010). It is well established that people in the earlier stages of dementia can provide valid self-reports of their well-being (Stoner et al., 2019; Clarke et al., 2020). Unfortunately, relying on self-report only may exclude people with more severe dementia, and reduce the possibility of longitudinal assessment throughout the degenerative course of the disease (Algar et al., 2016; Kaufmann and Engel, 2016). With increasing cognitive impairment, well-being is frequently assessed through *proxy-reports*. Proxy-reports refer to assessment of an individual based on the evaluations of informants other than the person themself. Studies have consistently found proxy-evaluations by family and professional caregivers to rate well-being lower as compared to self-reports (Sands et al., 2004; Kolanowski et al., 2007; Ferring and Boll, 2010; Schulz et al., 2013). The low correspondence between proxy-reports and self-report implies that well-being in dementia should be measured in face-to-face interviews for individuals able to give valid self-reports, together with observational measures by independent and neutral observers in those from whom self-reports may not be obtained (Ferring and Boll, 2010; Bowling et al., 2015).

It is well known that a measurement that relies on *retrospective* self-reports evaluating longer time-intervals is prone to bias because our autobiographical recall can be inaccurate and influenced by for example current mood (Shiffman et al., 2008). This may particularly be a source of bias in the dementia population due to impairments in memory, attention, insight, and communication skills (Ettema et al., 2007; Trigg et al., 2011). During retrospective self-report, the current emotional state may interfere with the judgment of the past (Kolanowski et al., 2014). Thus, the risk of substantial measurement error from self-reports is increased by the fluctuating nature of neuropsychiatric symptoms (Kales et al., 2015), as well as attention or awareness (Clare et al., 2012). Consequently, an alternative is to use Ecological Momentary Assessment (EMA) and assess well-being within a momentary timeframe that can detect clinically relevant variations occurring over short time intervals (Shiffman et al., 2008). EMA consists of several approaches - direct observation is one of them. Assessing well-being in dementia through observing behavior as it occurs is one approach that can omit several of the problems and sources of bias related to measurement in dementia as mentioned above (Ferring and Boll, 2010). In sum, observational methods are advantageous because (1) they can be used to assess subjects that struggle with self-report, (2) neutral observers may provide more accurate evaluations than proxies, (3) it is not dependent on memory of the past, and (4) it is sensitive to changes in state.

However, the well-being domains identified as central in dementia by Clarke et al. (2020) are not all available for assessment through observation. Assessing well-being through observation implies coding or rating behavioral expressions, bodily positions, verbal or non-verbal expressions, or facial expressions that are all assumed to indicate the inner state of the observed person. Thus, we suggest observable aspects in line with the model of Clarke et al. (2020) are operationalized expressions of well-being in terms of positive behavioral expressions, balance between positive and negative emotions, level of engagement, expressions of satisfaction, and quality of social relationships. These aspects reflect central domains from the perspective of people living with dementia (Clarke et al., 2020) and central theories of well-being (Diener, 1984) and well-being in dementia (Lawton et al., 1996; Kitwood, 1997). The remaining domains of Clarke et al. (2020) related to experiencing meaning, having a positive sense of self and a sense of agency, may better be assessed through self-report. Still, accessing these domains and describing them may be difficult for people with more moderate and severe dementia.

Former reviews have reported on a variety of observational measures for people living with dementia (Curyto et al., 2008), including observational instruments specific for well-being and QoL in dementia (Algar et al., 2016), and measurements of emotional expressions in dementia (Lee et al., 2019). However, there is a lack of systematic reviews evaluating measurement instruments assessing momentary well-being in dementia that includes an evaluation against quality criteria and risk of bias. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative is a relevant systematic approach for reviewing health related outcome instruments (Prinsen et al., 2018). COSMIN is developed through extensive Delphi-studies with experts and in concordance with well-established systematic approaches for conducting reviews such as the Cochrane Handbook, the PRISMA statement, and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) principles (Mokkink et al., 2017; Prinsen et al., 2018; Terwee et al., 2018).

Our objective is to systematically review the literature and inform researchers and practitioners about the current state of knowledge and clinical utility of observational instruments assessing momentary well-being, to support care and interventions for people living with dementia. Guided by the COSMIN-framework, this systematic review aims to:

1. Identify observational instruments assessing momentary well-being in people with dementia.
2. Evaluate study specific methodological quality of the included publications through risk of bias (RoB) ratings.
3. Evaluate and compare measurement properties against quality criteria at instrument level.
4. Summarize and grade the trustworthiness of the body of evidence for each instrument.
5. Assess feasibility and interpretability of the instruments.

METHODS

The protocol for this review was pre-registered in the international register of systematic reviews, PROSPERO (RRID:SCR_019061, ID: 176160). **Figure 1** describes the COSMIN-guideline for conducting systematic reviews on health-related outcome measures that was utilized in this review.

Inclusion and Exclusion Criteria

Criteria for inclusion were (a) observational measures of momentary well-being (b) assessed by independent observers (c) during direct observation or video-recordings, containing (d) observable operationalizations of well-being such as positive and negative emotions/affect, or behavioral displays of satisfaction or engagement. Instruments should assess well-being (e) before, during and/or after psychosocial interventions over (f) short time intervals (minutes or hours). At least one psychometric property should be reported, and g) instruments not exclusively assessing well-being could be included, but only the well-being

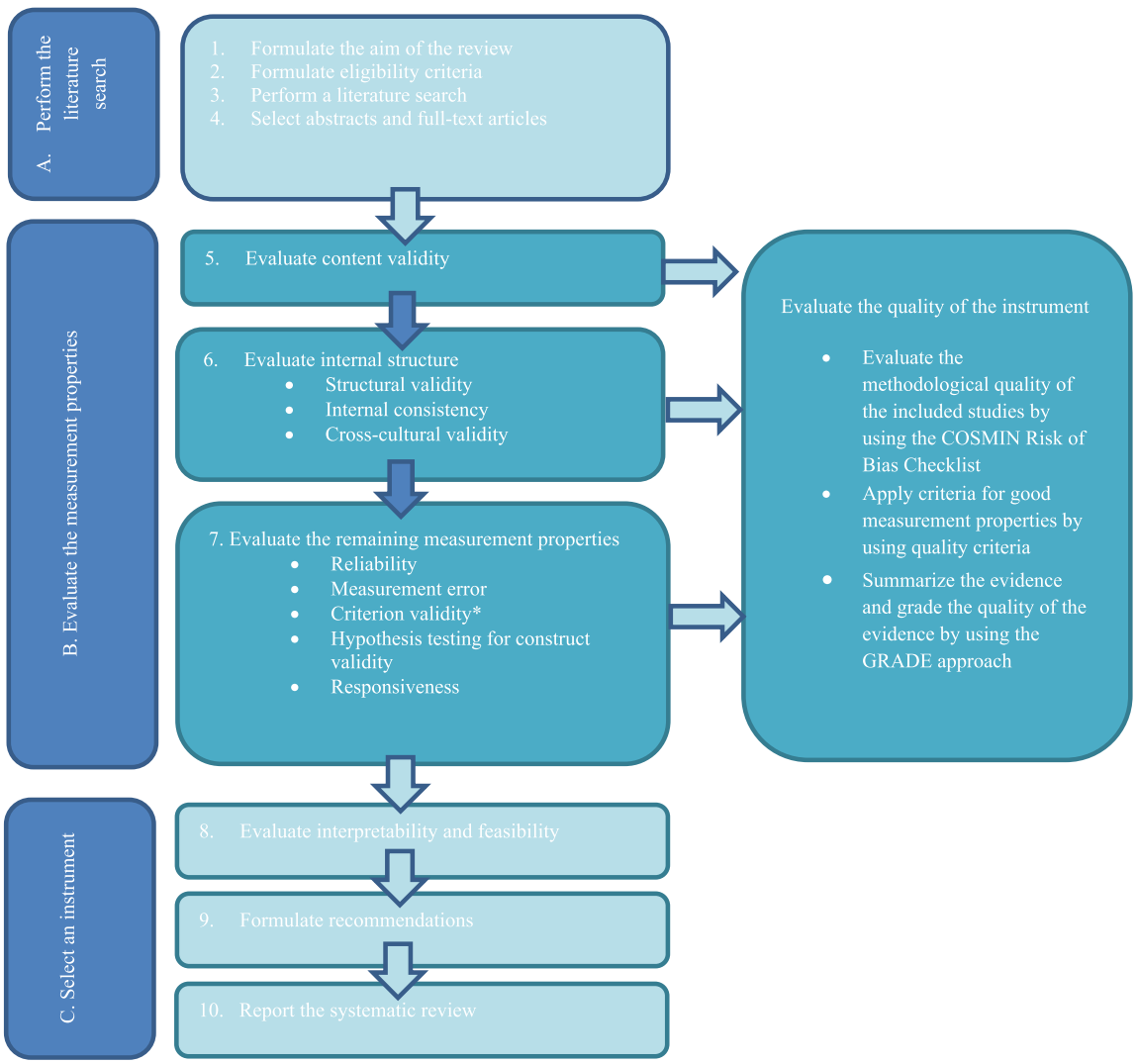

FIGURE 1 | COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guideline for systematic reviews of health-related measurement instruments. Reprint of this figure from Prinsen et al. (2018) is permitted under the Creative Commons Attribution 4.0. (http://creativecommons.org/licenses/by/4.0/). The acronym PROM (Patient Rated Outcome Instrument) is changed to "instrument" in this reprint. *Criterion validity was not assessed in this review, as no gold-standard instrument for comparison was identified.

domain would be assessed. Instruments developed for the general population could be included if they also were specifically tested in people with dementia. Only English peer-reviewed journal articles were included.

Exclusion criteria were observational instruments (a) focusing merely on ill-being, such as negative emotions, anxiety, depression or neuropsychiatric symptoms, and instruments measuring (b) observable physiological indicators of well-being only (such as biomarkers or startle reflex).

Search Strategy

Searches were performed on April 21st, 2020, and repeated on April 06th, 2021, in the databases MEDLINE, EMBASE, PsycINFO (all via OVD), Web of Science™, CINAHL (via EBSCOhost) and ProQuest® (Psychology and Nursing and Allied Health). A combination of the words "well-being," "dementia," "observation," "measurement," and "psychometric properties" were searched for, using both Boolean operators and truncations. We utilized the published search filter with words describing measurement properties of outcome instruments from Terwee et al. (2009). The full search strategy corresponding to the databases is available in **Supplementary Appendix A**.

The search was limited to peer-reviewed journal articles, searching in title, abstract and subject headings. In addition, we hand-searched reference lists of relevant reviews, investigated reference lists and forward chained citations of the included publications. Authors of relevant articles were contacted when the publication did not provide the full observation tool. Other publication types, such as conference proceedings, editorials and books were excluded (Prinsen et al., 2018), as were articles where the instrument was not accessible and lacked a full description of the operationalizations of the items in the publication.

Selection of Studies

The first author (KM) carried out the searches in the databases, imported the results to Endnote® (RRID:SCR_014001) where the results were checked, and duplicates removed. Next, KGM screened the titles and imported the records eligible for screening of abstract to Rayyan QCRI® (RRID:SCR_017584). KM also conducted hand searches of relevant records and imported these to Rayyan. The first (KM) and last (IN) author independently screened the records in Rayyan based on the eligibility criteria. Next, the results from the independent screening were compared, and all conflicts and their solutions of abstract screening were logged to ensure transparency. The next step was to evaluate the included publication based on full text. KM and IN read the full text independently and evaluated the publications against eligibility criteria in team meetings.

Data Extraction

Extraction was conducted by the first author (KM) and reviewed by a team including three of the authors (KM, EF-G, and IN). 20% of the data was extracted twice by the first author (KM) to ensue correct extraction. The extraction procedure was predefined and based on the COSMIN extraction tables (Prinsen et al., 2018). The first category addressed conceptualization (overarching conceptualization of well-being, population the instrument was developed in, and well-being domains assessed). The second category addressed central study characteristics (population, setting, methods, and results) for publications reporting on any of the measurement properties "content validity," "structural validity," "internal consistency," "cross-cultural validity/measurement invariance," "reliability," "measurement error," "construct validity" through hypothesis testing, and "responsiveness" (Mokkink et al., 2017; Prinsen et al., 2018; Terwee et al., 2018). The third category addressed feasibility (procedure, granularity, concreteness, training, requirements) and interpretability (measurement level and scoring, primary recording units, distribution, and sensitivity; Bakeman and Quera, 2012; Mokkink et al., 2017; Prinsen et al., 2018; Terwee et al., 2018).

Granularity refers to how fine grained and detailed the instrument is. Concreteness refers to how physically based the items are, where high concreteness involves bodily movement and low concreteness allows for interpretation of inner states. Measurement level defines which research questions may be asked, from nominal and ordinal to continuous output. Lastly, the primary recording unit defines how you sample the observations, from counting specific events in continuous or pre-specified intervals, to continuous recordings of duration (Bakeman and Quera, 2012; Chorney et al., 2015).

An overview of the COSMIN-definitions of central measurement properties of health-related instruments are provided in **Table 1**.

Evaluating Methodological Quality

Study specific RoB-ratings from multiple sources per instrument were ranked with the categories "very good," "adequate," "doubtful," "inadequate," and "not applicable." RoB-ratings were conducted by KM and IN in collaboration. Conflicting ratings were discussed with EF-G or NP. Rating criteria were based on the COSMIN RoB Checklist (Mokkink et al., 2017; Prinsen et al., 2018). The COSMIN-framework is created for patient-reported measurement instruments. To fit the COSMIN evaluations to the specific requirements for observational measures, some adaptations to the COSMIN-criteria were necessary. These mainly regarded the evaluation of content validity of the instruments. Our adaptations were based on recommendations from Bakeman and Quera (2011) and Bakeman and Quera (2012), and can be found in the **Supplementary Material** (**Supplementary Table 1**).

Consensus-based Standards for selection of health Measurement Instruments (COSMIN)-criteria for the *content validity* of self-reported measures are strongly based on feedback from the target group to assess relevance, comprehensiveness, and comprehensibility of the content of an instrument. Criteria for "relevance" requires items to be relevant for the construct of interest, the target population, and the context of use. To be "comprehensive," the items need to cover all key aspects of the construct (Terwee et al., 2018). We adapted the evaluations of content validity to observational measures based on Bakeman and Quera (2012); Chorney et al. (2015); and Perugia et al. (2018b). To get an "adequate" or "good" rating of content validity, our team decided at least two of the

TABLE 1 | COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) definitions of central terms.

Term	Definition[1]
Validity	*The degree to which an instrument measures the construct(s) it purports to measure*
Content validity	*The degree to which the content of an instrument is an adequate reflection of the construct(s) it purports to measure*
Construct validity[ab]	*The degree to which the scores of an instrument is consistent with hypotheses (for instance with regards to internal relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the instrument validly measures the construct to be measured*
Structural validity[c]	*The degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured*
Cross-cultural validity	*The degree to which the performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument*
Measurement invariance[2]	*Whether respondents from different groups with the same latent trait level (allowing for group differences) respond similarly to a particular item*
Reliability (extended definition)	*The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g., using different sets of items from the same* [instrument] *(internal consistency); over time (test–retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e., raters or responders) on different occasions (intra-rater)*
Internal consistency	*The degree of the interrelatedness among the items*
Measurement error	*The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured*
Reliability	*The proportion of the total variance in the measurement which is due to "true" differences between patients*
Responsiveness[b]	*The ability of an instrument to detect change over time in the construct to be measured*
Interpretability	*Interpretability is the degree to which one can assign qualitative meaning – that is, clinical or commonly understood connotations – to an instruments quantitative scores or change in scores*

[1] *Reprint of definitions permitted by the COSMIN-initiative. Original definitions are written in italics, and changes as regular text. (by the COSMIN team, all but,* [2] *available at https://cosmin.nl/wp-content/uploads/COSMIN-definitions-domains-measurement-properties.pdf.*

[2] *available at p. 51 https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf.*

[a] *As no gold standard for observing well-being in the field of dementia could be identified (Algar et al., 2016), criterion validity could not be evaluated (Prinsen et al., 2018). In this case, guidelines recommend to evaluate comparisons with other instruments as hypotheses testing for construct validity (Mokkink et al., 2017). These may be reported in the original publication as criterion validity, concurrent validity, convergent or divergent validity.*

[b] *While construct validity concerns hypothesis of correlations of single scores of similar instrument, responsiveness concerns testing hypotheses of correlations of change-scores of similar instruments to investigate the instruments ability to detect change (de Vet et al., 2011).*

[c] *In COSMIN, distinctions are made between reflective and formative instruments (de Vet et al., 2011). Reflective instruments (or subscales) are unidimensional, where increase in any item reflects an increase in the construct of interest. The evaluation of structural validity and internal consistency is only relevant for reflective scales with more than one item. Structural validity is the investigation of the expected unidimensionality of the instrument, and internal consistency is investigating the expected correlations between the items. Formative models have multidimensional structure and items may cause or form the construct independent of each other (de Vet et al., 2011).*

following approaches were required: theoretical approaches with literature reviews, qualitative field work and development of coding scheme or ethogram, and quantitative survey or qualitative interviews including the target group (people with dementia or their close care givers and/or experts from all relevant disciplines). In addition, lack of pilot field testing followed by evaluation and revision of the "comprehensibility" of the instrument lead to a rating of "inadequate."

Content validity is context- and population specific, implying that in this review the instruments' content validity is evaluated for the specific construct (well-being) in the specific context of evaluating psychosocial interventions for persons living with dementia (Terwee et al., 2018). Thus, evidence of content validity in other populations or contexts may not be generalizable and are not included.

As lack of *a priori* hypotheses is a common bias in health-related measurement development, we used a recommended generic hypothesis from COSMIN for evaluating construct validity and responsiveness (Prinsen et al., 2018, Table 4, p. 1154). COSMIN recommends *similar* constructs to be evaluated against a threshold of $\pm \geq 0.5$, and r*elated but dissimilar* constructs to be evaluated against a threshold of $\pm \geq 0.3$. Defining constructs as similar or only related *a priori* is a complex task. Relevant sources of measurement error identified in previous reviews are: (1) comparisons between state or trait dimensions (Curyto et al., 2008); (2) comparing self-, proxy- and observer-rated measures (Ferring and Boll, 2010); and (3) comparing instruments with different timeframes (Shiffman et al., 2008). Thus, we chose to use the recommended threshold of $\pm \geq 0.3$ as our threshold of comparison.

In addition, we did not expect decreasing well-being-scores to correlate with increasing dementia severity or cognitive impairment, as these constructs are found to be independent in several reviews (e.g., Missotten et al., 2008; Martyr et al., 2018).

Inter-rater reliability and agreement are particularly important properties of observational measures, and the new COSMIN-consensus regarding ratings of reliability and measurement error for clinician rated instruments was incorporated (Mokkink et al., 2020). The principle for overall quality scorings is 'the worst score counts', and one uses the lowest rating of the measurement property to indicate RoB (Mokkink et al., 2017; Prinsen et al., 2018; Terwee et al., 2018). COSMIN guidelines are available at www.cosmin.nl.

Data Synthesis

After the initial study specific evaluation, the total evidence provided for each instrument was rated against adapted COSMIN quality criteria using the ratings "good" (+), "unclear" (?), "inadequate" (-), = "conflicting" (±), "not evaluated" (NE), and "not applicable" (NA). **Table 2** provides an overview of the quality criteria. As most instruments were investigated in one publication only, no quantitative data synthesis was obtainable except for construct validity. For construct validity, the summarized number of hypotheses supporting the construct was divided by the sum of hypotheses (Prinsen et al., 2018).

The trustworthiness of the summarized quality criteria rating was ranked with Grading of Recommendations Assessment,

TABLE 2 | Adapted COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN)-quality criteria.

Property	Rating	Criteria
Content validity[a]	+	Both total relevance and comprehensiveness is rated as ' + ' and development study is not rated as 'inadequate.' An appropriate quantitative or qualitative data collection method used to identify relevant and comprehensive items for the instrument. At least two approaches used: theoretical approach with literature review, adaptations of other coding schemes, qualitative field work and development of coding scheme or ethogram, quantitative survey or qualitative interviews and focus groups including target group (experts from all relevant disciplines and/or patients and family care givers). Pilot test conducted. If there is a lack of evidence, the evaluation of the reviewers will determine overall rating
	−	Both total scores of relevance and comprehensiveness is rated '-'
	±	One of the two scores of relevance and comprehensiveness is rated '-' and the other is rated ' + '
Structural validity[b]	+	***CTT***: *CFA: CFI or TLI or comparable measure > 0.95 OR RMSEA < 0.06 OR SRMR < 0.08* ***IRT/Rasch***: *No violation of unidimensionality: CFI or TLI or comparable measure > 0.95 OR RMSEA < 0.06 OR SRMR < 0.08* *AND* *no violation of local independence: residual correlations among the items after controlling for the dominant factor < 0.20 OR Q3's < 0.37* *AND* *no violation of monotonicity: adequate looking graphs OR item scalability > 0.30* *AND* *adequate model fit IRT:* $\chi^2 > 0.001$ *Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardized values > −2 and < 2*
	?	***CTT***: *not all information for ' + ' reported* ***IRT/Rasch***: *model fit not reported*
	−	*Criteria for ' + ' not met*
Internal consistency[b]	+	*At least low evidence for sufficient structural validity AND Cronbach's alpha(s) ≥ 0.70 for each unidimensional scale or subscale*
	?	*Criteria for "At least low evidence for sufficient structural validity" not met*
	−	*At least low evidence for sufficient structural validity AND Cronbach's alpha(s) < 0.70 for each unidimensional scale or subscale*
Reliability[c]	+	*For continuous scores: ICC ≥ 0.70* *For ordinal or nominal scores: (weighted) Kappa ≥ 0.70*
	?	*ICC or (weighted) Kappa not reported*
	−	*ICC or (weighted) Kappa < 0.70*
Measurement error[c]	+	*For continuous scores: SDC or LoA or CV*$\sqrt{}$2*0.196 < M(C)IC* *For ordinal/nominal/dichotomous scores: Percentage specific (e.g., positive and negative) agreement calculated and above 80%*
	?	*MIC not defined*
	−	*For continuous scores: SDC or LoA or CV*$\sqrt{}$2*0.196 > M(C)IC* *For ordinal/nominal/dichotomous scores: Percentage specific (e.g., positive and negative) agreement calculated and above 80%*
Hypotheses-testing for construct validity[b]	+	*The results are in accordance with > 75% of the hypotheses,* and correlations with similar instruments are > 0.3
	?	Unclear hypotheses
	±	Results are in accordance with less than 75% of the hypotheses
	−	*The result is not in accordance with the hypotheses*, or all correlations are below > . 3
Cross-cultural validity/measurement invariance[b]	+	*No important differences found between group factors (such as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's* $R^2 < 0.02$*)*
	?	*No multiple group factor analysis OR DIF analysis performed*
	−	*Important differences between group factors OR DIF was found*
Responsiveness[b]	+	*The result is in accordance with > 75% of the hypotheses, OR AUC ≥ 0.70*
	?	Unclear hypotheses
	±	Results are in accordance with less than 75% of the hypotheses
	−	*The result is not in accordance with the hypotheses, OR AUC < 0.70*

[a] *Criteria is adapted from Terwee et al. (2018), available in the following COSMIN-manual (pp 58-59) https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf. Adaptations based on specific recommendations for development of observational instruments from Bakeman and Quera (2012).*

[b] *Criteria from Prinsen et al. (2018, p. 1152).*

[c] *Criteria from Mokkink et al. (2020) available in the following COSMIN-manual (p. 55) https://www.cosmin.nl/wp-content/uploads/user-manual-COSMIN-Risk-of-Bias-tool_v4_JAN_final.pdf. Reprint of tables from these three sources are permitted under the Creative Commons Attribution 4.0 (http://creativecommons.org/licenses/by/4.0/). Original criteria are written in italics, our adaptations are written as regular text.*

Abbreviations*: AUC, Area under the curve; CFA, confirmatory factor analysis; CFI, comparative fit index; CTT, classical test theory; CV, Coefficient of Variation; DIF, differential item functioning; ICC, intraclass correlation coefficient; IRT, Item response theory; LoA, Limits of Agreement; MIC, minimal important change; RMSEA, Root Mean Square Error of Approximation; SDC, Smallest detectable change; SRMR, Standardized Root Mean Residuals; TLI, Tucker-Lewis index.*

Ratings: *+, good; ?, unclear; −, inadequate; ±, conflicting; NE, not evaluated, NA, not applicable. Structural validity or internal consistency is reported as "not applicable" for instruments evaluated as formative.*

Development and Evaluation (GRADE) principles (GRADE Handbook, 2013), modified in the COSMIN approach for the context of health-related outcome measures (Prinsen et al., 2018). Four factors are assessed on instrument level: "risk of bias," "inconsistency," "imprecision," and "indirectness" of the evidence, graded as "high," "moderate," "low," or "very low". Ratings were conducted in team meetings with KM and IN, including EF-G if consensus was not met.

RESULTS

Search Results

Search results and reasons for exclusion is presented in **Figure 2**. After removing duplicates, KM screened 4309 records by title. Then, the 255 publications eligible for evaluation of abstracts was blind screened for inclusion by KM and IN (82% agreement). Additionally, 25 publications were added through hand search of relevant records. After full-text review of 87 records by KM and IN, 36 articles describing a total of 22 instruments were included, of which three originated from the hand-search.

Conceptualizations of the Included Instruments

Key characteristics of the instrument, target population and domains are presented in **Table 3**. The included instruments are sorted in the three (not mutually exclusive) categories (a) observations of emotions, (b) observations of positive behavioral expressions, and (c) observations of engagement. Instruments are presented in chronological order within each category.

Observations of Emotions

Five instruments were identified assessing emotion through operationalizations of facial, bodily, and behavioral expressions; *The Facial Action Coding System* (FACS, Ekman and Friesen, 1978; Ekman et al., 2002), *The Maximally Discriminative Facial Movement Coding System* (MAX, Izard, 1979, 1995), *The Observed Emotion Rating Scale*[1] (OERS, Lawton et al., 1996, 1999), *Observable Displays of Affect Scale* (ODAS, Vogelpohl and Beck, 1997), and *The Apparent Emotion Rating Instrument* (AER; Snyder et al., 1998). Two instruments employed generic approaches for emotion detection (FACS and MAX), two were dementia specific (OERS and ODAS), and one was developed to observe emotions in geriatric populations (AER).

Observations of Positive Expressions

Ten dementia-specific instruments that operationalized well-being as positive and negative expressions or responses to stimuli were identified; *Dementia Care Mapping* (DCM, Kitwood and Bredin, 1992), *The Positive Response Schedule* (PRS, Perrin, 1997), *Activity in Context and Time* (ACT; Wood, 2005), *Greater Cincinnati Chapter Well-Being Observational Tool* (GCC-WOT, Rentz, 2002), a revision of the former, named *Scripps Modified Greater Cincinnati Chapter Well-Being Observational Tool* (SM-GWW-WOT, Sauer et al., 2016), *AwareCare* (Clare et al., 2012), *The Behavior, Engagement and Affect Measure* (BEAM, Casey et al., 2014), *Maastricht Electronic Daily Life Observation tool* (MEDLO-tool, de Boer et al., 2016), COMMUNI-CARE (Lopez et al., 2016) and QUALIDEM-ILA (Junge et al., 2020).

Observations of Engagement

Seven instruments measuring engagement in dementia met the inclusion criteria; *Menorah Park Engagement Scale*[2] (MPES, Judge et al., 2000), *Observational Measurement of Engagement* (OME, Cohen-Mansfield et al., 2009), *Music in Dementia Assessment Scales* (MiDAS, McDermott et al., 2015), *Video coding – Incorporating Observed Emotion* (VC-IOE, Jones et al., 2015), *Engagement of a Person with Dementia Scale* (EPWDS, Jones et al., 2018), *Ethographic and Laban-Inspired Coding System of Engagement* (ELICSE, Perugia et al., 2018b), and *Music Therapy Engagement Scale for Dementia* (MTED, Tan et al., 2019).

Evaluating Measurement Properties

Extracted data on measurement properties and study characteristics are reported in **Supplementary Table 2** together with the *study specific* RoB-ratings. As most publications use data from repeated observations of the same subjects, both number of participants and number of observations are reported when available. Measurement properties are presented under three headings: (a) content validity, (b) construct validity, including structural validity, measurement invariance and hypothesis testing (for construct validity), and (c) reliability, including internal consistency, inter-rater, intra-rater or test–retest reliability, and measurement error. None of the included publications reported cross-cultural validity and responsiveness, using the methodological definition and criteria of COSMIN (see **Tables 1**, **2**).

The ratings against quality criteria for the available evidence of the measurement properties on *instrument level* are presented in **Table 4**. Ten of 22 instruments had only one publication describing the development and measurement properties. More than half of the instruments were developed or tested in small samples [11 of 36 studies have $n < 20$, mean $n = 89.4$ ($SD = 102$)]. The trustworthiness of the summarized result per property evaluated by the GRADE approach (GRADE Handbook, 2013; Prinsen et al., 2018) are presented in **Table 4**.

Content Validity

Seventeen of 22 instruments were rated as "good" when evaluated against quality criteria (MAX, OERS, PRS, DCM, ACT, GCWBT, SM-GCWBT, AwareCare, BEAM, MEDLO-tool, QUALIDEM-ILA, OME, MiDAS, VC-IOE, EPWDS, ELICSE/EMODEB, and MTED). Three instruments were rated as "conflicting" (ODAS, AER, and MPES), and two were rated as "inadequate" (FACS and COMMUNI-CARE). The study specific methodological

[1]Variations of the name of the OERS are Philadelphia Geriatric Center Affect Scales, Apparent Affect Rating Scale, Lawton's Modified Behavior Stream, Affect Rating Scale, and Observed Affect Scale (Lee et al., 2019).

[2]MPES is also referred to as Myers Research Institute Engagement Scale (Lee et al., 2007).

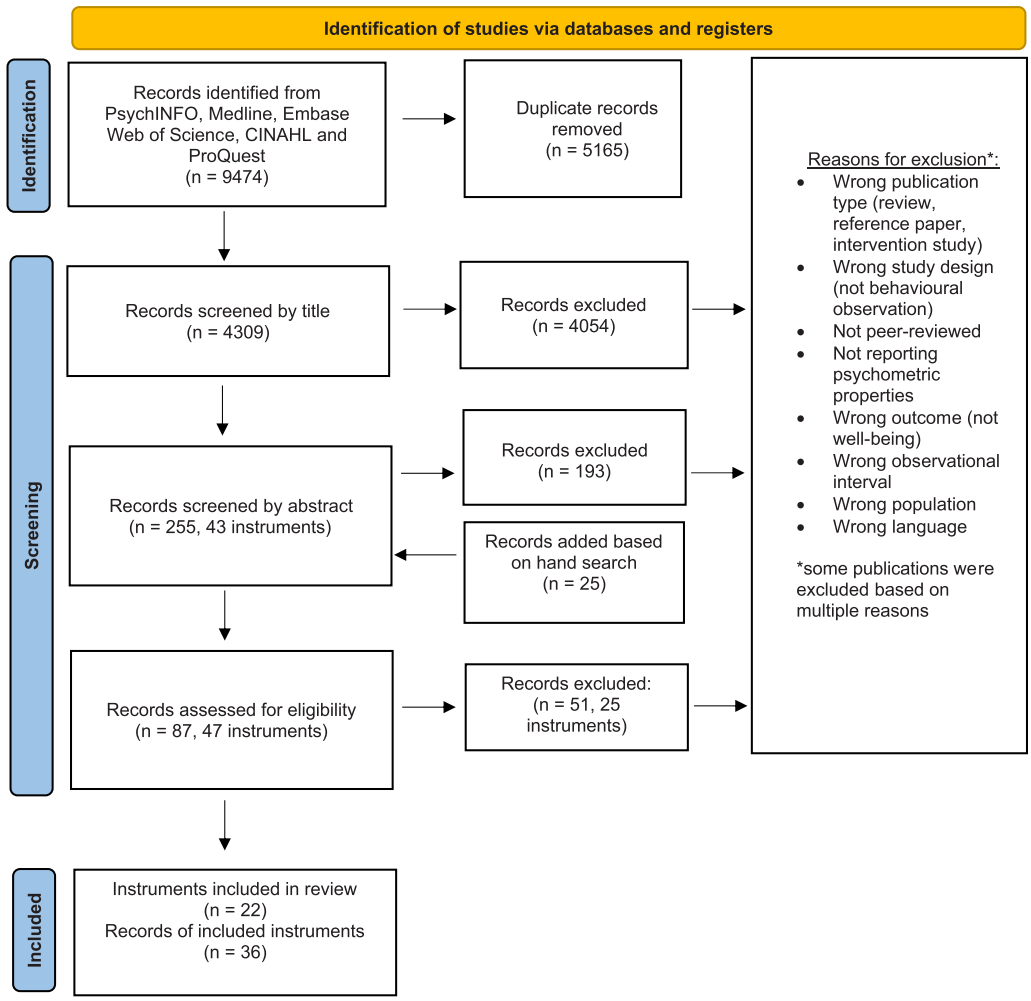

FIGURE 2 | PRISMA Flow chart of search results.

approach for establishing content validity is presented in **Supplementary Table 2**.

As presented in **Table 4**,11 of the 17 instruments meeting quality criteria, were supported with high quality ratings of evidence of content validity according to GRADE (MAX, OERS, DCM 8, ACT, GCWBT, AwareCare, BEAM, MEDLO-tool, QUALIDEM-ILA, MiDAS, and ELICSE/EMODEB). Inviting people with dementia and/or family caregivers to include their view on the content of the instruments is an advantage, but was only conducted for AwareCare, QUALIDEM-ILA, MiDAS, and EPWDS.

Structural Validity and Internal Consistency

Statistical methods to investigate structural validity are only developed for unidimensional and reflective instruments or subscales and require independent observations and large samples (de Vet et al., 2011). We identified nine scales as reflective (OERS, AER, GCWBT, SM-GCWBT, COMMUNI-CARE, QUALIDEM ILA, EPWDS, and MTED). Six of the nine scales used factor analysis to investigate structural validity. Except for OERS (Lawton et al., 1996) and QUALIDEM-ILA (Junge et al., 2020), all scales are at risk of bias due to small samples (<100, GCWBT; Gross et al., 2015; SM-GWWBT; Lokon et al., 2019; MiDAS; McDermott et al., 2014; MTED; Tan et al., 2019). Use of repeated (dependent) observations of the same individuals violates statistical assumptions of these methods as well (MiDAS; McDermott et al., 2014). As **Table 4** shows, no instruments have higher than "unclear"-rating of structural validity. This is mainly due to a lack of reporting model fit (OERS, QUALIDEM-ILA, MiDAS, and MTED). Investigations of structural validity for GCWBT (Gross et al., 2015) and SM-GCWBT (Lokon et al., 2019) did not confirm the theoretical factor structure.

Internal consistency was sometimes reported when no evidence of unidimensionality was provided (EPWDS; Jones et al., 2018; COMMUNI-CARE; Lopez et al., 2016; AER; Snyder et al., 1998). These results are rated as "unclear," as internal consistency is a reliability parameter relevant for reflective instruments known to be unidimensional only (Prinsen et al., 2018).

TABLE 3 | Characteristics of the included instruments.

Key references	Target population	Items/domains
EMOTIONS		
(Emotion) Facial Action Coding System (EMFACS/FACS) – describing positive and negative emotions based on facial behavior through action units (FACS) or systematic combination of action units expressing emotions (EMFACS)		
Ekman and Friesen (1978), Asplund et al. (1991), Asplund et al. (1995)	Generic instrument. Tested in people with moderate to severe dementia.	FACS – 27 descriptive action units EMFACS – combination of action units as emotions. Items tested in dementia research: Joy, surprise, sadness, anger, fear, disgust, and contempt
The Maximally Discriminative Facial Movement Coding System (MAX) – observing facial expressions of primary emotions		
Izard (1979), Izard (1995), Magai et al. (1996)	Generic instrument. Tested in persons with moderate to severe dementia.	13 descriptive units of facial behavior in mouth-lip region, 8 units in eyes-nose-cheek region, 6 units in brow region Formulas determine if one of eight emotions are detected: Joy, sadness, fear, anger, surprise, disgust, contempt, and interest
Observed Emotion Rating Scale (OERS) - Assessing emotions experienced by persons with Alzheimer's dementia		
Lawton et al. (1999)	Moderate to severe Alzheimer's dementia	Positive affect: Pleasure and interest Negative affect: Anger, anxiety/fear, and depression/sadness
The Apparent Emotion Rating Scale (AER) - Assessing positive and negative affect in geriatric populations.		
Snyder et al. (1998)	Geriatric populations with and without cognitive impairment in nursing homes, adult day care and research settings	Positive affect: Pleasure, interest, and tranquility Negative affect: Sadness, anxiety, and anger 15 verbal or non-verbal indicators for each domain.
Observable Displays of Affect (ODAS) - Behavioral displays of positive and negative affect following interventions		
Vogelpohl and Beck (1997), Beck et al. (2002)	People with dementia in nursing homes	41 behaviors of positive and negative affect categorized in six subscales 1. Facial positive displays 2. Facial negative displays 3. Vocal positive displays 4. Vocal negative displays 5. Body positive movement/posture 6. Body negative movement/posture
POSITIVE EXPRESSIONS		
Dementia Care Mapping version 8 (DCM-8) - Assessing psychological well-being and the quality of care in people with dementia in care settings		
Bradford Dementia Group (2005), Brooker and Surr (2006)	People with dementia in care settings	Combinations of Mood and Engagement (MEs) scores in correspondence to co-occurring Behavior Category Codes (BCCs) Additional: Personal Enhancers, Personal Detractions, and contextual field notes
Positive Response Schedule (PRS) – Assessing well-being in people with dementia through understanding occupational needs		
Perrin (1997)	People severely impaired by dementia	10 behavioral categories: Deliberate body movement, deliberate head movement, vocalization, looks at environment, looks at carer, initiates interaction, engagement, happy, sad, and fear
Activity in Context and Time (ACT) – Assessing environmental correlates of daily patterns of time use and well-being		
Wood (2005)	People with dementia in long term care settings	Environmental context domains (activity, social and physical) coded in relation to time use domains (positive behavior; gaze, mobility, conversation, and activity, negative behavior; agitation) and apparent affect (positive, negative, or null affect). Corresponding modifiers are created for each domain.
Greater Cincinnati Chapter Well-Being Observational Tool (GCWBT) – Assessing psychological well-being in people with dementia		
Kinney and Rentz (2005)	People with dementia in adult day programs, assisted living and long-term care. Assessing creative art interventions	Seven domains with 19 indicators of well-being: interest, sustained attention, pleasure, negative affect, sadness, self-esteem, and normalcy

(Continued)

TABLE 3 | (Continued)

Key references	Target population	Items/domains
Scripps Modified Greater Cincinnati Chapter Well-being Observation Tool (SM-GCWBT) - Psychological well-being and ill-being in people with dementia		
Sauer et al. (2016), Lokon et al. (2019)	Persons with moderate to advanced dementia in creative art interventions.	Two domains with 25 indicators Well-being: social interest, engagement, pleasure Ill-being: disengagement, negative affect, sadness, and confusion Domains scored on both frequency and intensity
AwareCare – Assessing behavioral signs of awareness and response to stimuli in people with severe dementia		
Clare et al. (2012)	People with severe dementia in care settings	10 different stimuli (not reviewed here) and 14 response categories: Eyes: eyes flicker, makes eye contact, explores with eyes Face: smiles, frowns, nods/shakes, moves head Limbs: reaches, grasps/holds Body: moves toward, moves away Vocalizations: single words, mumbles, shouts/moans
Behavior, Engagement and Affect Measure (BEAM) - Behavioral agitation, engagement and affect in people with dementia		
Casey et al. (2014)	People with mild to severe dementia living in long term care	Nine domains - Mobility status, activity context, agitation, positive behavior, engagement, affect, interaction: initiator, interaction: recipient, global contentment
Maastricht Electronic Daily Life Observational tool (MEDLO-tool) - Daily life aspects in long-term care, including emotional wellbeing		
de Boer et al. (2016)	Nursing home residents with moderate to severe dementia	Four domains: activity, physical environment, social interaction, and emotional well-being.
COMMUNI-CARE – Assessing psycho-emotional well-being in persons with dementia		
Lopez et al. (2016)	People with moderate to severe dementia during multi-sensory Snoezelen interventions	Five items – anxiety, communication, pleasure, adaptation to the surroundings, and affection
QUALIDEM for intensive longitudinal assessment (QUALIDEM-ILA) – Assessing momentary well-being of life in people with dementia		
Junge et al. (2020)	People with mild to severe dementia living in nursing homes.	Short version of QUALIDEM (Ettema et al., 2007) with 8 items in the following domains: restlessness, mood, anxiousness, body language, communication, happiness, sadness, and sociability
ENGAGEMENT		
Menorah Park Engagement Scale (MPES) - Engagement in activities		
Judge et al. (2000)	People with dementia in day care settings	Four categories of engagement: constructive engagement, passive engagement, non-engagement, and self-engagement
Observational measurement of Engagement (OME) - Engagement toward stimulus in persons with dementia		
Cohen-Mansfield et al. (2009)	People with dementia in long term care	Observations of response to stimuli: rate of refusal, duration of interest, attention, attitude, and activity
Music in Dementia Assessment Scales (MiDAS) - Musical engagement in music therapy for people with dementia		
McDermott et al. (2014), McDermott et al. (2015)	People with moderate to severe dementia receiving music therapy	Five visual analog subscales: interest, response, initiation, involvement, enjoyment Supplementary checklist of notable reactions during assessment (agitation/aggression, withdrawn/low in mood, restless/anxious, relaxed mood, attentive/interested, cheerful/smiling)
Video Coding – Incorporating Observed Emotion (VC-IOE) - Engagement toward stimulus (social robots)		
Jones et al. (2015)	People with dementia in care-settings	Six engagement-types with mutually exclusive operationalizations: emotion, verbal engagement, visual engagement, behavioral engagement, collective engagement, and agitation
Engagement of a Person with Dementia Scale (EPWDS) - Engagement toward an activity		
Jones et al. (2018)	People with dementia in acute, community and long-term care	Positive engagement or disengagement/negative engagement in the following five dimensions: affect, visual, verbal, behavioral and social

(Continued)

TABLE 3 | (Continued)

Key references	Target population	Items/domains
Ethographic and Laban Inspired Coding System of Engagement (ELICSE) and **Evidence-Based Model of Engagement-Related Behavior (EMODEB)** – Engagement naturally expressed through behaviors in activities of game-based and robot-based play		
Perugia et al. (2018b)	Mild to moderately severe dementia, nursing homes	13 different behaviors in three body parts. Head behavior, torso behavior and arms/hands behavior, and their following affective gestural support
Music therapy engagement scale (MTED) - Engagement in music therapy		
Tan et al. (2019)	Persons with dementia in acute hospital settings	Five domains of engagement: musical engagement, relatedness through music, verbal communication, emotional responsiveness, and overall responsiveness

Cross-Cultural Validity

No instruments reported cross-cultural validity. Nevertheless, instruments were developed in several different countries (see **Supplementary Table 2**), and eight reported the ethnicity of the included participants (BEAM; Casey et al., 2014; AwareCare; Clare et al., 2012; OME; Cohen-Mansfield et al., 2009; GCWBT; Kinney and Rentz, 2005; MAX; Magai et al., 2002; MiDAS; McDermott et al., 2014; MTED; Tan et al., 2019; ACT; Wood et al., 2005).

Measurement Invariance

Little evidence of measurement invariance was reported, when using COSMIN criteria. Only multiple group factor analysis and regression analysis are applicable approaches (Prinsen et al., 2018). An exception was FACS, where apathy explained lower frequency of facial emotions in people with mild to moderate dementia (Seidl et al., 2012).

Measurement invariance has important implications for interpretations of the scores of an instrument. Typical relevant covariates investigated were dementia severity, assessed with Pearson's or Spearman's correlations. Due to the methodological approach employed, these are reported under "construct validity" (Cfr. **Supplementary Table 2**). Lower well-being was correlated with dementia severity in MAX (Magai et al., 1997), AER (Snyder et al., 1998) and AwareCare (Clare et al., 2012). Evidence from earlier DCM-versions have shown well-being scores to vary due to level of cognitive impairment or dependency in the observed persons (Brooker and Surr, 2006; Chaudhury et al., 2013). QUALIDEM-ILA (Junge et al., 2020) and MTED (Tan et al., 2019) did not vary with dementia severity. Apathy correlated negatively with engagement in EPWDS (Jones et al., 2018).

Hypothesis Testing for Construct Validity

Sixteen of 22 instruments investigated construct validity through hypothesis testing. Nine instruments were thus rated as "good" ($>$ 75% of hypotheses supported; OERS, DCM 8, MEDLO-tool, COMMUNI-CARE, OME, MiDAS, EPWDS, ELICSE/EMODEB, and MTED). Five instruments were rated as "conflicting" (MAX, AER, AwareCare, BEAM, and QUALIDEM-ILA), and one as "inadequate" (FACS). Only OERS provided evidence rated as high quality according to GRADE.

A frequently detected risk of bias was lack of specific hypotheses about the strengths of correlations with similar or divergent measures, postulated *a priori* (Prinsen et al., 2018). According to our quality criteria, significant correlations ≤ 0.3 were discarded. Weak statistically significant correlations with instruments measuring similar constructs are not adequate evidence of construct validity (Mokkink et al., 2017), but were reported as evidence supporting construct validity in AER, BEAM, and QUALIDEM-ILA.

In three of the instruments with "conflicting" evidence (AwareCare, BEAM and QUALIDEM-ILA), proxy-reported long-term QoL ratings by staff and/or family members and momentary observations by independent observers did not correlate and consequently did not support construct validity (Clare et al., 2012; Casey et al., 2014; Junge et al., 2020). Overall, further investigation of construct validity with specific and *a priori* hypotheses is required for all instruments, except OERS.

Inter-Rater Reliability and Measurement Agreement

As **Table 4** demonstrates, some evidence of agreement between coders were reported in all but DCM 8 and QUALIDEM-ILA. Eight of 22 instruments (MAX, OERS, PRS, SM-GCWBT, COMMUNI-CARE, OME, EPWDS, and MTED) met quality criteria of inter-rater reliability (IRR, $>$ 0.70). Of these, only two (OERS and OME) were evaluated with high quality evidence according to GRADE. Some report IRR using invalid methods according to Prinsen et al. (2018) such as Spearman's Rho (BEAM; Casey et al., 2014) or Pearson's correlations (GCWOT; Gross et al., 2015). For instruments concerned about item levels, the items' specific Kappa values are the relevant parameters (Prinsen et al., 2018), but some report Kappa values on instrument level rather than an item-specific Kappa (GCWBT; Kinney and Rentz, 2005; COMMUNI-CARE; Lopez et al., 2016; SM-GCWBT; Sauer et al., 2016; PRS; Schall et al., 2015).

If the total sum of the scale is to be used, IRR should be assessed with intra class correlations (ICC), as the agreement of the *total sum* is the relevant reliability parameter (Prinsen et al., 2018). For most health measurement instruments, the preferred ICC formula is absolute agreement for random models with single measurements. This reflects whether different observers consistently reach the same conclusions (see de Vet et al., 2011; chapter 5). However, the formulae were often not reported and suboptimal calculations were often used.

For ordinal, nominal and dichotomous level scores, measurement error is defined as measurement agreement between raters. This was reported for 10 instruments, where

TABLE 4 | Rating against quality criteria and GRADE.

	Instrument	Content validity			Structural validity	Internal consistency	Cross-cultural validity	Measurement invariance	Construct validity	Reliability			Measurement error	Responsive-ness
		Rel	Comp	Total						IRR	Intra-r	TRR		
Emotions	FACS	±	–	–	NA	NA	NE	–	–	NE	NE	NE	±	NE
	MAX	+	+	+	NA	NA	NE	NE	±	+	NE	NE	NE	NE
	OERS	+	+	+	?	NE	NE	NE	+	+[1]	NE	NE	±	NE
	AER	±	+	±	NE	?	NE	NE	±	–	NE	NE	+	NE
	ODAS	±	+	±	NA	NA	NE	NE	NE	±	+	NE	+	NE
Positive expressions	DCM 8[2]	+	+	+	NA	NA	NE	NE	+	NE	NE	–	NE	NE
	PRS	+	+	+	NA	NA	NE	NE	NE	+	NE	NE	+	NE
	ACT	+	+	+	NA	NA	NE	NE	NE	±	NE	NE	NE	NE
	GCWBT	+	+	+	–	NE	NE	NE	NE	±[3]	NE	NE	NE	NE
	SM-GCWBT	+	+	+	–	?	NE	NE	NE	+	NE	NE	+	NE
	AWARECARE	+	+	+	NA	NA	NE	NE	±	±	NE	?	NE	NE
	BEAM	+	+	+	NA	NA	NE	NE	±	±	NE	NE	NE	NE
	MEDLO-tool[4]	+	+	+	NA	NA	NE	NE	+	±	NE	NE	±	NE
	COMMUNI-CARE	±	–	–	NE	?	NE	NE	+	+	NE	NE	NE	NE
	QUALIDEM-ILA	+	+	+	?	+	NE	NE	±	NE	NE	–	NE	NE
Engagement	MPES	+	–	±	NA	NA	NE	NE	NE	NE	NE	NE	+	NE
	OME	+	+	+	NA	NA	NE	NE	+	+	NE	NE	+	NE
	MiDAS	+[5]	+	+	?	?	NE	NE	+	±	NE	–	NE	NE
	VC-IOE	+	+	+	NA	NA	NE	NE	?	NE	NE	NE	+	NE
	EPWDS	+	+	+	NE	?	NE	NE	+	+	NE	+	NE	NE
	ELICSE/EMODEB	+	+	+	NA	NA	NE	NE	+	±	NE	NE	NE	NE
	MTED	+	+	+	?	+	NE	NE	+	+	NE	NE	NE	NE

High	Moderate	Low	Very low	Not evaluated

Ratings: +, good; ?, unclear; –, inadequate; ±, conflicting; NE, not evaluated; NA, not applicable.
Abbreviations: Rel, relevance; Comp, comprehensiveness; IRR, Intra-rater reliability; Intra-r, Intra-rater reliability; TRR, Test–retest reliability.
[1]OERS: for adequately trained independent observers, IRR is good.
[2]DCM 8– only data regarding the well-being subscale is evaluated, and the 8th version. However, evidence of former DCM versions covers problems with inter-rater reliability (Sloane et al., 2007) and measurement invariance with dependency (Brooker, 2005).
[3]IRR with extensive training met criteria (Kinney and Rentz, 2005) while shorter training did not (Gross et al., 2015).
[4]MEDLO-tool – only the well-being/agitation subscales are evaluated. [5]MiDAS is rated based on different timeframes in the staff ("today") versus the music-therapist ratings ("5 min"), and it is the "momentary" ratings we focus on in this review.

seven met the quality criteria (> 80%, AER, ODAS, PRS, SM-GCWBT, MPES, OME, and VC-IOE). Of these, only one was evaluated with high quality evidence (OME).

Low inter-rater agreement (IRR and measurement agreement) may reflect both lack of training and problems with content validity/poor operationalizations of the items. The amount of training will affect the level of inter-rater agreement, for instance as shown in OERS (Lawton et al., 1999) and when comparing inter-rater reliability for GCWBT with extensive training (Kinney and Rentz, 2005) and 30 min training (Gross et al., 2015). For MiDAS, the varying timeframes of the staff- and music-therapist ratings ("today" versus 5 min) may account for the low inter-rater reliability of the staff-ratings (McDermott et al., 2014). This may well reflect lower relevance of the items in the prolonged timeframe, and potentially issues concerning content validity.

Test–Retest Reliability and Measurement Error

Test–retest reliability was rarely investigated, and of the five scales reporting on this property, EPWDS was the only scale meeting the quality criteria. To validly evaluate test–retest reliability, the subjects need to be stable in the interim-period to ensure that any difference is caused by random measurement error (de Vet et al., 2011). In general, several studies showed fluctuating well-being scores (AwareCare; Clare et al., 2012; QUALIDEM-ILA; Junge et al., 2020; MiDAS; McDermott et al., 2014). Competing explanations of low test–retest reliability may include too long an interval between comparison measurements or may simply reflect qualities of the construct.

The low test–retest reliability detected for DCM 8 is prone to bias, as the assessments were three months apart (Villar et al., 2015).

For continuous level scores, measurement error is related to the test–retest reliability, and we need to know the smallest detectable change (SDC) or limits of agreement (LoA), as well as the minimal important change (MIC) defined by the target group, to apply the quality criteria (Prinsen et al., 2018). None of the instruments reported these outcomes.

Responsiveness

No instruments reported evidence of responsiveness.

Feasibility and Interpretability

Extracted data regarding feasibility and interpretability are reported in **Supplementary Table 3**. Additional publications from the search process describing use of the instrument in clinical settings or research were extracted here.

Feasibility

Four instruments require video-recordings (FACS, ODAS, VC-IOE, and ELICSE) and the latter may be used for direct observation. Several instruments allow for observing people simultaneously or sequentially (DCM, ACT, GCWBT, SM-GCWBT, BEAM, MEDLO-tool, MiDAS, and MTED). Some instruments were developed mainly as research tools (FACS, MAX, ODAS, PRS, ACT, VC-IOE, and ELICSE). Two instruments appear best suited for evaluation in care settings only (DCM 8 and MTED). Several instruments appear feasible for evaluating psychosocial interventions (FACS, MAX, ODAS, OERS, AER, PRS, ACT, MPES, BEAM, and QUALIDEM-ILA), and some are suited for care settings as well (OERS, AER, ACT, BEAM, QUALIDEM-ILA, DCM 8, AwareCare and MEDLO-tool). Some instruments are developed for *specific* interventional approaches, including art-interventions and other creative interventions (GCWBT and SM-GCWBT), multi-sensory interventions (COMMUNI-CARE), interaction with social robots (ELICSE, VC-IOE, and EPWDS), and music interventions (MiDAS and MTED). Most instruments are feasible for persons with mild, moderate, and severe dementia, but two instruments were specifically developed for very severe dementia (PRS and AwareCare). Personalized stimuli can be incorporated in two instruments (AwareCare and OME), and six instruments are easily adapted to other environmental contexts (OERS, GCWBT, SM-GCWBT, BEAM, MEDLO-tool, and ACT).

Interpretability

Skewed distributions of the negative expressions were commonly reported (FACS/EMFACS; Asplund et al., 1995; ODAS; Beck et al., 2002; Beerens et al., 2016; BEAM; Casey et al., 2014; MEDLO-tool; de Boer et al., 2016; MPES; Judge et al., 2000; GCWOT; Kinney and Rentz, 2005; OERS; Lawton et al., 1999; SM-GCWOT; Lokon et al., 2019; MAX; Magai et al., 1996, 2002; PRS; Perrin, 1997; Phillips et al., 2010; ACT; Wood, 2005). For AwareCare, infrequent items were removed during fieldwork to avoid skewness (Clare et al., 2012).

Sensitivity to detect statistically significant changes were demonstrated for FACS/EMFACS (in people with mild to moderate dementia; Seidl et al., 2012; but not for people with severe dementia; Asplund et al., 1995), MAX (Magai et al., 1996), OERS (when aggregating positive and negative affect; Hammar et al., 2011; except anger; Lawton et al., 1999), AER (Snyder et al., 2001), ODAS (for two of three subscales, Beck et al., 2002, or when aggregating scores to positive and negative affect; Lee et al., 2013, 2014, 2017), DCM 8 (Brooker, 2005), PRS (Hadley et al., 1999; Schall et al., 2015), ACT (Wood et al., 2005; Lassell et al., 2021), GCWBT (positive items only, Kinney and Rentz, 2005) and SC-GWBT (Sauer et al., 2016; Lokon et al., 2019), AwareCare (Clare et al., 2012, 2014), BEAM (for "happiness" and "agitation", Low et al., 2014), MEDLO-tool ("mood"; Beerens et al., 2016, 2018), MPES (Lee et al., 2007), OME (Cohen-Mansfield et al., 2011, 2012), MiDAS (Garrido et al., 2020) and EPWDS (Feng et al., 2020).

To ease interpretation, available sources for means and standard deviations of scores are reported in **Supplementary Table 3**. However, guidelines for interpretation of *clinically significant* scores or change scores are not identified in most instruments. DCM 8 offers calculating an individual or group level well-being profile. PRS gives a ratio, where higher ratios imply the setting triggers more well-being. AwareCare offers calculation of a "Responsiveness Index" for stimuli or for the individual, enabling the assessment of both individual processes and comparisons on group-level (Clare et al., 2012). COMMUNI-CARE provides a cut-off score of positive, indifferent, and negative effects of an intervention (Lopez et al., 2016). For ACT and EPWDS, creating an individual baseline is recommended to interpret change-scores.

DISCUSSION

In this review we investigated observational instruments assessing momentary well-being in the context of research, interventions and care for people living with dementia. We identified 22 instruments, and evaluated RoB on study level, and measurement properties, feasibility, and interpretability on instrument level. The content validity of many of the instruments reviewed was sound and supported by high quality evidence for 11 instruments. Meanwhile, the presence of high-quality evidence of other central psychometric aspects was sparse. This may in part be explained historically by the more recent development of stringent quality criteria. Hence, several instruments have the potential to meet these quality criteria if further investigated. To guide and advise further use of these instruments in care and research, we provide a general discussion of the most common methodological problems. Finally, we present instrument-specific recommendations.

Issues Regarding Measurement Properties, Feasibility, and Interpretability

Problems with skewed distributions or low frequencies of negative emotions, behaviors or expressions are reported for the majority of the instruments (Cfr. **Supplementary Table 3**). This complicates parametric approaches assuming a normal distribution of items. We suggest that assessing psychosocial interventions for people living with dementia should mainly focus on *increases* in well-being. Negative symptoms in dementia have a diversity of causes, some of which will necessarily be less modifiable by psychosocial interventions (Kales et al., 2015; Kolanowski et al., 2017; Livingston et al., 2017). However, momentary well-being is particularly achievable through modifying environmental factors (Lawton, 1994; Kolanowski et al., 2020). Moving the focus from ill-being (such as agitation or apathy) to well-being, has three advantages. First, it will decrease the labor intensiveness of the observational assessment because less items are assessed. Second, it will bring about data better fitted for statistical approaches because the distribution of ill-being items in the clinical studies using these instruments often were skewed and not normally distributed (see Asplund et al., 1995; Magai et al., 1996, 2002; Perrin, 1997; Lawton et al., 1999; Judge et al., 2000; Beck et al., 2002; Kinney and Rentz, 2005; Wood, 2005; Phillips et al., 2010; Casey et al., 2014; Beerens et al., 2016; de Boer et al., 2016; Lokon et al., 2019). Lastly, it will increase the likelihood of correct conclusions about the positive effects of the psychosocial interventions because this is operationally defined as an increase in positive expressions and not as a decrease in negative expressions. Ill-being should still be monitored during psychosocial interventions, but the absence of ill-being is not synonymous with well-being (Martyr et al., 2018).

While 15 of 22 instruments could detect statistically significant changes, definitions to guide *interpretation* of these change-scores were not provided. An option for future studies is to calculate MIC and the SDC or LoA (de Vet et al., 2011) for continuous level instruments. MIC is important because it is defined as the smallest clinical meaningful change as evaluated by patients or clinicians (de Vet and Terwee, 2010). SDC indicates whether change scores are reflecting a "true" change in the construct, as opposed to expected random error or natural fluctuation. Test–retest values may be used to calculate SDC for continuous scores (Prinsen et al., 2018; Mokkink et al., 2020). Several instruments were operationalized at a nominal or ordinal level, while using total score as continuous in statistical analyses. However, using the total score implies that the score reflects, predicts, or describes well-being validly. Although several instruments claim the total score to reflect level of well-being or engagement, adequate evidence of this relationship is rarely provided. Specifically, the formative instruments are hampered by unclear clinical interpretation.

Test–retest reliability reflects the instrument's measurement error in repeated measurement of stable constructs (de Vet et al., 2011). This required "stability" may be unattainable for fluctuating phenomena such as pain. In this review, several instruments provide evidence suggesting momentary well-being in dementia is a fluctuating phenomenon (Clare et al., 2012; McDermott et al., 2014; Junge et al., 2020). Fluctuations in the construct of interest between measurements creates an ambiguous reliability estimate (Jensen, 2003) and discarding instruments with a cut-off score < 0.70 (Prinsen et al., 2018) is not necessarily useful in this context. It is reasonable to assume test–retest scores reflect a natural fluctuation or variability in well-being in people with dementia, as the presence of neuropsychiatric symptoms such as apathy are episodic and fluctuating as well (Kales et al., 2015). Examining the natural variation of the construct by investigating test–retest reliability is nevertheless important, as the range of variation in fluctuating constructs influence the accuracy when interpreting scores of an instrument. Thus, a clinically significant score needs to be larger than the measurement error inflicted by this natural variation (de Vet et al., 2011). If test–retest reliability is not investigated, we cannot know if the measure can detect change in the observed persons beyond measurement error (Mokkink et al., 2020). This is a significant problem, that may lead to erroneous conclusions in both research and care. In addition, adjusting the interval of the repeated measurements to increase the likelihood of stability is essential, as longer time intervals may reflect the degenerative path of dementia and not instrument reliability.

Developing fine grained instruments used for ecological momentary sampling requires repeated assessment of the same subjects (Shiffman et al., 2008). Investigating behavior as it unfolds over time is labor intensive, and naturally includes smaller samples, often with numerous repeated observations. Standard approaches to develop self-rated instruments require large samples to investigate structural validity with factor analysis ($N > 100$), or scalability through for example Mokken analysis ($N > 2000$; Prinsen et al., 2018). Investigating large samples in labor intensive instruments is in many cases unrealistic. Additionally, using serially dependent repeated observations in the same subjects to increase the sample size violates basic assumptions required for these methods (Manolov and Moeyaert, 2017).

Most instruments in this review require further investigations of construct validity to ensure that the output is consistent with the underlying theoretical constructs. Comparisons with global rating scales are recommended when investigating the construct-validity of new instruments (de Vet et al., 2011). While developing COMMUNI-CARE, a validated clinician-rated global scale was used for this purpose (Lopez et al., 2016), but the same non-blinded investigator was rating both scales, contributing to a considerable risk of bias. In OME (Cohen-Mansfield et al., 2009), a similar approach is used, only with blinded ratings of a non-validated global engagement-scale. Thus, investigating construct validity through correlations with similar instruments is a challenge in the face of a lack of a "gold standard measure," as one must rely on existing instruments with their respective limitations (de Vet et al., 2011). Sometimes the hypothesized correlations included comparisons of well-being levels from long-term versus momentary instruments (Clare et al., 2012). Well-being states and traits do not necessary correlate (Curyto et al., 2008; Cohen-Mansfield, 2011). Therefore, investigating correlations with other momentary assessment approaches is recommended.

When assessing momentary well-being in dementia, two domains seem important to control for to interpret changes in well-being scores more accurately. Several of the instruments included in this review have a well-being score that is associated with (1) dementia severity or (2) level of function. However, research suggests that these constructs are not expected to be systematically related (Missotten et al., 2008; Barca et al., 2011; Cohen-Mansfield, 2011; Martyr et al., 2018). This has implications for how we interpret changes in well-being scores over time. If well-being scores of a particular instrument are lowered as a consequence of the dementia progressing, is this reflecting lack of treatment effect, poorer dementia care, or neurodegenerative development? Future studies assessing the measurement properties of these instruments should assess if a relationship between well-being and dementia severity or level of function is present. Such covariance may indicate that the instrument is tapping both cognitive functioning as well as well-being (for example if the score is relying on verbal expression). Understanding these relationships is required to accurately interpret changes in well-being scores during psychosocial interventions.

Personal well-being refers to a subjective evaluation, and observational measures use behavioral expressions to infer about an inner state. Hence, the most crucial property of a measurement instrument is *content validity*. Content validity will vary with the context, population, and construct to be measured, and affects all other psychometric properties of an instrument (Terwee et al., 2018). Together with agreement between observers, these two aspects are considered the most important for observational instruments (Bakeman and Quera, 2012; Chorney et al., 2015). Moreover, evidence of structural validity or construct validity, ensuring that an increase in the score reflects an increase in the construct, is important when making inferences about inner states. Cross-validating scores with other instruments, particularly self-report instruments, will strengthen this.

As no evidence of cross-cultural validity or responsiveness was detected, special attention to investigating this knowledge-gap and establishing these properties are important in future studies using any of the instruments in this review. In relation to cross-cultural validity, we make the following recommendation: Behavioral expressions of momentary well-being are likely to differ across cultures (Lim, 2016). Thus, securing cross cultural validity by *establishing content validity in new cultural contexts* is in our evaluation an alternative to statistical evaluation of cross-cultural validity for observational measures. This can be achieved through the recommended qualitative approaches involving clinical expertise from people with dementia, family- and professional caregivers, as well as clinical experts and field testing (Terwee et al., 2018).

In relation to the lack of responsiveness, we make the following recommendation: Several instruments have provided evidence of their capacity to statistically detect changes in intervention studies (Conf. **Supplementary Table 3**). However, this is not adequate evidence of responsiveness, as we do not know if the lack of detecting change is due to lack of responsiveness or lack of intervention effect. Responsiveness of these instruments needs to be investigated through correlations with change-scores in similar instruments (de Vet et al., 2011).

The *clinical utility* of an instrument is specific to the context and aims of the user, and is influenced by its feasibility, interpretability, benefits, and shortcomings (Smart, 2006; Terwee et al., 2018). To recommend a specific instrument to assess observed well-being is not our intention. However, we generally recommend identifying instruments with proper conceptualizations, which are feasible for the specific purpose, context, and target population (Terwee et al., 2018). Choosing instruments with acceptable content validity should be followed by investigation or adaptation to solve the additional instrument-specific issues addressed in this review. An overview of the issues of each instrument is provided in **Table 4**, **Supplementary Tables 2, 3**. Establishing or evaluating if the instrument has good content validity in the applied context is vital, especially in securing relevance and comprehensiveness (Chorney et al., 2015).

A final note worth commenting regards the large number of instruments identified in the hand search, of which three were included in this review. This suggests that researchers may not be choosing appropriate keywords when publishing articles relating to observational measures for people living with dementia.

Recommendations of Instruments

Of the instruments measuring *emotions* with acceptable content validity (OERS and MAX), OERS is the most frequently used (Lee et al., 2019) instrument with the most extensively documented psychometric properties (Lawton et al., 1996, 1999). MAX (and FACS) requiring a close view of the face; problems with interpreting facial movement in persons wearing glasses, having facial hair, or facing more than 45 degrees away from the camera (Cohn et al., 2007) reduces the clinical utility of these instruments in people living with dementia. Thus, the feasibility of instruments relying on facial expressions and excluding bodily expressions may decrease the instruments' sensitivity to detect

expressions of well-being in the dementia population (Seidl et al., 2012). However, as negative emotions are infrequent, the feasibility of the full OERS scale in research and clinical setting is limited (Algar et al., 2016). Thus, for investigating well-being in people with dementia, the positive emotions in OERS may be best suited. However, from these findings, emotions in people with mild to moderate dementia seem to be best measured through self-report (instruments are reviewed in Ferring and Boll, 2010; Stoner et al., 2019; and Clarke et al., 2020).

Users looking for instruments investigating *positive expressions* are recommended to consider any instruments with acceptable content validity (DCM 8, PRS, ACT, GCWBT, SM-GCWBT, AwareCare, BEAM, MEDLO-tool and QUALIDEM ILA). PRS and MEDLO-tool are instruments with high granularity, detecting changes on micro-levels that offer interval-sampling from 30 s to 2 min. While DCM, ACT, GCWBT, SM-GCWBT offers somewhat fine-grained observations (5-10 min), AwareCare offers fine-grained observations as they unfold over time, and BEAM consists of both fine-grained and aggregated scores. QUALIDEM-ILA is best suited for total evaluations of interventions (30-45 min). Users looking for behavioral or movement-anchored operationalizations of positive expressions with high levels of concreteness may look at PRS, ACT and AwareCare. DCM, GCWBT, SM-GCWBT, BEAM, MEDLO-tool and QUALIDEM-ILA offer more contextual cues and social interpretations.

AwareCare appears clinically useful for people with very severe dementia, and BEAM is feasible for moderate dementia. AwareCare detected signs of awareness in all participants and suggests a clinically useful index for interpretation as well (Clare et al., 2012). PRS needs to be investigated in a larger sample but is a promising tool in very severe dementia (Perrin, 1997). BEAM covers behavior, engagement, and affect, through direct observation in various settings without being very labor intensive and while avoiding observer's fatigue (Casey et al., 2014). Further investigation of its construct validity may, however, be required, in addition to an improved evaluation of inter-rater reliability. The clinical sensitivity of DCM has been questioned (Cooke and Chaudhury, 2013), and the well-being (ME-score) of DCM 8 is probably not sensitive enough to detect clinical change reliably in intervention studies on a group level. DCM 8 seems better suited for clinical practice (Villar et al., 2015) on an individual level (Brooker and Surr, 2006). MEDLO-tool's mood score is based on DCM as well, and shows the same problems (Beerens et al., 2016; de Boer et al., 2016), lowering the utility of this instrument for assessing well-being. ACT is based on a thorough development (Wood, 2005), and seems like a feasible and clinically useful instrument, but needs further investigation of construct validity. GCWBT should be omitted due to evidence of low structural validity (Gross et al., 2015), but the revised SM-GCWBT needs further modification and investigation of a proposed two-factor structure, as well as exclusion of some unrelated items (Lokon et al., 2019). Further investigation of QUALIDEM-ILA, in terms of both inter-rater reliability and use in clinical/research contexts are required (Junge et al., 2020). Still, QUALIDEM-ILA is one of the most recent instruments included in this review, and further publications are expected.

Of the instruments assessing *engagement* with acceptable content validity (OME, VC-IOE, EPWDS, ELICSE, MiDAS, and MTED), users searching for instruments with high granularity may look at VC-IOE or ELICSE (both continuous sampling), EPWDS or MiDAS (5-min intervals), or OME (15 min including both duration-based and aggregated scores). MTED provides an aggregated score based on the intervention-session. ELICSE and VC-IOE offers the highest level of concreteness, and EPWDS, OME, MiDAS, and MTED is less concrete and more interpretative. However, higher levels of concreteness will often increase labor intensiveness (Bakeman and Quera, 2011) and offer broader generalizability, at the cost of lower sensitivity to individual variations. In clinical contexts, allowing for interpreting idiographic expressions of well-being may sometimes be an advantage.

Ethographic and Laban Inspired Coding System of Engagement (ELICSE) is based on an exemplary solid development-phase with subsequent theoretical and conceptual development (Perugia et al., 2018a,b, 2020). Nevertheless, the system is highly context specific to the manipulation of objects when sitting down and may not be as easily adaptable to other activities or clinical contexts. Developers of OME describe a need for further work on increasing the clinical utility of the scale (Cohen-Mansfield et al., 2011, 2012), and it is critiqued for lack of interpretability (Jones et al., 2015, 2018; Perugia et al., 2018b). VC-IOE needs further evaluation of reliability and construct validity (Jones et al., 2015). MiDAS strength is the inclusion of the target group in the development (McDermott et al., 2015), but needs further investigation of psychometric properties and is hampered by low intra-rater reliability (McDermott et al., 2014). MTED appear to be a good option when evaluating engagement in clinical music therapy processes, but the scale is not intended for evaluating intervention effect (Tan et al., 2019).

Engagement of a Person with Dementia Scale (EPWDS) stands out as a feasible, easily administered scale that may allow for assessing engagement in contexts other than robot-based play (Jones et al., 2018). Formal evaluation of its structural validity is required, but indications of test–retest reliability are promising given the common problems of low stability between assessments in this population.

Strengths and Limitations

The first strength of this review is that the protocol was pre-registered in PROSPERO. The second strength is that we used the most relevant systematic approach, the COSMIN-guidelines (Prinsen et al., 2018; Terwee et al., 2018; Mokkink et al., 2020). The third strength is that when required, these guidelines were adapted for evaluating observational instruments based on relevant literature (Bakeman and Quera, 2011, 2012; Chorney et al., 2015; Perugia et al., 2018b). The fourth strength is the extensive review of study-specific and instrument-specific evaluation and overarching methodological issues that provides relevant knowledge to both researchers and practitioners.

A first limitation of this review is that by including instruments reporting at least one psychometric property, instruments describing promising content validity only were not evaluated (such as Morse and Chatterjee, 2018). A second

limitation is that the COSMIN-criteria of construct validity requires at least 75% of hypotheses to be supported. This may lead to somewhat unbalanced ratings, as publications reporting only one or two supportive correlations may be given a more positive rating than studies examining multiple correlations. However, testing several hypotheses provides more detailed knowledge about construct validity. A third limitation is the use of correlations of > 0.3 as the cutoff for supporting construct validity. This cutoff may seem low, and less conservative than the original suggestion of correlations ≥ 0.5 with instruments measuring similar constructs (Prinsen et al., 2018). However, the majority of the correlated instruments were assessing related and not similar constructs, indicating that correlations > 0.3 are an adequate expectation. Finally, the blinding procedure within our team of raters could have been more extensive, as completely blinded ratings are considered the gold standard (Mokkink et al., 2017).

Conclusion

Several instruments may validly assess well-being through observation in people with dementia. Evaluating their context specific clinical utility and content validity are more important than choosing the instrument with the best ratings or psychometric properties. However, piloting the instruments, investigating the effects of cultural context and study-specific inter-rater agreement and measurement error is advised. Moreover, utilizing an instrument in a clinical study provides the opportunity to investigate hypotheses that may further inform the construct validity. All measurement approaches come with some strengths and some weaknesses, and observational measures are vulnerable to misinterpretation when they are used to infer about inner states. Nevertheless, observations offer unique opportunities to investigate associations between external stimuli and well-being that can provide important knowledge of the usefulness of various interventions for people living with dementia.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

KM conducted the literature searches, initial screening of records and imported these to the data-management-tools, extracted data, assessed RoB, evaluated against quality criteria, and conducted GRADE-ratings, and wrote methods and result-section, with ideas and commentaries from EF-G and IN. IN and KM blind-screened the abstracts for inclusion and adapted the COSMIN guidelines to observational measures. These were consecutively reviewed in consensus-meetings with KM and IN. EF-G were included in discussions if consensus was not met. KM, EF-G, and IN reviewed and consolidated extracted data into the current tables and wrote the introduction and discussion in collaboration. NP edited the document for conceptual clarity and discussed methodological and quantitative considerations regarding the measures. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the University of Bergen, Norway.

ACKNOWLEDGMENTS

We wish to acknowledge librarian Kjersti Aksnes-Hopland at the University of Bergen Library for her important advice about search strategies, databases and tools for deduplication and management of references. We appreciate the COSMIN-initiative that offers freely available tools with high standards for reviewing and assessing health related measurement instruments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.742510/full#supplementary-material

REFERENCES

Algar, K., Woods, R. T., and Windle, G. (2016). Measuring the quality of life and well-being of people with dementia: a review of observational measures. *Dementia* 15, 832–857. doi: 10.1177/1471301214540163

Asplund, K., Jansson, L., and Norberg, A. (1995). Facial expressions of patients with dementia: a comparison of two methods of interpretation. *Int. Psychogeriatr.* 7, 527–534. doi: 10.1017/S1041610295002262

Asplund, K., Norberg, A., Adolfsson, R., and Waxman, H. M. (1991). Facial expressions in severely demented patients: a stimulus-response study of four patients with dementia of the Alzheimer type. *Int. J. Geriatr. Psychiatry* 6, 599–606. doi: 10.1002/gps.930060809

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

Bakeman, R., and Quera, V. (2012). "Behavioral observation," in *APA Handbook of Research Medhods in Psychology*, eds H. Cooper, M. Paul, L. L. Debra, A. T. Panter, D. Rindskopf, and K. J. Sher (Washington: American Psychological Association), 207–225. doi: 10.1037/13619-013

Barca, M. L., Engedal, K., Laks, J., and Selbæk, G. (2011). Quality of Life among elderly patients with dementia in institutions. *Dement. Geriatr. Cogn. Disord.* 31, 435–442. doi: 10.1159/000328969

Beck, C. K., Vogelpohl, T. S., Rasin, J. H., Uriri, J. T., O'sullivan, P., Walls, R., et al. (2002). Effects of behavioral interventions on disruptive behavior and affect in demented nursing home residents. *Nurs. Res.* 51, 219–228. doi: 10.1097/00006199-200207000-00002

Beerens, H. C., De Boer, B., Zwakhalen, S. M. G., Tan, F. E. S., Ruwaard, D., Hamers, J. P. H., et al. (2016). The association between aspects of daily life and quality of life of people with dementia living in long-term care facilities: a

momentary assessment study. *Int. Psychogeriatr.* 28, 1323–1331. doi: 10.1017/S1041610216000466

Beerens, H. C., Zwakhalen, S. M. G., Verbeek, H., E.S. Tan, F., Jolani, S., Downs, M., et al. (2018). The relation between mood, activity, and interaction in long-term dementia care. *Aging Ment. Health* 22, 26–32. doi: 10.1080/13607863.2016.1227766

Bowling, A., Rowe, G., Adams, S., Sands, P., Samsi, K., Crane, M., et al. (2015). Quality of life in dementia: a systematically conducted narrative review of dementia-specific measurement scales. *Aging Ment. Health* 19, 13–31. doi: 10.1080/13607863.2014.915923

Bradford Dementia Group. (2005). *DCM 8 User's Manual: The DCM Method*. Bradford: University of Bradford.

Brooker, D. (2005). Dementia care mapping: a review of the research literature. *Gerontologist* 45, 11–18. doi: 10.1093/geront/45.suppl_1.11

Brooker, D., and Surr, C. (2006). Dementia Care Mapping (DCM): initial validation of DCM 8 in UK field trials. *Int. J. Geriatr. Psychiatry* 21, 1018–1025. doi: 10.1002/gps.1600

Casey, A.-N., Low, L.-F., Goodenough, B., Fletcher, J., and Brodaty, H. (2014). Computer-assisted direct observation of behavioral agitation, engagement, and affect in long-term care residents. *J. Am. Med. Dir. Assoc.* 15, 514–520. doi: 10.1016/j.jamda.2014.03.006

Chaudhury, H., Cooke, H., and Frazee, K. (2013). "Developing a physical environmental evaluation component of the Dementia Care Mapping (DCM) tool," in *Environmental Gerontology: Making Meaningful Places in Old Age*, eds G. D. Rowles and M. Bernard (New York: Springer Publishing Company), 153–172.

Chorney, J. M., Mcmurtry, C. M., Chambers, C. T., and Bakeman, R. (2015). Developing and modifying behavioral coding schemes in pediatric psychology: a Practical Guide. *J. Pediatr. Psychol.* 40, 154–164. doi: 10.1093/jpepsy/jsu099

Clare, L., Quinn, C., Hoare, Z., Whitaker, R., and Woods, R. T. (2014). Care staff and family member perspectives on quality of life in people with very severe dementia in long-term care: a cross-sectional study. *Health Qual. Life Outcomes* 12, 175–175. doi: 10.1186/s12955-014-0175-3

Clare, L., Whitaker, R., Quinn, C., Jelley, H., Hoare, Z., Woods, B., et al. (2012). AwareCare: development and validation of an observational measure of awareness in people with severe dementia. *Neuropsychol. Rehabil.* 22, 113–133. doi: 10.1080/09602011.2011.640467

Clarke, C., Woods, B., Moniz-Cook, E., Mountain, G., Øksnebjerg, L., Chattat, R., et al. (2020). Measuring the well-being of people with dementia: a conceptual scoping review. *Health Qual. Life Outcomes* 18:249. doi: 10.1186/s12955-020-01440-x

Cohen-Mansfield, J. (2011). "The shifting baseline theory of well-being: Lessons from across the aging spectrum," in *Understanding Well-Being In The Oldest Old*, eds L. W. Poon and J. Cohen-Mansfield (New York: Cambridge University Press), 46–64.

Cohen-Mansfield, J., Dakheel-Ali, M., Jensen, B., Marx, M. S., and Thein, K. (2012). An analysis of the relationships among engagement, agitated behavior, and affect in nursing home residents with dementia. *Int. Psychogeriatr.* 24, 742–752. doi: 10.1017/S1041610211002535

Cohen-Mansfield, J., Dakheel-Ali, M., and Marx, M. S. (2009). Engagement in persons with dementia: the concept and its measurement. *Am. J. Geriatr. Psychiatry* 17, 299–307. doi: 10.1097/JGP.0b013e31818f3a52

Cohen-Mansfield, J., Marx, M., Freedman, L., Murad, H., Regier, N., Thein, K., et al. (2011). The Comprehensive Process Model of Engagement. *Am. J. Geriatr. Psychiatry* 19, 859–870. doi: 10.1097/JGP.0b013e318202bf5b

Cohn, J. F., Ambadar, Z., and Ekman, P. (2007). "Observer-Based Measurement of Facial Expression With The Facial Action Coding System," in *Handbook of Emotion Elicitation and Assessment*, eds J. A. Coan and J. J. B. Allen (New York: Oxford University Press), 203–221.

Cooke, H. A., and Chaudhury, H. (2013). An examination of the psychometric properties and efficacy of Dementia Care Mapping. *Dementia* 12, 790–805. doi: 10.1177/1471301212446111

Curyto, K., Haitsma, K., and Vriesman, D. (2008). Direct ob servation of behavior: a review of current measures for use with older adults with dementia. *Res. Gerontol. Nurs.* 1, 52–76. doi: 10.3928/19404921-20080101-02

de Boer, B., Beerens, H. C., Zwakhalen, S. M. G., Tan, F. E. S., Hamers, J. P. H., and Verbeek, H. (2016). Daily lives of residents with dementia in nursing homes: development of the Maastricht electronic daily life observation tool. *Int. Psychogeriatr.* 28, 1333–1343. doi: 10.1017/S1041610216000478

de Vet, H. C. W., and Terwee, C. B. (2010). The minimal detectable change should not replace the minimal important difference. *J. Clin. Epidemiol.* 63, 804–805. doi: 10.1016/j.jclinepi.2009.12.015

de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., and Knol, D. L. (2011). *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press.

Diener, E. (1984). Subjective well-being. *Psychol. Bull.* 95, 103–157.

Dröes, R. M., Chattat, R., Diaz, A., Gove, D., Graff, M., Murphy, K., et al. (2016). Social health and dementia: a European consensus on the operationalization of the concept and directions for research and practice. *Aging Ment. Health* 21, 4–17. doi: 10.1080/13607863.2016.1254596

Ekman, P., and Friesen, W. V. (1978). *Facial Action Coding System (FACS): Manual*. Palo Alto: Consulting Psychologist Press.

Ekman, P., Friesen, W. V., and Hager, J. C. (2002). *Facial Action Coding System: The Manual on CD ROM*. Salt Lake City: A Human Face.

Ettema, T. P., Dröes, R.-M., De Lange, J., Mellenbergh, G. J., and Ribbe, M. W. (2007). QUALIDEM: development and evaluation of a dementia specific quality of life instrument—-validation. *Int. J. Geriatr. Psychiatry* 22, 424–430. doi: 10.1002/gps.1692

Feng, Y., Barakova, E. I., Yu, S., Hu, J., and Rauterberg, G. W. M. (2020). Effects of the level of interactivity of a social robot and the response of the augmented reality display in contextual interactions of people with dementia. *Sensors* 20:3771. doi: 10.3390/s20133771

Ferring, D., and Boll, T. (eds) (2010). *Subjective Well-Being in Older Adults: Current State and Gasp of Research*. London: Palgrave Macmillan.

Garrido, S., Dunne, L., Stevens, C. J., and Chang, E. (2020). Music playlists for people with dementia: trialing a guide for caregivers. *J. Alzheimers Dis.* 77, 219–226. doi: 10.3233/JAD-200457

GRADE Handbook (2013). *Handbook for Grading the Quality of Eidence and the Strength of Reccomendations Using the GRADE Approach*. Available online at: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html (Accessed April 21, 2020)

Gross, S. M., Danilova, D., Vandehey, M. A., and Diekhoff, G. M. (2015). Creativity and dementia: does artistic activity affect well-being beyond the art class?. *Dementia* 14, 27–46. doi: 10.1177/1471301213488899

Hadley, C., Brown, S., and Smith, A. (1999). Evaluating interventions for people with severe dementia: using the Positive Response Schedule. *Aging Ment. Health* 3, 234–240. doi: 10.1080/13607869956190

Hammar, L. M., Gotell, E., and Engstrom, G. (2011). Singing while caring for persons with dementia. *Arts Health* 3, 39–50. doi: 10.1080/17533015.2010.481289

Izard, C. E. (1979). *The Maximally Discriminative Facial Movement Coding System (MAX)*. Newark: University of Delaware.

Izard, C. E. (1995). *The Maximally Discriminative Facial Movement Coding System (MAX) (Rev. ed)*. Newark: University of Delaware.

Jensen, M. P. (2003). Questionnaire validation: a brief guide for readers of the research literature. *Clin. J. Pain* 19, 345–352. doi: 10.1097/00002508-200311000-00002

Jones, C., Sung, B., and Moyle, W. (2015). Assessing engagement in people with dementia: a new approach to assessment using video analysis. *Arch. Psychiatr. Nurs.* 29, 377–382. doi: 10.1016/j.apnu.2015.06.019

Jones, C., Sung, B., and Moyle, W. (2018). Engagement of a Person with Dementia Scale: establishing content validity and psychometric properties. *J. Adv. Nurs.* 74, 2227–2240. doi: 10.1111/jan.13717

Judge, K. S., Camp, C. J., and Orsulic-Jeras, S. (2000). Use of Montessori-based activities for clients with dementia in adult day care: effects on engagement. *Am. J. Alzheimers Dis. Other Demen.* 15, 42–46. doi: 10.1177/153331750001500105

Junge, S., Gellert, P., O'sullivan, J. L., Moller, S., Voigt-Antons, J.-N., Kuhlmey, A., et al. (2020). Quality of life in people with dementia living in nursing homes: validation of an eight-item version of the qualidem for intensive longitudinal assessment. *Qual. Life Res.* 29, 1721–1730. doi: 10.1007/s11136-020-02418-4

Kales, H. C., Gitlin, L. N., and Lyketsos, C. G. (2015). Assessment and management of behavioral and psychological symptoms of dementia. *BMJ* 350:h369. doi: 10.1136/bmj.h369

Kaufmann, E. G., and Engel, S. A. (2016). Dementia and well-being: a conceptual framework based on Tom Kitwood's model of needs. *Dementia* 15, 774–788. doi: 10.1177/1471301214539690

Kinney, J. M., and Rentz, C. A. (2005). Observed well-being among individuals with dementia: memories in the Making, an art program, versus other structured activity. *Am. J. Alzheimers Dis. Other Demen.* 20, 220–227. doi: 10.1177/153331750502000406

Kitwood, T. (1997). *Dementia Reconsidered: The Person Comes First*. Buckingham: Open University Press.

Kitwood, T., and Bredin, K. (1992). A new approach to the evaluation of dementia care. *J. Adv. Health Nurs. Care* 1, 41–60.

Kolanowski, A., Behrens, L., Lehman, E., Oravecz, Z., Resnick, B., Boltz, M., et al. (2020). Living well with dementia: factors associated with nursing home residents' affect balance. *Res. Gerontol. Nurs.* 13, 21–30. doi: 10.3928/19404921-20190823-01

Kolanowski, A., Boltz, M., Galik, E., Gitlin, L. N., Kales, H. C., Resnick, B., et al. (2017). Determinants of behavioral and psychological symptoms of dementia: a scoping review of the evidence. *Nurs. Outlook* 65, 515–529. doi: 10.1016/j.outlook.2017.06.006

Kolanowski, A., Hoffman, L., and Hofer, S. M. (2007). Concordance of self-report and informant assessment of emotional well-being in nursing home residents with dementia. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 62, 20–27. doi: 10.1093/geronb/62.1.P20

Kolanowski, A., Van Haitsma, K., Meeks, S., and Litaker, M. (2014). Affect balance and relationship with well-being in nursing home residents with dementia. *Am. J. Alzheimers Dis. Other Demen.* 29, 457–462. doi: 10.1177/1533317513518657

Lassell, R., Wood, W., Schmid, A. A., and Cross, J. E. (2021). A comparison of quality of life indicators during two complementary interventions: adaptive gardening and adaptive riding for people with dementia. *Complement. Ther. Med.* 57:102658. doi: 10.1016/j.ctim.2020.102658

Lawton, M. P. (1994). Quality of life in Alzheimer's disease. *Alzheimer Dis. Assoc. Disord.* 8, 138–150. doi: 10.1097/00002093-199404000-00015

Lawton, M. P., Van Haitsma, K., and Klapper, J. (1996). Observed affect in nursing home residents with Alzheimer's disease. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 51:3. doi: 10.1093/geronb/51b.1.p3

Lawton, M. P., Van Haitsma, K., Perkinson, M., and Ruckdeschel, K. (1999). Observed affect and quality of life in dementia: further affirmations and problems. *Aging Ment. Health* 5, 69–81.

Lee, K. H., Algase, D. L., and Mcconnell, E. S. (2013). Daytime observed emotional expressions of people with dementia. *Nurs. Res.* 62, 218–225. doi: 10.1097/NNR.0b013e31829999d7

Lee, K. H., Algase, D. L., and Mcconnell, E. S. (2014). Relationship between observable emotional expression and wandering behavior of people with dementia. *Int. J. Geriatr. Psychiatry* 29, 85–92. doi: 10.1002/gps.3977

Lee, K. H., Boltz, M., Lee, H., and Algase, D. L. (2017). Is an engaging or soothing environment associated with the psychological well-being of people with dementia in long-term care?. *J. Nurs. Scholarsh.* 49, 135–142. doi: 10.1111/jnu.12263

Lee, K. H., Lee, J. Y., Boltz, M., and Mcconnell, E. S. (2019). Emotional expression of persons with dementia: an integrative review with implications for evidence-based practice. *Worldviews Evid. Based Nurs.* 16, 344–351. doi: 10.1111/wvn.12395

Lee, M. M., Camp, C. J., and Malone, M. L. (2007). Effects of intergenerational Montessori-based activities programming on engagement of nursing home residents with dementia. *Clin. Interv. Aging* 2, 477–483.

Lim, N. (2016). Cultural differences in emotion: differences in emotional arousal level between the East and the West. *Integr. Med. Res.* 5, 105–109. doi: 10.1016/j.imr.2016.03.004

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., et al. (2017). Dementia prevention, intervention, and care. *Lancet* 390, 2673–2734. doi: 10.1016/S0140-6736(17)31363-6

Lokon, E., Sauer, P. E., and Li, Y. (2019). Activities in dementia care: a comparative assessment of activity types. *Dementia* 18, 471–489. doi: 10.1177/1471301216680890

Lopez, J. J. B., Bolivar, J. C. C., and Perez, M. S. (2016). COMMUNI-CARE: assessment tool for reactions and behaviours of patients with dementia in a multisensory stimulation environment. *Dementia* 15, 526–538. doi: 10.1177/1471301214528346

Low, L.-F. P., Goodenough, B. P., Fletcher, J. P., Xu, K. P., Casey, A.-N. B., Chenoweth, L. R. N. P., et al. (2014). The effects of humor therapy on nursing home residents measured using observational methods: the SMILE cluster randomized trial. *J. Am. Med. Dir. Assoc.* 15, 564–569. doi: 10.1016/j.jamda.2014.03.017

Madsø, K. G., and Nordhus, I. H. (2021). Implementation of quality of life assessment in long-term care. *Int. Psychogeriatr.* 33, 861–863. doi: 10.1017/S1041610221000983

Magai, C., Cohen, C., Gomberg, D., Malatesta, C., and Culver, C. (1996). Emotional expression during mid- to late-stage dementia. *Int. Psychogeriatr.* 8, 383–395. doi: 10.1017/S104161029600275X

Magai, C., Cohen, C. I., Culver, C., Gomberg, D., and Malatesta, C. (1997). Relation between premorbid personality and patterns of emotion expression in mid- to late-stage dementia. *Int. J. Geriatr. Psychiatry* 12, 1092–1099. doi: 10.1002/(SICI)1099-1166(199711)12:11<1092::AID-GPS690<3.0.CO2-X

Magai, C., Cohen, C. I., and Gomberg, D. (2002). Impact of training dementia caregivers in sensitivity to nonverbal emotion signals. *Int. Psychogeriatr.* 14, 25–38. doi: 10.1017/S1041610202008256

Manolov, R., and Moeyaert, M. (2017). How can single-case data be analyzed? Software resources, tutorial, and reflections on analysis. *Behav. Modif.* 41, 179–228. doi: 10.1177/0145445516664307

Martyr, A., Nelis, S. M., Quinn, C., Wu, Y.-T., Lamont, R. A., Henderson, C., et al. (2018). Living well with dementia: a systematic review and correlational meta-analysis of factors associated with quality of life, well-being and life satisfaction in people with dementia. *Psychol. Med.* 48, 2130–2139. doi: 10.1017/S0033291718000405

McDermott, O., Orgeta, V., Ridder, H. M., and Orrell, M. (2014). A preliminary psychometric evaluation of Music in Dementia Assessment Scales (MiDAS). *Int. Psychogeriatr.* 26, 1011–1019. doi: 10.1017/S1041610214000180

McDermott, O., Orrell, M., and Ridder, H. M. (2015). The development of Music in Dementia Assessment Scales (MiDAS). *Nord. J. Music Ther.* 24, 232–251. doi: 10.1080/08098131.2014.907333

Missotten, P., Squelard, G., Ylieff, M., Di Notte, D., Paquay, L., De Lepeleire, J., et al. (2008). Relationship between quality of life and cognitive decline in dementia. *Dement. Geriatr. Cogn. Disord.* 25, 564–572. doi: 10.1159/000137689

Mokkink, L. B., Boers, M., Van Der Vleuten, C. P. M., Bouter, L. M., Alonso, J., Patrick, D. L., et al. (2020). COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med. Res. Methodol.* 20:293. doi: 10.1186/s12874-020-01179-5

Mokkink, L. B., De Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., et al. (2017). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual. Life Res.* 27, 1171–1179. doi: 10.1007/s11136-017-1765-4

Morse, N., and Chatterjee, H. (2018). Museums, health and wellbeing research: co-developing a new observational method for people with dementia in hospital contexts. *Perspect. Public Health* 138, 152–159. doi: 10.1177/1757913917737588

National Institute for Health and Care Excellence [NICE] (2018). *Dementia: Assessment, management and support for people living with dementia and their carers*. London: National Institute for Health and Care Excellence.

Øksnebjerg, L., Diaz-Ponce, A., Gove, D., Moniz-Cook, E., Mountain, G., Chattat, R., et al. (2018). Towards capturing meaningful outcomes for people with dementia in psychosocial intervention research: a pan-European consultation. *Health Expect.* 21, 1056–1065. doi: 10.1111/hex.12799

Perrin, T. (1997). The Positive Responses Schedule for severe dementia. *Aging Ment. Health* 1, 184–191. doi: 10.1080/13607869757290

Perugia, G., Diaz-Boladeras, M., Catala-Mallofre, A., Barakova, E. I., and Rauterberg, M. (2020). "ENGAGE-DEM: A Model of engagement of people with dementia," in *IEEE Transactions on Affective Computing*, (Piscataway: IEEE). doi: 10.1109/TAFFC.2020.2980275

Perugia, G., Van Berkel, R., Díaz-Boladeras, M., Català-Mallofré, A., Rauterberg, M., and Barakova, E. (2018b). Understanding engagement in dementia through behavior. The ethographic and Laban-inspired coding system of engagement (ELICSE) and the Evidence-based Model of Engagement-related Behavior (EMODEB). *Front. Psychol.* 9:690. doi: 10.3389/fpsyg.2018.00690

Perugia, G., Roíguez-Martín, D., Díaz-Boladeras, M., Català-Mallofré, A., Barakova, E. I., and Rauterberg, G. W. M. (2018a). Quantity of movement as a measure of engagement for dementia : the influence of motivational

disorders. *Am. J. Alzheimers Dis. Other Demen.* 33, 112–121. doi: 10.1177/1533317517739700

Phillips, L. J., Reid-Arndt, S. A., and Pak, Y. (2010). Effects of a Creative Expression Intervention on Emotions, Communication, and Quality of Life in Persons With Dementia. *Nurs. Res.* 59, 417–425. doi: 10.1097/NNR.0b013e3181faff52

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C. W., et al. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual. Life Res.* 27, 1147–1157. doi: 10.1007/s11136-018-1798-3

Reilly, S. T., Harding, A. J. E., Morbey, H., Ahmed, F., Williamson, P. R., Swarbrick, C., et al. (2020). What is important to people with dementia living at home? A set of core outcome items for use in the evaluation of non-pharmacological community-based health and social care interventions. *Age Ageing* 49, 664–671. doi: 10.1093/ageing/afaa015

Rentz, C. A. (2002). Memories in the Making© : outcome-based evaluation of an art program for individuals with dementing illnesses. *Am. J. Alzheimers Dis. Other Demen.* 17, 175–181. doi: 10.1177/153331750201700310

Sands, L. P., Ferreira, P., Stewart, A. L., Brod, M., and Yaffe, K. (2004). What explains differences between dementia patients' and their caregivers' ratings of patients' quality of life?. *Am. J. Geriatr. Psychiatry* 12, 272–280. doi: 10.1097/00019442-200405000-00006

Sauer, P. E., Fopma-Loy, J., Kinney, J. M., and Lokon, E. (2016). "It makes me feel like myself": person-centered versus traditional visual arts activities for people with dementia. *Dementia* 15, 895–912. doi: 10.1177/1471301214543958

Schall, A., Haberstroh, J., and Pantel, J. (2015). Time series analysis of individual music therapy in dementia: effects on communication behavior and emotional well-being. *GeroPsych* 28, 113–122. doi: 10.1024/1662-9647/a000123

Schulz, R., Cook, T. B., Beach, S. R., Lingler, J. H., Martire, L. M., Monin, J. K., et al. (2013). Magnitude and causes of bias among family caregivers rating Alzheimer disease patients. *Am. J. Geriatr. Psychiatry* 21, 14–25. doi: 10.1016/j.jagp.2012.10.002

Seidl, U., Lueken, U., Thomann, P. A., Kruse, A., and Schröder, J. (2012). Facial Expression in Alzheimer's Disease: impact of Cognitive Deficits and Neuropsychiatric Symptoms. *Am. J. Alzheimers Dis. Other Demen.* 27, 100–106. doi: 10.1177/1533317512440495

Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological Momentary Assessment. *Annu. Rev. Clin. Psychol.* 4, 1–32. doi: 10.1146/annurev.clinpsy.3.022806.091415

Skevington, S. M., and Böhnke, J. R. (2018). How is subjective well-being related to quality of life? Do we need two concepts and both measures?. *Soc. Sci. Med.* 206, 22–30. doi: 10.1016/j.socscimed.2018.04.005

Sloane, P. D., Brooker, D., Cohen, L., Douglass, C., Edelman, P., Fulton, B. R., et al. (2007). Dementia care mapping as a research tool. *Int. J. Geriatr. Psychiatry* 22, 580–589. doi: 10.1002/gps.1721

Smart, A. (2006). A multi-dimensional model of clinical utility. *Int. J. Qual. Health Care* 18, 377–382. doi: 10.1093/intqhc/mzl034

Snyder, M., Ryden, M. B., Shaver, P., Wang, J.-J., Savik, K., Gross, C. R., et al. (1998). The Apparent Emotion Rating Instrument: assessing affect in cognitively impaired elders. *Clin. Gerontol.* 18, 17–29. doi: 10.1300/J018v18n04_03

Snyder, M., Tseng, Y.-H., Brandt, C., Croghan, C., Hanson, S., Constantine, R., et al. (2001). A glider swing intervention for people with dementia. *Geriatr. Nurs.* 22, 86–90. doi: 10.1067/mgn.2001.115197

Stoner, C. R., Stansfeld, J., Orrell, M., and Spector, A. (2019). The development of positive psychology outcome measures and their uses in dementia research: a systematic review. *Dementia* 18, 2085–2106. doi: 10.1177/1471301217740288

Tan, J., Wee, S.-L., Yeo, P. S., Choo, J., Ritholz, M., and Yap, P. (2019). A new music therapy engagement scale for persons with dementia. *Int. Psychogeriatr.* 31, 49–58. doi: 10.1017/S1041610218000509

Terwee, C. B., Jansma, E. P., Riphagen, I. I., and De Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual. Life Res.* 18:1115. doi: 10.1007/s11136-009-9528-5

Terwee, C. B., Prinsen, C., Chiarotto, A., Westerman, M., Patrick, D., Alonso, J., et al. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual. Life Res.* 27, 1159–1170. doi: 10.1007/s11136-018-1829-0

Trigg, R., Watts, S., Jones, R., and Tod, A. (2011). Predictors of quality of life ratings from persons with dementia: the role of insight. *Int. J. Geriatr. Psychiatry* 26:83. doi: 10.1002/gps.2494

Villar, F., Vila-Miravent, J., Celdran, M., Fernandez, E., and D'arcy, L. (2015). Assessing the well-being of patients with dementia: validity and reliability of the Spanish version of the Dementia Care Mapping (DCM) tool. *Estud. Psicol.* 36, 389–412. doi: 10.1080/02109395.2015.1026124

Vogelpohl, T. S., and Beck, C. K. (1997). Affective responses to behavioral interventions. *Semin. Clin. Neuropsychiatry* 2, 102–112. doi: 10.1053/SCNP00200102

Wood, W. (2005). Toward developing new occupational science measures: an example from dementia care research. *J. Occup. Sci.* 12, 121–129. doi: 10.1080/14427591.2005.9686555

Wood, W., Harris, S., Snider, M., and Patchel, S. A. (2005). Activity situations on an Alzheimer's disease special care unit and resident environmental interactions, time use, and affect. *Am. J. Alzheimers Dis. Other Demen.* 20, 105–118. doi: 10.1177/153331750502000210

World Health Organization [WHO] (2017). *Global Action Plan On the Public Health Response to Dementia 2017–2025*. Geneva: World Health Organisation.

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

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

Copyright © 2021 Madsø, Flo-Groeneboom, Pachana and Nordhus. 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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 05 January 2022
doi: 10.3389/fpsyg.2021.795328

The Work Gratitude Scale: Development and Evaluation of a Multidimensional Measure

Carolyn M. Youssef-Morgan[1], Llewellyn E. van Zyl[2,3,4,5] and Barbara L. Ahrens[6]*

[1] College of Business, Bellevue University, Bellevue, NE, United States, [2] Department of Industrial Engineering and Innovation Sciences, University of Eindhoven, Eindhoven, Netherlands, [3] Optentia Research Unit, North-West University, Vanderbijlpark, South Africa, [4] Department of Human Resource Management, University of Twente, Enschede, Netherlands, [5] Department of Social Psychology, Institut für Psychologie, Goethe University, Frankfurt am Main, Germany, [6] Waukee Community School District, Waukee, IA, United States

This study explores gratitude as a multidimensional and work-specific construct. Utilizing a sample of 625 employees from a variety of positions in a medium-sized school district in the United States, we developed and evaluated a new measure, namely the Work Gratitude Scale (WGS), which encompasses recognized conative (intentional), cognitive, affective, and social aspects of gratitude. A systematic, six-phased approach through structural equation modeling (SEM) was used to explore and confirm the factorial structure, internal consistency, measurement invariance, concurrent, convergent, and discriminant validity of the WGS. The results supported a 10-item measure with three dimensions: "grateful appraisals" (three items), "gratitude toward others" (four items), and "intentional attitude of gratitude" (three items). Thereafter, first-order, second-order, and bifactor confirmatory models were estimated and compared. Work gratitude was found to be best described by a second-order construct with three underlying first-order dimensions. Measurement invariance was supported in relation to gender. Concurrent validity was supported in relation to two existing dispositional gratitude scales, namely the Gratitude Questionnaire and the Gratitude, Resentment, and Appreciation Scale (GRAT). Convergent validity was supported in relation to the Core Self-Evaluations Scale (CSES) and the Psychological Capital Questionnaire. Discriminant validity was supported in relation to various demographic factors such as age, gender, occupation, and tenure. The findings support the WGS as a multidimensional measure that can be used in practice to measure overall work-related gratitude and to track the effectiveness of gratitude-related workplace interventions.

Keywords: gratitude, work gratitude, positive psychology, positive organizational behavior, scale development, measurement, positive psychological assessment

OPEN ACCESS

Edited by:
Chiara Consiglio, Sapienza University of Rome, Italy

Reviewed by:
Ana Cláudia Souza Vazquez, Federal University of Health Sciences of Porto Alegre, Brazil
Katarzyna Danuta Durniat, University of Wrocław, Poland

****Correspondence:***
Carolyn M. Youssef-Morgan
cyoussefmorgan@bellevue.edu

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

Received: *14 October 2021*
Accepted: *17 November 2021*
Published: *05 January 2022*

Citation:
Youssef-Morgan CM, van Zyl LE and Ahrens BL (2022) The Work Gratitude Scale: Development and Evaluation of a Multidimensional Measure. Front. Psychol. 12:795328. doi: 10.3389/fpsyg.2021.795328

INTRODUCTION

As both a state and a psychological strength, gratitude has become one of the fundamental building blocks of positive psychology (van Zyl et al., 2021). Emmons (2004) defines gratitude as "a sense of thankfulness and joy in response to receiving a gift, whether the gift can be a tangible benefit from a specific other or a moment of peaceful bliss evoked by natural beauty" (p. 554).

There are many recognized benefits of gratitude. For example, gratitude promotes health, wellbeing and life satisfaction (Dickens, 2017). Gratitude is also related to effective coping (Wood et al., 2007), development of social support, and reduction of stress and depression (Wood et al., 2008a). Additionally, gratitude leads to prosocial behaviors that can promote healthy, satisfying, and productive relationships and connections (Portocarrero et al., 2020).

To-date, with very few exceptions, much of the empirical gratitude research pertains to non-work domains (Emmons, 2004; Watkins, 2014). On the other hand, in the context of positive organizational behavior, Luthans et al. (2015) propose gratitude as an evidence-based positive psychological resource that is open to development and management in the workplace. Relatedly, Fehr et al. (2017) emphasize that "organizations are not simply extensions of everyday social interactions. Rather, the organizational context introduces a unique set of constraints and affordances that influence how individual employees feel, think, and act on a daily basis" (p. 361) and make a case for a multifaceted work gratitude model that incorporates several personal, situational, and organizational contingencies. It follows that work gratitude may also exhibit unique dimensionality that may necessitate the purposeful design and validation of dedicated measures. This study seeks to fill this gap, by developing a work-specific and multidimensional measure of gratitude.

Conceptualizations of Gratitude

Gratitude has been conceptualized and measured as a *general, stable dispositional trait* (Portocarrero et al., 2020) or an enduring virtue (Peterson and Seligman, 2004; Morgan et al., 2017). Seligman et al. (2005) conceptualize character strengths such as gratitude as "trait-like—an individual difference with demonstrable generality and stability" (p. 411). In other words, within this perspective, grateful people have a general tendency to experience thankfulness or appreciation more frequently, more intensely, for longer periods, and across a broader range of people and situations than their less grateful counterparts (McCullough et al., 2002). In the context of the workplace, Cain et al. (2019) conceptualize gratitude as a dispositional trait or an enduring "tendency to notice and be thankful for how various aspects of a job affect one's life" (p. 441).

However, alternative, more malleable perspectives of gratitude also abound and are supported by abundant empirical evidence. Gratitude has been shown to involve several cognitive, affective, social, and situational factors (Wood et al., 2008b, 2009). Research shows that gratitude varies with situations and events (Wood et al., 2008a). Gratitude can also change over time (Froh et al., 2010; Chopik et al., 2019). Importantly, gratitude is open to development through a variety of relatively simple and practical interventions (*c.f.,* Dickens, 2017; Richter et al., 2021). Interestingly, recent studies utilizing functional neuromagnetic imaging (fMRI) provide tangible neuroimaging evidence that gratitude, previously thought to be a "hardwired" tendency, is indeed malleable and responsive to basic, short, well-recognized development interventions such as gratitude journaling (Karns et al., 2017). Thus, gratitude cannot be just a general, dispositional trait.

Within the malleable perspectives of gratitude, at least four views can be found: (a) affective/emotional, (b) cognitive/evaluative, (c) social/other-focused, and (d) conative/intentional. First, gratitude has been supported as a quick, intense, and constantly fluctuating *emotional state*, with substantial within-person variability (Emmons and Mishra, 2011; Spence et al., 2014). Second, gratitude can be viewed as a *situation-specific cognition* that involves positive appraisals of various aspects of a particular situation. For example, according to Emmons and Crumpler (2000), gratitude is the appreciation of an altruistic gift. This appreciation requires recognizing the gift, recognizing the goodness of the gift, recognizing the goodness of the giver, and recognizing the benefits of the gift that go beyond one's social expectations of others (Watkins, 2014). In other words, the extent of gratitude one experiences is often dependent on a combination of perceptions and evaluations of the value of the gift to the recipient, the cost of the gift to the giver, and the benevolence (i.e., altruistic intentions) of the giver toward the recipient of the gift. This multifaceted cognitive perspective goes beyond both the enduring trait and the purely transient affective state conceptualizations of gratitude.

Third is the *social/other-focused* perspective. Conceptualizations of this perspective vary in the literature along several dimensions. As indicated in the opening definition, the object of gratitude may another person's generosity, or it can be a less tangible "moment" of thankfulness or appreciation of a specific event, or of one's blessings and fortune in general (Emmons, 2004). When the object of gratitude is another person's actions, another social dimension of gratitude is whether it involves reciprocating the gift or benevolent act. Furthermore, if gratitude involves reciprocity, the next social dimension is whether reciprocity is directed toward the original giver or another recipient, often referred to as "paying it forward" or prosocial behaviors in general (*c.f.*, Ma et al., 2017).

The fourth and final perspective is the *conative/intentional* view of gratitude. In this view, a distinction is made between reactive gratitude, in which a person may experience gratitude as a result of the benevolence of others or being overwhelmed by life's abundant blessings, and a proactive attitude, which is more intentional in nature. Being proactively grateful entails consciously choosing to be grateful and intentionally finding ways to do so. Examples of intentional gratitude include purposefully counting one's blessings, enjoying life's simple pleasures, and stopping to smell the roses (Watkins et al., 2003). Indeed, the most recognized gratitude development interventions closely follow these purposeful and intentional practices (Seligman et al., 2005).

Importantly, the above four perspectives or views of gratitude can and do overlap. For example, cognitive appraisals and conative aspects of gratitude may be necessary to determine social reciprocity actions, and all views of gratitude are likely to produce positive emotions. Thus, an integrative and multidimensional perspective of gratitude is both necessary and more accurate than narrower definitions and conceptualizations (Morgan et al., 2017). Furthermore, synergistic integration of these four perspectives (affective, cognitive, social, and conative) has also been conceptually supported in the context of similar positive

psychological resources, such as psychological capital (PsyCap), a higher order construct that includes hope, efficacy, resilience, and optimism (Youssef and Luthans, 2013; Youssef-Morgan and Luthans, 2013; Luthans and Youssef-Morgan, 2017).

Integrating this four-pronged conceptualization of gratitude, we define *work gratitude* as "the intentional choice to engage in positive appraisals and feelings of thankfulness and appreciation toward the characteristics, situations, and people currently present in one's work context." Specifically, this definition synthesizes the conative (intentional choice), cognitive (positive appraisals), affective (feelings), and social (people) aspects of gratitude. Further, it takes into consideration that gratitude is a situational and context-specific state, rather than just a general disposition.

Relevant Approaches for Measuring Work Gratitude

The most commonly used measures of general gratitude are the *Gratitude Questionnaire* (GQ6; McCullough et al., 2002) and the *Gratitude, Resentment, and Appreciation Scale* (GRAT; Watkins et al., 2003). The GQ6 is a six-item scale that measures dispositional gratitude. Although dispositional gratitude is conceptualized in terms of frequency, intensity, span, and density, this scale assesses dispositional gratitude as a unidimensional factor (McCullough et al., 2002).

The GRAT also measures dispositional gratitude (Watkins et al., 2003). However, this 44-item scale, or its 16-item short version (Thomas and Watkins, 2003; Diessner and Lewis, 2007), are both multidimensional. They measure trait gratitude along three dimensions: sense of abundance (lack of sense of deprivation), simple appreciation (appreciation for simple pleasures), and appreciation for others (social appreciation).

Gratitude has also been measured as an affective state (Spence et al., 2014). This five-item scale measures state gratitude as a unidimensional, discrete emotion. While Spence et al. (2014) study is work-related (linked gratitude to organizational citizenship behaviors), the scale they developed is an measure of state gratitude in general, not work gratitude. Gratitude has also been conceptualized and measured as a multidimensional moral virtue (Morgan et al., 2017), with dimensions somewhat consistent with earlier views.

Recent studies demonstrate the relevance of gratitude in the workplace, both conceptually and empirically (Spence et al., 2014; Luthans et al., 2015; Fehr et al., 2017; Cain et al., 2019). However, measures of gratitude as a work-specific construct are lacking. A notable exception is Cain et al.'s (2019) recently developed Gratitude at Work Scale, which measures gratitude as a dispositional trait along two dimensions: gratitude for supportive work environment, and gratitude for meaningful work. While these two dimensions are highly relevant, we believe that they do not encompass the full range of conative, cognitive, affective, and social aspects of gratitude. Importantly, conceptualizing and measuring work gratitude as a dispositional trait or enduring tendency contradicts the extensive literature supporting the efficacy of short and simple gratitude interventions in changing participants' gratitude levels (including Study 3 of Cain et al., 2019), which point to the malleability and state-like nature of gratitude (Luthans et al., 2015). Because the scale focuses on gratitude as a dispositional trait, it does not cover the volitional or intentional aspects of gratitude. It simply asks participants to recount how often they are grateful for various aspects of the workplace (e.g., coworkers, supervisors, clients, salary and benefits, work-life balance, autonomy, accomplishments, and growth opportunities).

We draw from the PsyCap literature, where a multidimensional work-specific measure of a positive psychological construct was developed and validated. PsyCap is supported as a higher-order construct, with hope, efficacy, resilience, and optimism as lower-order constructs. The Psychological Capital Questionnaire (PCQ-24, Luthans et al., 2007) was developed by adapting items from dispositional measures of hope, efficacy, resilience, and optimism. The wording of each item was adapted in two distinct ways. First, because the original scales were designed to measure the trait counterparts of each resource, they were adapted by adding language that reflects the state-like, "here and now" timeframe typically experienced in the work context. This approach was also used by Snyder et al. (1996) in adapting the dispositional hope scale (Snyder et al., 1991) to create a state hope scale. Examples of language added include: "at the present time," "currently," and "right now." Second, with the exception of efficacy, the original scales were designed to measure general, not context-specific psychological constructs. Thus, the words "work," "at work," or "as it pertains to my work" were added to each item to make it work-specific.

Subsequently, a shorter, 12-item version (PCQ-12) of the PCQ-24 was developed (Avey et al., 2011). The PCQ-12 became widely used, particularly in cross-cultural research due to the reduced cost and increased ease of translation. One of the most notable changes is that the PCQ-12 does not include any reverse-scored items. Research shows that reverse-scored and negatively worded items tend to form a separate factor and change the dimensionality of a construct (Tomas and Oliver, 1999). Also they tend to lower the reliability of multi-item scales by as much as 20% (Barnette, 2000). This problem is particularly prevalent in measures of positive constructs, because seemingly opposite positive and negative constructs may not necessarily be opposite ends of a single construct, but rather distinct constructs (Peterson and Chang, 2002; Merritt, 2012). These method effects are evident in investigations of a number of established scales (DiStephano and Motl, 2006; Salazar, 2015). Thus, there is an abundance of evidence against the use of reverse-scored or negatively worded items.

The PCQ-12 was further adapted to other contexts beyond the workplace, such as health PsyCap, relationships PsyCap, overall (life in general) PsyCap (Luthans et al., 2013), and academic PsyCap (Martínez et al., 2019, 2021). Adaptation to most contexts is relatively easy. It can be accomplished simply by replacing the word "work" with the desired context. Context adaptations were not found to compromise the validity, reliability, or dimensionality of the scale.

TABLE 1 | Demographic characteristics of participants (N = 625).

Item	Category	Frequency (*f*)	Percentage (%)
Gender	Male	108	17.3
	Female	514	82.2
	Missing	3	0.5
Age	18–20	0	0
	21–25	122	19.5
	26–30	196	31.4
	31–35	177	28.3
	36–40	112	17.9
	41–50	17	2.7
	51+	0	0
Tenure (years)	0–5	262	41.9
	6–10	153	24.5
	11–15	98	15.7
	16–20	59	9.4
	21+	53	8.5
Occupation	Leadership team	43	6.88
	Administrative support staff	44	7.04
	Teacher	424	67.84
	Associate	58	9.28
	Food preparation and serving	12	1.92
	Building/grounds/maintenance	5	0.8
	Other	39	6.24

Based on the lessons learned from the development and adaptation of PsyCap measures, it follows that a plausible approach to developing, or at least generating an initial list of items for, a work gratitude scale (WGS) is to adapt the items in existing trait gratitude scales. This can be accomplished by adding wording to reflect (a) the "here and now" timeframe and (b) the "at work" context. Additionally, (c) the adapted items should be positively worded, rather than reverse-scored.

With this approach in mind, the purpose of this study was to develop, evaluate and validate a robust measure for work gratitude. Specifically, it aimed to explore the factorial structure, internal consistency, measurement invariance (gender), concurrent, convergent and discriminant validity of the WGS.

MATERIALS AND METHODS

Research Approach

A cross-sectional survey-based research design was utilized to collect the data for this study.

Participants

A convenience sample of 625 participants volunteered to take part in the study. Participants held a variety of positions in a medium-sized school district in the United States. At the time of the study, the school district employed about 900 employees, all of whom were invited to participate. **Table 1** provides a descriptive overview of the demographic characteristics of the sample.

Measures

The Work Gratitude Scale

The WGS was developed based on the approach employed by Luthans et al. (2007), by adapting existing trait measures of positive psychological constructs to the present, "here and now," and "at work" context. We also consulted recommended assessment approaches to-date for state and context-specific gratitude (McCullough et al., 2002; Wood et al., 2008a; Watkins et al., 2003), and heeded advice regarding avoiding negatively worded (reverse-scored) items, particularly when measuring positive psychological constructs (Tomas and Oliver, 1999; Barnette, 2000; Peterson and Chang, 2002; Merritt, 2012).

The initial pool of items was drawn from the GQ-6 (6 items) and the GRAT short version (16 items), which yielded 22 items for further consideration. Seven items were reverse-scored (negatively worded) and thus excluded, reducing the number to 15 items. Two items of the GRAT could not be adapted to the work context ("oftentimes I have been overwhelmed at the beauty of nature" and "every Fall I really enjoy watching the leaves change colors"). Thus, 13 items were adapted and used on the survey. Before survey administration, the 13 items were reviewed by an expert panel of positive psychology scholars to assess face validity and for final fine-tuning. The full list of items is shown in **Table 4** and discussed further in subsequent sections. Each item was rated on a seven-point Likert scale ranging from 1 (Strongly Disagree) to 7 (Strongly Agree). The final measure was comprised of 10 items (*c.f.,* **Appendix A**).

Concurrent Validity Measures

The GQ6 was used to measure the frequency, intensity and density of dispositional gratitude (McCullough et al., 2002). The six item, self-report questionnaire assesses dispositional gratitude as a unidimensional factor and is rated on a seven-point Likert type scale ranging from 1 ("Strongly Disagree") to 7 ("Strongly Agree"). A sample item is "I am grateful to a wide variety of people." The GQ-6 has shown to be a reliable measure in various contexts with an average Cronbach of 0.81 across 58 samples (Card, 2019).

The GRAT was used to measure dispositional gratitude as a stable trait (Watkins et al., 2003). The 16-item GRAT short version (Thomas and Watkins, 2003; Diessner and Lewis, 2007) is a multidimensional measure that assesses three components of gratitude: sense of abundance [lack of sense of deprivation. e.g., "There never seems to be enough to go around and I never seem to get my share (r)"], simple appreciation (appreciation for simple pleasures, e.g., "Oftentimes I have been overwhelmed at the beauty of nature"), and appreciation for others (social appreciation; e.g., "I feel deeply appreciative for the things others have done for me in my life") on a nine-point Likert scale ranging from 1 ("Strongly Disagree") to 9 ("Strongly Agree"). A meta-analysis has shown that the GRAT is a reliable measure in various contexts with an average Cronbach of 0.92 across 5 samples (Card, 2019).

Convergent Validity Measures

The Core Self-Evaluations Scale (CSES) was used to measure participants overall perception of self (Judge et al., 2003). The

12-item scale is a higher-order construct that measures four established personality traits: self-esteem ("Overall, I am satisfied with myself"), generalized efficacy (e.g., "I am confident I get the success I deserve in life"), locus of control ("I determine what will happen in my life"), and neuroticism ("There are times when things look pretty bleak and hopeless to me"). The CSES measures these as a unidimensional construct. Each item is rated on a five-point Likert type scale ranging from 1 ("Strongly Disagree") to 5 ("Strongly Agree"). The instrument has shown to have high levels on internal consistency with Cronbach Alphas greater than 0.80 and test-retest reliability of 0.81 (Gardner and Pierce, 2010).

The PCQ-12 was used to measure PsyCap (Avey et al., 2011). The 12-item scale measures a higher-order psychological capital construct that is comprised of four first order factors: hope (e.g., "Right now I see myself as being pretty successful at work"), efficacy (e.g., "I feel confident in representing my work area in meetings with management"), resilience (e.g., "I usually take stressful things at work in stride"), and optimism (e.g., "I always look on the bright side of things regarding my job"). Each item is rated on a six-point Likert type scale ranging from 1 ("Strongly Disagree") to 7 ("Strongly Agree"). The questionnaire has shown to be a reliable measure with McDonalds Omega's ranging from 0.72 to 0.90 on the various sub-scales (Rice et al., 2021).

Discriminant Validity Measures

Discriminant validity was assessed by relating the WGS to various demographic factors such as age (in years), gender, type of occupation, and tenure (in years).

Statistical Analyses

Data were analyzed using SPSS v26 (IBM SPSS, 2019), JASP v. 0.14.1 (JASP, 2021), and Mplus v 8.6 (Muthén and Muthén, 2021). A systematic, six-phased approach through structural equation modeling (SEM) was used to determine the factorial validity, internal consistency, measurement invariance with gender, concurrent, convergent and discriminant validity of the WGS.

First, to explore the factorial structure and item loadings of the WGS, an exploratory factor analytical (EFA) approach through SEM was employed. To determine the factorability of the instrument, the Kaiser-Meyer-Olkin (KMO) approach and Bartlett's sphericity test were estimated. According to Kaiser and Rice (1974), both a KMO value larger than 0.60 and a significant chi-square on Bartlett's sphericity assessment would indicate that meaningful factorial structures could be extracted from the data. Thereafter, a competing EFA modeling approach with the maximum likelihood estimation (ML) method, and a direct Oblimin rotation was used through the SEM framework. Here, competing exploratory factorial models were specified to be extracted from the data, based on Eigenvalues larger than 1 (Muthén and Muthén, 2021). Competing EFA models were compared based on model fit statistics with associated cut-off criteria (*c.f.,* **Table 2**), items were required to load statistically significantly (factor loading > 0.40; $p < 0.05$) on their respective extracted factors, cumulatively all factors needed to declare at least 50% total variance, and items should not represent multiple factors. Dual loadings

TABLE 2 | Model fit statistics.

Fit indices	Cut-off criterion	Sensitive to *N*	Penalty for model complexity
Absolute fit indices			
Chi-square (χ^2)	• Lowest comparative value between measurement models	Yes	No
	• Non-significant ($p > 0.01$)		
	• Significant difference in Chi-square between models		
	• For model comparison: retain model with lowest Chi-square		
χ^2/df	• <3 = Excellent and <5 = Acceptable	No	No
	• For model comparison: retain model with $\Delta\chi^2/df > 1$		
Approximate fit indices			
Root-Means-Square Error of Approximation (RMSEA)	• 0.06–0.08 (Marginally acceptable); 0.01–0.05 (excellent)	No	Yes
	• Not-significant ($p > 0.01$)		
	• 90% confidence interval range should not include zero		
	• For model comparison: retain model where ΔRMSEA ≤ 0.015		
Standardized Root Mean Square Residual (SRMR)	• 0.06–0.08 (Marginally acceptable); 0.01–0.05 (excellent)	Yes	No
	• For model comparison: retain model where ΔSRMR ≤ 0.015		
Incremental fit indices			
Comparative Fit Index (CFI)	• 0.90–0.95 (Marginally acceptable fit); 0.96–0.99 (excellent)	No	No
	• For model comparison: retain model with highest CFI value (ΔCFI > 0.01)		
Tucker-Lewis Index (TLI)	• 0.90–0.95 (Marginally acceptable fit); 0.96–0.99 (excellent)	No	Yes
	• For model comparison: retain model with highest TLI value (ΔTLI > 0.01)		
Akaike Information Criterion (AIC)	• Lowest value in comparative measurement models	Yes	Yes
Bayes Information Criterion (BIC)	• Lowest value in comparative measurement models	Yes	Yes

Adapted from Hu and Bentler (1999) and Wong and Wong (2020).

were systematically removed, and models were re-estimated (Wong and Wong, 2020).

Second, the factorial validity of the WGS was explored. Based on the best fitting EFA model, a competing confirmatory factor analytical measurement modeling strategy with the ML estimator was used to explore different, theoretically informed factorial permutations of the WGS. Both traditional independent cluster modeling confirmatory factor analytical (ICM-CFA) models and a bifactor model were estimated and sequentially compared. Observed items were used as indicators for latent factors. Observed indicators were only permitted to load onto their *a priori* theoretical factors and cross-loadings were not permitted. For the bifactor model, a single general factor (Work Gratitude) and three specific factors ("grateful appraisals," "gratitude toward others," and "intentional attitude of gratitude") were specified. All items were estimated to load on the general factor. For the specific factors, items were targeted to load onto their *a priori* factors. Here an orthogonal targeted rotation was used, and all covariances between specific factors were constrained to zero. To determine the best fitting model for the data, both the Hu and Bentler's (1999) criteria for model fit indices (*c.f.,* **Table 2** for the criteria) as well as indicators of measurement quality were employed. Measurement quality for the best fitting measurement models was assessed through evaluating the standardized factor loadings ($\lambda > 0.40$; $p < 0.05$), item uniqueness (>0.1 but <0.9; $p < 0.05$), and the presence of no multiple cross-loadings (Kline, 2010). Models that showed both excellent fit and measurement quality were retained for further analyses (McNeish et al., 2018).

Third, for the best fitting traditional ICM-CFA measurement models, item-level descriptive statistics (means, standard deviations, skewness, and kurtosis), standardized factor loadings, corrected item-total correlations (CITC), average variance extracted (AVE), and levels of internal consistency were estimated. To determine the multivariate normality of each item, Kim (2013) suggested that absolute values for skewness (<2) and kurtosis (<2) be employed for samples larger than 500. CITC represents each individual item's relationship to the overall factor on which it loads (Zijlmans et al., 2019). A CITC lower than $r = 0.30$ indicates that a specific item might not accurately represent the overall factor on which it is specified (Zijlmans et al., 2019). To determine the level of internal consistency of the WGS, three indicators were computed: point-estimate composite reliability (upper-bound; $\rho > 0.80$; Raykov, 2009), McDonald's Omega ($\omega > 0.80$; Hayes and Coutts, 2020), and Cronbach's alpha (lower-bound; $\alpha > 0.70$; Nunnally and Bernstein, 1994). Further, AVE was used as an indicator of the average level of reliability of each item within the scale. Here, levels of 50% or higher are deemed acceptable (Kline, 2010).

Additionally, for the bifactor model, the explained common variance (ECV), the item level explained common variance (IECV), the Average Relative Parameter Bias (ARPB), as well as Omega_s for the specific factor ($\omega_{specific} > 0.80$) and Omega_h ($\omega_{hierarchical} > 0.80$) for the general factor as indicators of reliability were computed. ECV refers to the proportion of total common variance explained by a general factor within bifactor models (Stucky and Edelen, 2015). For the specific factors, ECV represents the strength of the factor relative to all the explained variance within each specific factor. An ECV value for the general factor should exceed 0.50 (Stucky et al., 2013). Similarly, the IECV refers to "the extent to which an item's responses are accounted for by variation on the latent general dimension alone, and thus acts as an assessment of unidimensionality at the individual item level" (Stucky et al., 2013, p. 51) which should also exceed 0.50. ARPB represents an indicator of item bias within bifactor models. Here an ARPB value of less than 0.15 (i.e., 15%) is acceptable and therefore "poses no serious concern" (Rodriguez et al., 2016).

Fourth, we also investigated the factor equivalence or "measurement invariance" for gender (males vs. females). Here, increasingly restrictive equality constraints were placed on the best fitting measurement models. Configural (similar factor structures), metric (similar factor loadings) and scalar (similar intercepts) invariance models were estimated and sequentially compared. Invariance was established when the following criteria are met: (a) a non-significant difference in χ^2 between increasingly restrictive models, and (b) non-significant changes in RMSEA ($\Delta < 0.015$), SRMR ($\Delta < 0.015$), CFI (<0.01), TLI (<0.01), and χ^2/df (<1) between increasingly constrained models (Cheung and Rensvold, 2002; Reise et al., 2013). Once invariance was established, we proceeded to estimate latent mean differences between males and females. Here, the latent mean score for the reference group (males) was constrained to zero. The mean score for the female group was then freely estimated. If the latent mean score differed significantly ($p < 0.05$) from zero, it would indicate meaningful differences between genders (Wickrama et al., 2016; Wong and Wong, 2020).

Fifth, to establish convergent validity, the best fitting models of the WGS were then entered into a measurement model with both the GQ6 and the GRAT scales. The measurement models needed

TABLE 3 | Model fit statistics for competing exploratory factorial models.

Model	Type	χ^2	*df*	χ^2/df	CFI	TLI		RMSEA	SRMR	AIC	BIC	aBIC	Meets criteria
Model 0	One factor model	1365.23	65.00	21.00	0.78	0.74	0.18	[0.172–0.189]	0.09	19147.92	19320.36	19196.54	No
Model 1	Two factor model	744.81	53.00	14.05	0.88	0.83	0.15	[0.137–0.155]	0.06	18551.50	18777.01	18615.09	No
Model 2	Three factor model	188.791	42.00	4.50	0.98	0.95	0.08	[0.065–0.086]	0.02	18017.48	18291.62	18094.78	Partially
Model 3	Three factor model (removed items)	**23.45**	18.00	1.30	1.00	1.00	**0.02**	[0.000–0.045]	0.01	14260.60	14468.57	14319.35	Yes

χ^2 = Chi-square; df = degrees of freedom; TLI = Tucker-Lewis Index; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation [90% CI]; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; BIC = Bayes Information Criterion; aBIC = Adjusted Bayes Information Criterion. Bold: Non-significant p > 0.001.

to fit the data based on the criteria in **Table 3**, as well as produce a statistically significant ($p < 0.05$) standardized correlation between WGS and the GQ6 and GRAT scales (Kline, 2010).

Finally, to investigate concurrent validity and discriminant validity, structural models were used. Here two separate structural models were estimated for the ICM-CFA and bifactor models. For concurrent validity, the WGS (as an exogenous factor) was regressed on the CSES and the PCQ-12 (as endogenous factors). We established discriminant validity within the same structural model by estimating correlations between the WGS and predefined demographic factors (gender, age, occupation, and tenure). For the bifactor model, a similar approach was employed, however, both the general factor as well as the three specific factors were estimated to be related to the endogenous factors and correlated with the demographic factors. A significance level of $p < 0.05$ (95% confidence interval) was set for each regressive path and correlation.

RESULTS

Exploratory Factor Analysis

An EFA approach was employed to explore the factorial structure of the WGS. First, the KMO measure and Bartlett's test for sphericity were used to determine the factorability of the instrument. Results showed that meaningful factors could be extracted from the data because the KMO value was larger than 0.60 (KMO = 0.93) and a significant chi-square [$\chi^2_{(625)} = 5996.35$, $df = 78$, $p < 0.01$] was produced. We therefore proceeded to estimate the EFA models in Mplus.

Initially five factorial models were specified to be extracted from the data. Results showed that three factors could be extracted with eigenvalues larger than 1. Further, only three of the five models converged. Therefore, only the one, two, and three factorial models could be inspected (*c.f.,* **Table 3**). EFA Model 0, a *single first order factorial model* [$\chi^2_{(615)} = 1365.22$; $p < 0.01$; $df = 65$; $\chi^2/df = 21.00$; CFI = 0.78; TLI = 0.74; RMSEA = 0.18 [0.172, 0.189], $p < 0.01$; SRMR = 0.09; AIC = 19147.92; BIC = 19320.36; Eigenvalue = 7.57; $R^2 = 58.23\%$] did not fit the data and was therefore rejected. EFA Model 1, a *two first order factorial model* [$\chi^2_{(615)} = 744.814$; $p < 0.01$; $df = 53$; $\chi^2/df = 14.05$; CFI = 0.88; TLI = 0.83; RMSEA = 0.15 [0.137, 0.155], $p < 0.01$; SRMR = 0.06; AIC = 18551.50; BIC = 18777.01; Eigenvalue Factor 1 = 7.57; $R^2 = 58.23\%$; Eigenvalue Factor 2 = 1.25; $R^2 = 11.67\%$], also did not fit the model fit criteria, and was also rejected. Only EFA Model 2, the *three first order factorial model* [$\chi^2_{(615)} = 188.79$; $p < 0.01$; $df = 42$; $\chi^2/df = 4.95$; CFI = 0.98; TLI = 0.95; RMSEA = 0.08 [0.065, 0.086], $p < 0.01$; SRMR = 0.02; AIC = 18017.48; BIC = 18291.62; Eigenvalue Factor 1 = 7.57; $R^2 = 58.23\%$; Eigenvalue Factor 2 = 1.25; $R^2 = 11.67\%$; Eigenvalue Factor 3 = 1.02; $R^2 = 7.82\%$], fitted the data the best.

The three-factor EFA Model 3 also showed to fit the data significantly better than the one and two first-order factorial models. The item loadings and declared variance for this model are presented in **Table 4**. The results showed that three items (WGS3, WGS4, and WGS13) needed to be removed due to dual loadings, and therefore the three-factor model was respecified to produce new fit statistics. EFA Model 3, the *three first order factorial model with the three items removed* [$\chi^2_{(615)} = 23.45$; $p > 0.01$; $df = 18$; $\chi^2/df = 1.30$; CFI = 1.00; TLI = 1.00; RMSEA = 0.02 [0.000, 0.045], $p > 0.01$; SRMR = 0.01; AIC = 14260.60; BIC = 14468.57] fitted the data the best. **Table 4** shows that all the items loaded significantly on their respective factors, with factor loadings exceeding the 0.40 threshold. The first factor was labeled "grateful appraisals," the second factor "gratitude toward others," and the third factor "intentional attitude of gratitude." The Oblimin factorial correlation showed that all factors were strongly correlated (with a range of r between 0.59 and 0.66; $p < 0.01$).

Competing Confirmatory Factor Analytical Measurement Models

Next, a theoretically informed competing measurement modeling strategy was employed to further explore the factorial validity of the WGS. Measured items were used as indicators for latent factors, no items were removed, and error terms were not permitted to correlate. Four measurement models were estimated and compared:

- **Model 1**: A single first-order factorial model was specified where all 10 items loaded directly on to a single factor called "Work Gratitude."
- **Model 2**: A three first-order correlated factor model was estimated for the factors labeled "grateful appraisals" (comprised of three items: SWGS1, SWGS2, and SWGS6), "gratitude toward others" (comprised of four items: SWGS5, SWGS7, SWGS8, and SWGS12), and "intentional attitude of gratitude" (comprised of three items: SWGS9, SWGS10, and SWGS11).
- **Model 3**: A second-order factorial model comprised of the three first-order factors specified in the previous model was specified to directly load onto an overall "Work Gratitude" factor.
- **Model 4**: A bifactor model with one general Work Gratitude factor (on which all 10 items directly loaded) and three specific first-order factors (as mentioned in Model 2) was estimated.

The results, summarized in **Table 5**, showed that Models 2, 3, and 4 fitted the data best. Models 2 and 3 produced the same fit statistics [$\chi^2_{(625)} = 133.21$; $df = 32$; $\chi^2/df = 4.16$; CFI = 0.98; TLI = 0.97; RMSEA = 0.07 [0.059, 0.084]; SRMR = 0.07]. This is expected if the three factors fully represent a second-order factorial model. Further, the bifactor model, Model 4, fitted the data significantly better than the second-order factor model, Model 2 ($\Delta\chi^2 = -86.26$; $\Delta df = -7$; $\chi^2/df = -2.28$; ΔCFI = 0.02; ΔTLI = 0.03; ΔRMSEA = −0.03; ΔSRMR = −0.02; ΔAIC = −72.26; ΔBIC = −41.29). Both models further fitted the data significantly better than Model 1. Therefore, only Model 3 and 4 was retained for further inspection and analysis.

TABLE 4 | Exploratory factor analysis-factor loadings and variance.

Label	Item	EFA Model 2				EFA Model 3 (removed items)			
		λ_1	λ_2	λ_3	R^2 (%)	λ_1	λ_2	λ_3	R^2 (%)
Grateful appraisals									
WGS 1	Right now, I have so much at work to be thankful for.	**0.91**	−0.06	0.02	58.23	**0.94**	−0.06	−0.01	58.19
WGS 2	At this present time, if I had to list everything that I felt grateful for at work, it would be a very long list.	**0.81**	0.04	0.06	–	**0.83**	0.05	0.04	–
WGS 6	At the present time, life has been good to me at work.	**0.71**	0.12	0.01	–	**0.69**	0.12	0.00	–
Gratitude toward others									
WGS 5	Currently, I couldn't have gotten where I am today at work without the help of many people.	−0.05	**0.84**	−0.05	9.59	−0.03	**0.81**	−0.04	11.67
WGS 7	Although I think it's important to feel good about my current work accomplishments, I think that it's also important to remember how others have contributed to my accomplishments.	0.05	**0.85**	−0.01	–	0.08	**0.85**	−0.03	–
WGS 8	Although I'm basically in control of my work at the present time, I can't help but think about all those who have supported me and helped me along the way.	−0.07	**0.86**	0.10	–	−0.05	**0.89**	0.06	–
WGS 12	Right now, I feel deeply appreciative for the things others have done for me at work.	0.19	**0.51**	0.17	–	0.19	**0.49**	0.17	–
Intentional attitude of gratitude									
WGS 9	Currently, I think that it's important to "Stop and smell the roses" as it pertains to my work.	−0.10	0.04	**0.82**	7.82	−0.09	0.04	**0.82**	9.67
WGS 10	Currently, I believe that it's important to pause often to "count my blessings" at work.	0.06	0.06	**0.83**	–	0.09	0.07	**0.80**	–
WGS 11	Right now, I think it's important to enjoy the simple things that pertain to my work.	0.02	−0.06	**0.88**	–	0.03	−0.07	**0.89**	–
Double loadings									
WGS 3	At this time, I am grateful to a wide variety of people at work.	**0.38**	**0.47**	−0.01	–	–	–	–	–
WGS 4	Right now, I find myself able to appreciate the people, events, and situations that have been part of my work history.	**0.35**	**0.44**	0.05	–	–	–	–	–
WGS 13	Today, I believe it's important to appreciate each day at work.	**0.36**	0.07	**0.50**	–	–	–	–	–

λ = Factor Loading; R^2 = Variance. Bold = Significant loading with $p < 0.01$.

TABLE 5 | Competing confirmatory factor analytical models.

Model	Type	χ^2	*df*	χ^2/df	CFI	TLI	RMSEA		SRMR	AIC	BIC	aBIC	Meets criteria
Model 1	Single first-order factor model	1113.4	35	31.81	0.75	0.68	0.22	[0.212–0.235]	0.09	15316.6	15449.3	15354.07	No
Model 2	Three first-order factor model	133.21	32	4.16	0.98	0.97	**0.07**	[0.059–0.084]	0.04	14342.4	14488.4	14383.6	Yes
Model 3	Second-order factor model	133.21	32	4.16	0.98	0.97	**0.07**	[0.059–0.084]	0.04	14342.4	14488.4	14383.6	Yes
Model 4	Bifactor model	**46.95**	25	1.88	1.00	1.00	**0.04**	[0.020–0.054]	0.02	14270.10	14447.1	14320.10	Yes

χ^2 = Chi-square; df = degrees of freedom; TLI = Tucker-Lewis Index; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation [90% CI]; SRMR = Standardized Root Mean Square Residual; AIC = Akaike Information Criterion; BIC = Bayes Information Criterion; aBIC = Adjusted Bayes Information Criterion. Bold: Non-significant $p > 0.001$.

Item Level Descriptive Statistics, Standardized Factor Loadings, and Internal Consistencies

As shown in **Table 6**, all items were normally distributed (skewness and kurtosis <2; Kim, 2013). Each item was clearly associated with the overall factor being assessed (CITC $r > 0.30$; Zijlmans et al., 2019), and each of the three sub-factors and overall work gratitude was reliable at both the upper- ($\rho > 0.80$; $\omega > 0.80$) and lower- bound level of internal consistency ($\alpha > 0.70$). For the bifactor model, both the Omega for the specific factors and the Hierarchical Omega for the general factor were higher than 0.80.

For the second-order factorial model, Model 3, all items in each subscale loaded statistically significantly onto their respective *a priori* factors with standardized factor loadings ranging 0.74 to 0.93 ($\lambda > 0.40$; $p < 0.01$; Kline, 2010). Further, the AVE for both factors (and the overall Work Gratitude factor) was higher than the suggested 0.50 cutoff point (Kline, 2010). Further, all three first-order factors, grateful appraisals ($\lambda = 0.82$, SE = 0.02, $p < 0.01$), gratitude toward others ($\lambda = 0.79$, SE = 0.03, $p < 0.01$) and intentional attitude of gratitude ($\lambda = 0.86$, SE = 0.03,

TABLE 6 | Item level descriptive statistics, factor loadings and internal consistencies of the second-order factorial and bifactor models.

Factor	Item	Model 3–second order factorial model													Model 4–bifactor model											
							G_{factor}								S_{factor}											
		$\bar{x}$	σ	Skw	Kurt	CITC	λ	S.E.	R^2	δ	AVE	ρ	α	ω	Meets criteria	λ	S.E.	R^2	λ	S.E.	R^2	δ	IECV	ARPB	ECV	OmegaS	OmegaH	Meets criteria
Grateful appraisal											0.73	0.89	0.88	0.89	Yes										0.32	0.89	–	Yes
SWGS 1		5.72	1.09	−1.20	1.54	0.82	0.88	0.01	0.77	0.23	–	–	–	–	Yes	0.70	0.03	0.49	0.59	0.04	0.35	0.16	0.58	0.043	–	–	–	Yes
SWGS 2		5.42	1.28	−0.83	0.17	0.80	0.90	0.01	0.81	0.19	–	–	–	–	Yes	0.76	0.02	0.57	0.45	0.04	0.21	0.22	0.74	0.020	–	–	–	Yes
SWGS 6		5.80	1.03	−1.35	2.46	0.72	0.77	0.02	0.59	0.41	–	–	–	–	Yes	0.67	0.03	0.44	0.41	0.04	0.14	0.41	0.76	0.044	–	–	–	Yes
Gratitude toward others											0.67	0.89	0.88	0.88	Yes										0.36	0.90	–	Yes
SWGS 5		5.67	1.20	−1.16	1.47	0.71	0.76	0.02	0.57	0.43					Yes	0.55	0.03	0.30	0.54	0.04	0.29	0.41	0.51	0.155	–	–	–	Yes
SWGS 7		5.80	1.00	−0.99	0.99	0.80	0.88	0.01	0.77	0.24	–	–	–	–	Yes	0.68	0.03	0.47	0.55	0.04	0.30	0.24	0.61	0.097	–	–	–	Yes
SWGS 8		5.73	1.03	−1.02	1.17	0.81	0.89	0.01	0.79	0.21	–	–	–	–	Yes	0.68	0.03	0.46	0.59	0.03	0.35	0.19	0.57	0.109	–	–	–	Yes
SWGS 12		5.75	1.03	−0.96	1.08	0.67	0.74	0.02	0.55	0.45	–	–	–	–	Yes	0.73	0.03	0.53	0.40	0.04	0.06	0.41	0.90	0.025	–	–	–	Yes
Intentional attitude of gratitude																									0.31	0.89	–	Yes
SWGS 9		5.57	1.19	−1.13	1.73	0.73	0.77	0.02	0.59	0.41	0.72	0.88	0.88	0.88	Yes	0.61	0.03	0.38	0.49	0.04	0.24	0.38	0.61	0.096	–		–	Yes
SWGS 10		5.78	1.06	−0.85	0.71	0.80	0.93	0.01	0.86	0.14	–	–	–	–	Yes	0.81	0.02	0.66	0.42	0.04	0.17	0.17	0.79	0.014	–		–	Yes
SWGS 11		5.85	0.96	−0.96	1.26	0.79	0.84	0.02	0.71	0.29	–	–	–	–	Yes	0.70	0.03	0.49	0.51	0.04	0.26	0.25	0.65	0.053	–		–	Yes
Work gratitude		5.70	0.83	−0.62	0.22	–	–	–	–	–	0.68	0.82	0.82	0.82	Yes	–	–	–	–	–	–	–	–	–	0.67	–	0.82	Yes
Grateful appraisal		5.65	1.02	−0.96	0.80	–	0.82	0.02	0.67	0.33	–	–	–	–	Yes	–	–	–	–	–	–	–	–	–	–	–	–	Yes
Gratitude toward others		5.74	0.92	−0.76	0.60	–	0.79	0.03	0.62	0.38	–	–	–	–	Yes	–	–	–	–	–	–	–	–	–	–	–	–	Yes
Intentional attitude of gratitude		5.73	0.96	−0.77	0.26	–	0.86	0.02	0.74	0.26	–	–	–	–	Yes	–	–	–	–	–	–	–	–	–	–	–	–	Yes

$\bar{x}$, Mean; σ, Standard deviation; Skw, Skewness; Kurt, Kurtosis; CICT, Corrected item total correlation; λ, Standardized factor loadings; S.E., Standard Error; R^2, Variance; δ, Item Uniqueness; AVE, Average Variance Extracted; ρ, Composite Reliability; α, Cronbach's Alpha; ω, McDonald's Omega.

$p < 0.01$) loaded statistically significantly onto the second-order Work Gratitude factor.

For the bifactor factorial Model 4, similar trends were observed. All items loaded statistically significantly onto both the general and specific factors with standardized factor loadings ranging from 0.40 to 0.81 ($\lambda > 0.40$; $p < 0.01$; Kline, 2010). Additionally, the IECV for all items exceeded the 0.50 threshold (Stucky et al., 2013). Further, the ARPB was below 15% on each item, therefore, no item-related bias was evident (Rodriguez et al., 2016). The ECV for the general factor was larger than the suggested 0.50 threshold, however, those for the specific factors ranged from 0.31 to 0.36. This implies that the general factor for the WGS is more representative of overall Work Gratitude, than the individual factors. This lower level of ECV is still acceptable, given that the threshold for the General Factor exceeds the limits.

Therefore, both models showed excellent levels of measurement quality and were therefore retained for further analysis.

Measurement Invariance and Mean Comparisons

Next, the factorial equivalence of both Model 3 and Model 4 was tested with respect to males vs. females. The results, summarized in **Table 7**, showed that all models fitted the data based on the criteria described in **Table 3**. Further, no statistically significant differences in χ^2 as well as RMSEA ($\Delta < 0.015$), SRMR ($\Delta < 0.015$), CFI (< 0.01), TLI (<0.01), and χ^2/df (<1) between the configural, metric, scalar, and strict invariance models could be established (Cheung and Rensvold, 2002; Wong and Wong, 2020). Therefore, both models showed to be invariant between genders and meaningful mean comparisons can be made.

For Model 3's first-order factors, the results showed that no differences between males and females could be found for grateful appraisals ($\Delta\ \bar{x} = -0.02$; SE = 0.12; $p = 0.87$), gratitude toward others ($\Delta\ \bar{x} = 0.08$; SE = 0.13; $p = 0.50$) or intentional attitude of gratitude ($\Delta\ \bar{x} = 0.18$; SE = 0.12; $p = 0.12$). No mean differences for overall Work Gratitude between males and females could be established ($\Delta\ \bar{x} = 0.09$; SE = 0.12; $p = 0.44$).

Similar results were present for the Bifactor Model 4. No mean differences between males and females could be found on the three specific factors: grateful appraisals ($\Delta\ \bar{x} = -0.19$; SE = 0.28; $p = 0.48$), gratitude toward others ($\Delta\ \bar{x} = 0.02$; SE = 0.19; $p = 0.93$) or intentional attitude of gratitude ($\Delta\ \bar{x} = 0.17$; SE = 0.22; $p = 0.44$). No mean differences between genders on the general Work Gratitude could be established ($\Delta\ \bar{x} = 0.11$; SE = 0.17; $p = 0.53$).

Convergent Validity With the Trait Gratitude Scales (GQ6 and Gratitude, Resentment, and Appreciation Scale)

To establish convergent validity, the two best fitting factorial models of the WGS were entered into a measurement model with two closely associated other measures of gratitude. The GQ6 was estimated as a unidimensional factor, with all items loading directly onto an overall dispositional gratitude factor. Further, the GRAT scale was estimated as a multidimensional factor comprised of three first order factors namely simple appreciation. appreciation for others and lack of a sense of deprivation. All items were specified to load onto their prior theoretical factorial models.

The results in **Table 8** showed that for the second-order factorial Model 3 and the two trait gratitude scales, the data fitted the model adequately [$\chi^2_{(601)} = 1195.659$; $df = 452$; $\chi^2/df = 2.65$; CFI = 0.90; TLI = 0.90; RMSEA = 0.05 [0.049, 0.056]; SRMR = 0.07]. Both gratitude factors were strongly ($r > 0.50$) and statistically significantly ($p < 0.05$) related to the WGS.

For the bifactor Model 4 including the two trait gratitude scales, the data also fitted the model adequately [$\chi^2_{(601)} = 1118.47$; $df = 440$; $\chi^2/df = 2.54$; CFI = 0.91; TLI = 0.90; RMSEA = 0.05 [0.047, 0.054]; SRMR = 0.07]. The results showed that the general Work Gratitude factor was significantly related to both the GRAT ($r = 0.57$; $p < 0.05$) and the GQ6 ($r = 0.34$; $p < 0.05$). All the specific factors, excluding gratitude toward others ($r = -0.11$; $p = 0.08$), related statistically significantly to GQ6 and GRAT. However, these relationships ranged from small to marginal. The results imply that the second order factorial model seems to be better associated with the trait gratitude scales than the bifactor model.

Concurrent and Discriminant Validity

To establish concurrent and discriminant validity, separate structural models were estimated for the second-order Model 3 and the bifactor Model 4, and CSES and PCQ-12. For concurrent validity, CSES was specified as a unidimensional construct, where all items loaded directly onto a single factor. PCQ-12 was specified as second-order factorial model comprised of four specific factors (hope, efficacy, resilience, and optimism). In both models, demographic factors were used to establish discriminant validity. For concurrent validity, the WGS was regressed on both CSES and the second order PsyCap factor. For discriminant validity, the demographic factors were specified to correlate with the WGS.

The results in **Table 9** showed that the second-order factorial Model 3, with demographic factors, CSES and PCQ-12, fitted the data adequately [$\chi^2_{(607)} = 1722.41$; $df = 652$; $\chi^2/df = 2.64$; CFI = 0.90; TLI = 0.90; RMSEA = 0.05 [0.049, 0.055]; SRMR = 0.06]. The results showed that none of the demographic factors related significantly to Work Gratitude ($p > 0.05$). Further, Work Gratitude was positively and significantly related to CSES ($\beta = 0.38$; SE = 0.05; $p < 0.05$; $R^2 = 0.15$) and PCQ-12 ($\beta = 0.61$; SE = 0.04; $p < 0.05$; $R^2 = 0.37$). Both concurrent and discriminant validity could therefore be established for the second order factorial Model.

For the Bifactor Model 4, with demographic factors, CSES and PCQ-12, the results showed that the model fitted the data adequately [$\chi^2_{(607)} = 1358.03$; $df = 614$; $\chi^2/df = 2.22$; CFI = 0.91; TLI = 0.90; RMSEA = 0.05 [0.041, 0.058]; SRMR = 0.05]. The results showed that most of the demographic variables were not associated with neither the general nor the specific factors of the WGS, with three exceptions. Age ($r = 0.24$, $p < 0.05$) and tenure ($r = 0.18$, $p < 0.05$) were significantly related to intentional attitude of gratitude. Occupation was related to gratitude toward others ($r = -0.16$, $p < 0.05$). Therefore, discriminant validity

TABLE 7 | Measurement invariance for gender.

Model	χ^2	*df*	χ^2/df	CFI	TLI	RMSEA	SRMR	Model comparison	Δχ2	$\Delta\chi^2/df$	ΔCFI	ΔTLI	ΔRMSEA	ΔSRMR	Meets criteria
Second-order factorial model															
M1 Configural invariance	165.16	64	2.58	0.98	0.97	0.07 [0.058–0.085]	0.04	–	–	–	–	–	–	–	Yes
M2 Metric invariance: first order	182.66	71	2.57	0.97	0.97	0.07 [0.059–0.084]	0.06	M2 vs. M1	17.50	−0.01	0.00	0.00	0.00	0.02	Yes
M3 Metric invariance: second order	186.56	73	2.56	0.97	0.97	0.07 [0.059–0.084]	0.07	M3 vs. M2	3.90	−0.02	0.00	0.00	0.00	0.01	Yes
M4 Scalar invariance: first order	197.88	80	2.47	0.97	0.97	0.07 [0.057–0.081]	0.08	M4 vs. M3	11.32	−0.08	0.00	0.00	0.00	0.00	Yes
M5 Scalar invariance: second order	202.29	82	2.47	0.97	0.97	0.07 [0.057–0.081]	0.08	M5 vs. M4	4.41	−0.01	0.00	0.00	0.00	0.00	Yes
Bifactor model															
M1 Configural invariance	67.11	50	1.34	1.00	0.99	0.03 (0.000–0.053)	0.02	–	–	–	–	–	–	–	Yes
M2 Metric invariance	103.31	66	1.57	0.99	0.99	0.04 (0.026–0.058)	0.07	M2 vs. M1	36.20	0.22	−0.01	−0.01	0.01	0.05	Yes
M3 Scalar invariance	116.18	72	1.61	0.99	0.99	0.04 (0.029–0.059)	0.07	M3 vs. M2	12.87	0.05	0.00	0.00	0.00	0.00	Yes

χ^2 = Chi-square; df = degrees of freedom; TLI = Tucker-Lewis Index; CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation [90% CI]; SRMR = Standardized Root Mean Square Residual.

TABLE 8 | Convergent validity with the GQ6 and GRAT.

	Relationships	Standardized				Validity established
		r	S.E	*t*-value	*p*	
Second order model	WGS ←→ GQ6	0.50	0.06	8.97	0.00	Yes
	WGS ←→ GRAT	0.69	0.03	20.26	0.00	Yes
Bifactor model	Grateful appraisals ←→ GQ6	0.15	0.07	2.33	0.02	Yes
	Gratitude toward others ←→ GQ6	0.33	0.07	4.69	0.00	Yes
	Intentional attitude of gratitude ←→ GQ6	0.18	0.07	2.46	0.01	Yes
	Work gratitude ←→ GQ6	0.34	0.07	5.18	0.00	Yes
	Grateful appraisals ←→ GRAT	−0.11	0.06	−1.73	0.08	No
	Gratitude toward others ←→ GRAT	0.37	0.06	6.30	0.00	Yes
	Intentional attitude of gratitude ←→ GRAT	0.19	0.07	2.66	0.01	Yes
	Work gratitude ←→ GRAT	0.57	0.06	10.37	0.00	Yes

←→ = Correlation; r = correlation coefficient; S.E = Standard Error; p = statistical significance.

could not be established. However, concurrent validity for the bifactor Model 4 was established as all specific and general factors related positively and significantly to CSES and PCQ-12 ($p < 0.05$).

Therefore, the results show that only the second-order factorial model was supported to be both concurrently and discriminately valid.

DISCUSSION

Despite its numerous recognized benefits, to-date there are limited applications of gratitude in the workplace. Furthermore, current gratitude measures pose a number of challenges that hamper effective measurement of gratitude at work. This study sought to fill this gap by developing and evaluating a WGS. Integrating the extant gratitude literature, we defined work gratitude as "the intentional choice to engage in positive appraisals and feelings of thankfulness and appreciation toward the characteristics, situations, and people currently present in one's work context." We used a systematic, six-phased approach to determine the factorial validity, internal consistency, measurement invariance, concurrent, convergent and discriminant validity of the WGS. Furthermore, we compared first-order, second-order, and bifactor competing models for Work Gratitude.

The results supported a 10-item measure for a second-order factorial model of work gratitude comprised of three dimensions: "grateful appraisals" (three items), "gratitude toward others" (four items), and "intentional attitude of gratitude" (three items). This second-order factorial model showed significantly better model fit, measurement quality, internal consistency, measurement invariance in relation to gender, concurrent validity in relation to two existing dispositional gratitude scales (GQ-6; McCullough et al., 2002, and GRAT;

TABLE 9 | Concurrent and discriminant validity.

Relationships	Type of validity	Standardized						Validity established
		β	*r*	S.E	*t*-value	*p*	R^2	
Second order model								
Work gratitude ←→ Gender	Discriminant	–	0.02	0.04	0.45	0.65	–	Yes
Work gratitude ←→ Age	Discriminant	–	0.06	0.04	1.34	0.18	–	Yes
Work gratitude ←→ Occupation	Discriminant	–	−0.09	0.04	−2.24	0.06	–	Yes
Work gratitude ←→ Tenure	Discriminant	–	0.01	0.04	0.21	0.83	–	Yes
Work gratitude → CSES	Concurrent	0.38	–	0.04	9.13	0.00	0.15	Yes
Work gratitude → PCQ-12	Concurrent	0.61	–	0.04	17.19	0.00	0.37	Yes
Bifactor model								
Grateful appraisals ←→ Gender	Discriminant	–	−0.14	0.07	−2.04	0.06	–	Yes
Gratitude toward others ←→ Gender	Discriminant	–	−0.07	0.06	−1.07	0.28	–	Yes
Intentional attitude of gratitude ←→ Gender	Discriminant	–	−0.02	0.07	−0.31	0.75	–	Yes
Work gratitude ←→Gender	Discriminant	–	0.09	0.05	1.69	0.09		Yes
Grateful appraisals ←→ Age	Discriminant	–	0.04	0.06	0.70	0.49	–	Yes
Gratitude toward others ←→ Age	Discriminant	–	0.03	0.06	0.62	0.54	–	Yes
Intentional attitude of gratitude ←→ Age	Discriminant	–	0.24	0.07	3.56	0.00	–	No
Work gratitude ←→ Age	Discriminant	–	−0.02	0.05	−0.35	0.73	–	Yes
Grateful appraisals ←→ Occupation	Discriminant	–	−0.11	0.07	−1.62	0.11	–	Yes
Gratitude toward others ←→ Occupation	Discriminant	–	−0.16	0.06	−2.60	0.01	–	No
Intentional attitude of gratitude ←→ Occupation	Discriminant	–	−0.07	0.06	−1.13	0.26	–	Yes
Work gratitude ←→ Occupation	Discriminant	–	−0.01	0.05	−0.25	0.81	–	Yes
Grateful appraisals ←→ Tenure	Discriminant	–	0.01	0.08	0.10	0.92	–	Yes
Gratitude toward others ←→ Tenure	Discriminant	–	0.11	0.06	1.86	0.06	–	Yes
Intentional attitude of gratitude ←→ Tenure	Discriminant	–	0.18	0.08	2.41	0.02	–	No
Work gratitude ←→ Tenure	Discriminant	–	−0.01	0.06	−0.23	0.82	–	Yes
Grateful appraisals ←→ CSES	Concurrent	0.50	–	0.08	6.28	0.00	0.40	Yes
Gratitude toward others ←→ CSES	Concurrent	0.21	–	0.08	2.71	0.01	0.40	Yes
Intentional attitude of gratitude ←→ CSES	Concurrent	0.30	–	0.09	3.25	0.00	0.40	Yes
Work gratitude ←→ CSES	Concurrent	0.14	–	0.06	2.24	0.03	0.40	Yes
Grateful appraisals ←→ PCQ-12	Concurrent	0.63	–	0.07	8.68	0.00	0.66	Yes
Gratitude toward others ←→ PCQ-12	Concurrent	0.25	–	0.07	3.72	0.00	0.66	Yes
Intentional attitude of gratitude ←→ PCQ-12	Concurrent	0.30	–	0.08	3.61	0.00	0.66	Yes
Work gratitude ←→ PCQ-12	Concurrent	0.34	–	0.06	5.54	0.00	0.66	Yes

→ = Regression; ←→ = Correlation; β = Standardized Beta; r = correlation coefficient; S.E = Standard Error; p = statistical significance; R^2 *= Variance.*

Thomas and Watkins, 2003; Watkins et al., 2003), convergent validity in relation to core self-evaluations (Judge et al., 2003) and PsyCap (Luthans et al., 2007; Avey et al., 2011), and discriminant validity in relation to demographic factors (age, gender, occupation and tenure) than other *a priori* factorial permutations.

This study, therefore, supports work gratitude as a second-order construct with three underlying first-order dimensions. From this perspective, work gratitude is operationalized and measured as a function of (a) *grateful appraisals of work* (i.e., positive, cognitive appraisals of work characteristics and situations), (b) *gratitude toward others at work* (i.e., social appreciation toward the contributions of others at one's work), and (c) *an intentional attitude of gratitude* (i.e., purposefully enumerating, enjoying, and being mindful of positive aspects of one's work). The results showed that this multidimensional conceptualization of work gratitude is related to, yet empirically distinct from other existing gratitude measures. In line with the findings of Cain et al. (2019), this means that employees who are generally inclined to show gratitude toward others may or may not necessarily show feelings of gratitude at work, and those experiences of gratitude at work may not necessarily extend to other life domains. Thus, work gratitude is related to, but conceptually distinct from gratitude in other life domains, or gratitude in general.

This multidimensional measurement strategy is closely aligned with the gratitude literature (Watkins et al., 2003; Morgan et al., 2017), as well as the positive organizational literature (Youssef and Luthans, 2013; Youssef-Morgan and Luthans, 2013; Luthans and Youssef-Morgan, 2017). This three-dimensional self-other-environment characterization of gratitude is also consistent with established psychological frameworks such as social cognitive theory, where agentic actions occur at the intersection of self-reflection, observation

and learning from others, and influencing while being simultaneously influenced by one's environment (Bandura, 2001, 2012; Bandura and Locke, 2003).

Specifically, gratitude is not just a unidimensional, deterministic dispositional trait (McCullough et al., 2002), or a transient, momentary affective state (Spence et al., 2014). It involves intentional, cognitive, affective, social mechanisms through which one reflects upon, evaluates, and appreciates various aspects of a specific context, such as meaningful and enjoyable work experiences, as well as the contributions of leaders, mentors, and colleagues at work (Luthans et al., 2015). The agentic and intentional components of gratitude are essential. Here, an employee purposefully chooses to be mindful of these various positive aspects of work and react to them positively and gratefully, rather than taking them for granted. Lyubomirsky (2007) posits that about 50% of positivity is trait-based, and circumstances determine only 10%, but 40% is open to growth and development through one's intentional choices, thoughts and actions. This malleability and intentionality of gratitude is particularly relevant in the work context. For example, widely recognized gratitude development interventions (Davis et al., 2016) can be beneficial and effective if applied in the workplace to promote well-being, prosocial behaviors, and other desirable work outcomes. With the contribution of the current study, employee levels of work gratitude can be regularly assessed, monitored, and targeted for short and effective training interventions to promote grateful appraisals, gratitude toward others, and intentional attitudes of gratitude in employees.

Strengths and Limitations

Among the notable strengths of this study are sample size and the availability and utilization of highly relevant constructs with valid and reliable measures (GQ-6, GRAT, CSES, and PCQ-12) to facilitate item generation and assess the concurrent, convergent, and discriminant validity of the WGS. Another strength is the rigorous and systematic testing of competing models and the highly consistent results supporting the two-factor model.

An important limitation of this study is that the sample was drawn from an educational environment, namely a medium-sized school district in the United States. This may limit the generalizability of the findings to other contexts. More effort should be taken by future researchers to investigate the factorial structure of the instrument in other work-related contexts. Furthermore, the majority of participants were young, female, teachers, with short tenure, which limits the external validity of the findings. However, the positions of the participants still varied widely, resembling a wide range of jobs in other industries and professions. Males, older, and longer tenured participants were adequately represented. Measurement invariance was supported for gender, and there was sufficient variation in gender, age, tenure and occupation to support discriminant validity.

Another limitation is that this study did not test for predictive or incremental validity of the WGS. Future studies can utilize the WGS to predict meaningful work attitudes, behaviors, performance, and other important work outcomes. They can also incorporate established predictors of such outcomes to assess the incremental validity of the WGS.

Future Directions

Future research should examine the WGS in a wide range of work contexts such as manufacturing, services, and non-profit organizations of varying sizes to establish external validity. Furthermore, gratitude and other character strengths and virtues (Peterson and Seligman, 2004) are perceived and expressed differently across cultures. Thus, the WGS should be examined in other countries and cultures beyond the United States.

The availability of a valid and reliable measure of work gratitude is an important step in expanding the gratitude literature and research to the work context, linking it to strategic workplace initiatives such as human resource selection and development, and utilizing it to promote desirable work outcomes such as employee productivity, wellbeing, and prosocial behaviors. In terms of practical implications, there are many recognized and easy-to-implement gratitude interventions, which can be readily implemented in the workplace. Developing gratitude can yield highly desirable prosocial behaviors (Ma et al., 2017), which can promote employee wellbeing and a positive organizational culture.

Finally, the development of a context-specific measure of gratitude is an important step in operationalizing, measuring, and developing gratitude in other important life domains. Specifically, the WGS can be easily adapted to other contexts, by replacing the word "work" with other contexts of interest. This approach is similar to adaptations of the PCQ-12 to measure PsyCap in a variety of contexts (e.g., academic PsyCap, Martínez et al., 2019, 2021; relationship and health PsyCap, Luthans et al., 2013). Future research should rigorously examine and evaluate the psychometric characteristics of such adaptations of the WGS, but the availability of a context-specific measure of gratitude offers a valuable starting point.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Western IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CY-M was primarily responsible for the conceptual framework of the manuscript, as well as leading and integrating the work of the research team. LZ was primarily responsible for the data analysis. BA was primarily responsible for the data collection. All authors contributed to the article and approved the submitted version.

REFERENCES

Avey, J. B., Avolio, B. J., and Luthans, F. (2011). Experimentally analyzing the impact of leader positivity on follower positivity and performance. *Leadersh. Q.* 22, 282–294.

Bandura, A. (2001). Social cognitive theory: an agentic perspective. *Annu. Rev. Psychol.* 52, 1–26.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *J. Manag.* 38, 9–44.

Bandura, A., and Locke, E. (2003). Negative self-efficacy and goal effects revisited. *J. Appl. Psychol.* 88, 87–99. doi: 10.1037/0021-9010.88.1.87

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: if you feel the need, there is a better alternative to using those negatively worded stems. *Educ. Psychol. Measur.* 60, 361–370.

Cain, I. H., Cairo, A., Duffy, M., Meli, L., Rye, M. S., and Worthington, E. L. Jr. (2019). Measuring gratitude at work. *J. Posit. Psychol.* 14, 440–451.

Card, N. A. (2019). Meta-analyses of the reliabilities of four measures of gratitude. *J. Posit. Psychol.* 14, 576–586.

Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Model.* 9, 233–255.

Chopik, W. J., Newton, N. J., Ryan, L. H., Kashdan, T. B., and Jarden, A. J. (2019). Gratitude across the life span: age differences and links to subjective well-being. *J. Posit. Psychol.* 14, 292–302. doi: 10.1080/17439760.2017.1414296

Davis, D. E., Choe, E., Meyers, J., Wade, N., Varjas, K., Gifford, A., et al. (2016). Thankful for the little things: a meta-analysis of gratitude interventions. *J. Counsel. Psychol.* 63, 21–31. doi: 10.1037/cou0000107

Dickens, L. R. (2017). Using gratitude to promote positive change: a series of meta-analyses investigating the effectiveness of gratitude interventions. *Basic Appl. Soc. Psychol.* 39, 193–208. doi: 10.1080/01973533.2017.1323638

Diessner, R., and Lewis, G. (2007). Further validation of the gratitude, resentment, and appreciation test (GRAT). *J. Soc. Psychol.* 147, 445–447. doi: 10.3200/socp.147.4.445-448

DiStephano, C., and Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Struct. Equ. Model.* 13, 440–464.

Emmons, R. A. (2004). "Gratitude," in *Character Strengths and Virtues: A Handbook and Classification*, eds C. Peterson and M. Seligman (Oxford: Oxford University Press), 553–568.

Emmons, R. A., and Crumpler, C. A. (2000). Gratitude as a human strength: appraising the evidence. *J. Soc. Clin. Psychol.* 19, 56–69. doi: 10.1017/S1041610220000964

Emmons, R. A., and Mishra, A. (2011). "Why gratitude enhances well-being: what we know, what we need to know," in *Designing Positive Psychology: Taking Stock and Moving Forward*, eds K. M. Sheldon, T. B. Kashdan, and M. F. Steger (New York, NY: Oxford University Press), 248–262.

Fehr, R., Fulmer, A., Awtrey, E., and Miller, J. A. (2017). The grateful workplace: a multilevel model of gratitude in organizations. *Acad. Manag. Rev.* 42, 361–381. doi: 10.5465/amr.2014.0374

Froh, J. J., Bono, G., and Emmons, R. (2010). Being grateful is beyond good manners: gratitude and motivation to contribute to society among early adolescents. *Motiv. Emot.* 34, 144–157.

Gardner, D. G., and Pierce, J. L. (2010). The core self-evaluation scale: further construct validation evidence. *Educ. Psychol. Measur.* 70, 291–304.

Hayes, A. F., and Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But… *Commun. Methods Measur.* 14, 1–24. doi: 10.1080/19312458.2020.1718629

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

IBM SPSS (2019). *IBM SPSS Software Version 26*. Indianapolis, Ind: IBM Press.

JASP (2021). *Jeffery's Amazing Stats Program (v.0.14.1). [Statistical Software]*. Available online at: https://jasp-stats.org

Judge, T. A., Erez, A., Bono, J. E., and Thoresen, C. J. (2003). The core self-evaluations scale (CSES): development of a measure. *Pers. Psychol.* 56, 303–331. doi: 10.1111/j.1744-6570.2003.tb00152.x

Kaiser, H. F., and Rice, J. (1974). Little Jiffy, Mark IV. *Educ. Psychol. Measur.* 34, 111–117.

Karns, C. M., Moore, W. E., and Mayr, U. (2017). The cultivation of pure altruism via gratitude: a functional MRI study of change with gratitude practice. *Front. Hum. Neurosci.* 11:599. doi: 10.3389/fnhum.2017.00599

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor. Dent. Endodont.* 38, 52–54. doi: 10.5395/rde.2013.38.1.52

Kline, R. B. (2010). *Principles and Practices of Structural Equation Modelling*, 3rd Edn, New York, NY: Gilford Press.

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.

Luthans, F., Youssef, C. M., Sweetman, D. S., and Harms, P. D. (2013). Meeting the leadership challenge of employee well-being through relationship PsyCap and health PsyCap. *J. Leadersh. Organ. Stud.* 20, 118–133.

Luthans, F., and Youssef-Morgan, C. M. (2017). Psychological capital: an evidence-based positive approach. *Annu. Rev. Organ. Psychol. Organ. Behav.* 4, 339–366. doi: 10.1146/annurev-orgpsych-032516-113324

Luthans, F., Youssef-Morgan, C. M., and Avolio, B. (2015). *Psychological Capital and Beyond*. New York, NY: Oxford University Press.

Lyubomirsky, S. (2007). *The How of Happiness*. New York, NY: Penguin.

Ma, L. K., Tunney, R. J., and Ferguson, E. (2017). Does gratitude enhance prosociality? A meta-analytic review. *Psychol. Bull.* 143, 601–635. doi: 10.1037/bul0000103

Martínez, I. M., Meneghel, I., Carmona, M., and Youssef-Morgan, C. M. (2021). Adaptation and validation to Spanish of the psychological capital questionnaire–12 (PCQ-12) in academic contexts. *Curr. Psychol.* 40, 3409–3416.

Martínez, I. M., Youssef-Morgan, C. M., Chambel, M. J., and Marques-Pinto, A. (2019). Antecedents of academic performance in university students: psychological capital and academic engagement. *Educ. Psychol.* 39, 1047–1067. doi: 10.1080/01443410.2019.1623382

McCullough, M. E., Emmons, R. A., and Tsang, J. A. (2002). The grateful disposition: a conceptual and empirical topography. *J. Pers. Soc. Psychol.* 82:112. doi: 10.1037/0022-3514.82.1.112

McNeish, D., An, J., and Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cut-offs in latent variable models. *J. Pers. Assess.* 100, 43–52. doi: 10.1080/00223891.2017.1281286

Merritt, S. M. (2012). The two-factor solution to Allen and Meyer's (1990) affective commitment scale: effects of negatively worded items. *J. Bus. Psychol.* 27, 421–436. doi: 10.1007/s10869-011-9252-3

Morgan, B., Gulliford, L., and Kristjánsson, K. (2017). A new approach to measuring moral virtues: the multi-component gratitude measure. *Pers. Individ. Differ.* 107, 179–189.

Muthén, L. K., and Muthén, B. O. (2021). *Mplus (Version 8.4). [Statistical Software]*. Los Angeles, CA: Muthén and Muthén.

Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*, 3rd Edn, New York, NY: McGraw-Hill.

Peterson, C., and Chang, E. (2002). "Optimism and flourishing," in *Flourishing: Positive Psychology and the Life Well-Lived*, eds C. Keyes and J. Haidt (Washington, DC: American Psychological Association), 55–79.

Peterson, C., and Seligman, M. E. P. (2004). *Character Strengths and Virtues: A Handbook and Classification*. Oxford: Oxford University Press.

Portocarrero, F. F., Gonzalez, K., and Ekema-Agbaw, M. (2020). A meta-analytic review of the relationship between dispositional gratitude and well-being. *Pers. Individ. Differ.* 164:110101. doi: 10.1016/j.paid.2020.110101

Raykov, T. (2009). Evaluation of scale reliability for unidimensional measures using latent variable modeling. *Measur. Eval. Counsel. Dev.* 42, 223–232.

Reise, S. P., Scheines, R., Widaman, K. F., and Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling a bifactor perspective. *Educ. Psychol. Measur.* 73, 5–26.

Rice, S. P., Probst, T. M., and Bohle, S. L. (2021). Psychometric properties and measurement invariance of psychological capital, grit, and gratitude in selected samples from chile and the USA. *J. Well Being Assess.* 1–22. doi: 10.1007/s41543-021-00039-3

Richter, S., Van Zyl, L. E., Lara, C., and Stander, M. W. (2021). Positive psychological coaching tools and techniques: a systematic review and classification. *Front. Psychiatry* 12:667200. doi: 10.3389/fpsyt.2021.667200

Rodriguez, A., Reise, S. P., and Haviland, M. G. (2016). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychol. Methods* 21, 137–150.

Salazar, M. S. (2015). The dilemma of combining positive and negative items in scales. *Psicothema* 27, 192–199. doi: 10.7334/psicothema2014.266

Seligman, M. E. P., Steen, T. A., Park, N., and Peterson, C. (2005). Positive psychology progress: empirical validation of interventions. *Am. Psychol.* 60, 410–421.

Snyder, C. R., Harris, C., Anderson, J. R., Holleran, S. A., Irving, L. M., Sigmon, S. T., et al. (1991). The will and the ways: development and validation of an individual-differences measure of hope. *J. Pers. Soc. Psychol.* 60, 570–585. doi: 10.1037//0022-3514.60.4.570

Snyder, C. R., Sympson, S. C., Ybasco, F. C., Borders, T. F., Babyak, M. A., and Higgins, R. L. (1996). Development and validation of the state hope scale. *J. Pers. Soc. Psychol.* 70, 321–335.

Spence, J. R., Brown, D. J., Keeping, L. M., and Lian, H. (2014). Helpful today, but not tomorrow? Feeling grateful as a predictor of daily organizational citizenship behaviors. *Pers. Psychol.* 67, 705–738.

Stucky, B. D., and Edelen, M. O. (2015). "Using heierarchical IRT models to create unidimensional measures from multidimensional data," in *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*, eds S. P. Reise and D. A. Revicki (New York, NY: Routledge), 183–206.

Stucky, B. D., Thissen, D., and Edelen, M. O. (2013). Using logistic approximations of marginal trace lines to develop short assessments. *Appl. Psychol. Measur.* 37, 41–57. doi: 10.1177/0146621612462759

Thomas, M., and Watkins, P. (2003). Measuring the grateful trait: development of revised GRAT. *Paper Presented at the Annual Convention of the Western Psychological Association*, Vancouver, BC.

Tomas, J. M., and Oliver, A. (1999). Rosenberg's self-esteem scale: two factors or method effects. *Struct. Equ. Model.* 6, 84–98.

van Zyl, L. E., Arijs, D., Cole, M. L., Glinska, A., Roll, L. C., Rothmann, S., et al. (2021). The strengths use scale: psychometric properties, longitudinal invariance and criterion validity. *Front. Psychol.* 12:676153. doi: 10.3389/fpsyg.2021.676153

Watkins, P. C. (2014). *Gratitude and the Good Life: Toward a Psychology of Appreciation*. Dordrecht: Springer.

Watkins, P. C., Woodward, K., Stone, T., and Kolts, R. L. (2003). Gratitude and happiness: development of a measure of gratitude, and relationships with subjective well-being. *Soc. Behav. Pers. Intern. J.* 31, 431–451. doi: 10.1037/pas0000209

Wickrama, K., Lee, T. K., O'Neal, C. W., and Lorenz, F. (2016). *Higher-Order Growth Curves and Mixture Modeling With Mplus: A Practical Guide*. Milton Park: Routledge.

Wong, J., and Wong, X. (2020). *Structural Equation Modelling: Applications Using Mplus*, 2nd Edn, Chichester: Wiley & Sons.

Wood, A. M., Joseph, S., and Linley, P. A. (2007). Coping style as a psychological resource of grateful people. *J. Soc. Clin. Psychol.* 26, 1076–1093. doi: 10.1016/j.cjtee.2020.11.007

Wood, A. M., Joseph, S., Lloyd, J., and Atkins, S. (2009). Gratitude influences sleep through the mechanism of pre-sleep cognitions. *J. Psychosom. Res.* 66, 43–48. doi: 10.1016/j.jpsychores.2008.09.002

Wood, A. M., Maltby, J., Gillett, R., Linley, P., and Joseph, S. (2008a). The role of gratitude in the development of social support, stress, and depression: two longitudinal studies. *J. Res. Pers.* 42, 854–871. doi: 10.1016/j.jrp.2007.11.003

Wood, A. M., Maltby, J., Stewart, R., Linley, P., and Joseph, S. (2008b). A social-cognitive model of trait and state levels of gratitude. *Emotion* 8, 281–290. doi: 10.1037/1528-3542.8.2.281

Youssef, C. M., and Luthans, F. (2013). "Managing psychological capital in organizations: cognitive, affective, conative, and social mechanisms of happiness," in *Oxford Handbook of Happiness*, eds S. A. David, I. Boniwell, and A. C. Ayers (New York, NY: Oxford University Press), 751–766.

Youssef-Morgan, C. M., and Luthans, F. (2013). "Psychological capital theory: toward a positive holistic model," in *Advances in Positive Organizational Psychology*, ed. A. Bakker (Bingley: Emerald), 145–166. doi: 10.1108/s2046-410x(2013)0000001009

Zijlmans, E. A. O., Tijmstra, J., van der Ark, L. A., and Sijtsma, K. (2019). Item-score reliability as a selection tool in test construction. *Front. Psychol.* 9:2298. doi: 10.3389/fpsyg.2018.02298

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

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

Copyright © 2022 Youssef-Morgan, van Zyl and Ahrens. 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.

APPENDIX A: THE WORK GRATITUDE SCALE

Instructions: For the following set of questions, think about your current working environment. Consider, the nature and function of your work, your colleagues/peers and the role work plays in your life. Then indicate to what extent you agree with the following 10 questions.

No	Item	Strongly disagree	Disagree	Slightly disagree	Neutral	Slightly agree	Agree	Strongly agree
Grateful appraisals								
1	Right now, I have so much at work to be thankful for.	1	2	3	4	5	6	7
2	At this present time, if I had to list everything that I felt grateful for at work, it would be a very long list.	1	2	3	4	5	6	7
3	At the present time, life has been good to me at work.	1	2	3	4	5	6	7
Gratitude toward others								
4	Currently, I couldn't have gotten where I am today at work without the help of many people.	1	2	3	4	5	6	7
5	Although I think it's important to feel good about my current work accomplishments, I think that it's also important to remember how others have contributed to my accomplishments.	1	2	3	4	5	6	7
6	Although I'm basically in control of my work at the present time, I can't help but think about all those who have supported me and helped me along the way.	1	2	3	4	5	6	7
7	Right now, I feel deeply appreciative for the things others have done for me at work.	1	2	3	4	5	6	7
Intentional attitude of gratitude								
8	Currently, I think that it's important to "Stop and smell the roses" as it pertains to my work.	1	2	3	4	5	6	7
9	Currently, I believe that it's important to pause often to "count my blessings" at work.	1	2	3	4	5	6	7
10	Right now, I think it's important to enjoy the simple things that pertain to my work.	1	2	3	4	5	6	7

Scoring Create an average or "mean" score of the following items to create a score for each of the components of the GWS

1. **Grateful appraisals** = (Item 1 + Item 2 + Item 3)/3
2. **Gratitude toward others** = (Item 4 + Item 5 + Item 6 + Item 7)/4
3. **Intentional attitude of gratitude** = (Item 8 + Item 9 + Item 10)/3

To create an overall score of work gratitude, create an average score of the means for each of the aforementioned components

1. **Overall Gratitude at Work** = (Grateful appraisal + Gratitude toward others + Intentional attitude of gratitude)/3

frontiers in Psychiatry

METHODS
published: 07 January 2022
doi: 10.3389/fpsyt.2021.795672

Exploratory Structural Equation Modeling: Practical Guidelines and Tutorial With a Convenient Online Tool for Mplus

Llewellyn E. van Zyl[1,2,3,4]* *and Peter M. ten Klooster*[5]

[1] *Department of Industrial Engineering, University of Eindhoven, Eindhoven, Netherlands,* [2] *Optentia Research Focus Area, North-West University, Vanderbijlpark, South Africa,* [3] *Department of Human Resource Management, University of Twente, Enschede, Netherlands,* [4] *Institut für Psychologie, Goethe University, Frankfurt am Main, Germany,* [5] *Psychology, Health and Technology, University of Twente, Enschede, Netherlands*

OPEN ACCESS

Edited by:
Ashwani Kumar Mishra, All India Institute of Medical Sciences, India

Reviewed by:
Marco Guicciardi, University of Cagliari, Italy
Jose A. Piqueras, Miguel Hernández University of Elche, Spain
Herb Marsh, Australian Catholic University, Australia

***Correspondence:**
Llewellyn E. van Zyl
llewellyn101@gmail.com

Specialty section:
This article was submitted to Public Mental Health, a section of the journal Frontiers in Psychiatry

Received: *15 October 2021*
Accepted: *30 November 2021*
Published: *07 January 2022*

Citation:
van Zyl LE and ten Klooster PM (2022) Exploratory Structural Equation Modeling: Practical Guidelines and Tutorial With a Convenient Online Tool for Mplus. Front. Psychiatry 12:795672. doi: 10.3389/fpsyt.2021.795672

Critics of positive psychology have questioned the validity of positive psychological assessment measures (PPAMs), which negatively affects the credibility and public perception of the discipline. Psychometric evaluations of PPAMs have shown that various instruments produce inconsistent factor structures between groups/contexts/times frames, that their predictive validity is questionable, and that popular PPAMs are culturally biased. Further, it would seem positive psychological researchers prioritize date-model-fit over measurement quality. To address these analytical challenges, more innovative and robust approaches toward the validation and evaluation of PPAMs are required to enhance the discipline's credibility and to advance positive psychological science. Exploratory Structural Equation Modeling (ESEM) has recently emerged as a promising alternative to overcome *some* of these challenges by incorporating the best elements from exploratory- and confirmatory factor analyses. ESEM is still a relatively novel approach, and estimating these models in statistical software packages can be complex and tedious. Therefore, the purpose of this paper is to provide novice researchers with a practical tutorial on how to estimate ESEM with a convenient online tool for Mplus. Specifically, we aim to demonstrate the use of ESEM through an illustrative example by using a popular positive psychological instrument: the *Mental Health Continuum-SF*. By using the MHC-SF as an example, we aim to provide (a) a brief overview of ESEM (and different ESEM models/approaches), (b) guidelines for novice researchers on how to estimate, compare, report, and interpret ESEM, and (c) a step-by-step tutorial on how to run ESEM analyses in Mplus with the De Beer and Van Zy ESEM syntax generator. The results of this study highlight the value of ESEM, over and above that of traditional confirmatory factor analytical approaches. The results also have practical implications for measuring mental health with the MHC-SF, illustrating that a bifactor ESEM Model fits the data significantly better than any other theoretical model.

Keywords: exploratory structural equation modeling, psychometrics, factor analysis, mental health continuum, statistical tutorials

INTRODUCTION

Positive psychology emerged in the late 1990s to counterbalance the dominating psychopathological focus of the time (1). In the 22 years since its inception, positive psychology's strive to apply the scientific method to investigate the positive states, -traits, and -behaviors that enhance mental health, has spawned a magnitude of new theories, models and constructs (2, 3). The growth of the discipline resulted in a rapid rise in the development and use of "positive psychological assessment measures" (PPAMs) aimed at measuring these positive psychological constructs validly and reliably (4). However, critics of positive psychology have questioned the validity of PPAMs (5–8). Psychometric evaluations of PPAMs have shown that various instruments produce inconsistent factor structures, that reliability estimates vary significantly between groups/contexts/times frames, that the predictive validity is questionable and that popular PPAMs are culturally biased [cf. (9–11)]. Although these challenges apply to all self-report psychometric instruments aimed at measuring psychological phenomena, it is particularly damaging to the discipline as it fuels current scientific critiques of positive psychology [cf. (12, 13)]. These critiques, in turn, negatively affect the credibility and public perception of the discipline.

A typical example of these challenges can be seen with the *Mental Health Continuum-Short Form* [MHC-SF; (14)]. The MHC-SF is one of the most popular PPAMs aimed at measuring mental health, has shown to produce various factorial models ranging from a correlated three first-order factorial model (comprised of emotional-, psychological-, and social well-being) through to various types of bifactor models with varying ranges of reliability (15–18). Zemojtel-Piotrowska et al. (18) also showed that the MHC-SF is not equivalent between cultures and required various modifications of the factorial model to ensure that partial invariance could be established. Further, the item relating to "positive relationships" on the psychological well-being sub-scale, is strongly related to the social well-being subscale and has shown to load on both constructs in several contexts [cf. (18)]. In individualistic cultures, a clear distinction between these three factors is apparent, however in collectivistic cultures psychological- and social well-being seem to be tied more closely together (16, 18). Therefore, limiting the cross-cultural comparisons which could be made with the instrument. A final conceptual issue also pertains to how mental health is defined vs. how it's measured. Keyes (19) indicated that mental health lies on a continuum ranging from languishing to flourishing. However, the MHC-SF measures mental health as a function of "a dynamic interaction between three factors" classified into three categories (languishing, moderate mental health, and flourishing). When cross-loadings are constrained to zero in its estimation, no "dynamic interaction" between these factors can be captured. Further, the categorization of mental health into categories is not aligned with the idea that mental health ranges on a continuum. Therefore, there is a disconnect between the conceptual formation of mental health as a continuum and the psychometric measurement (or estimation) thereof as categorical. Given that factors like mental health cannot directly be observed but only inferred through behavioral observation, Morin et al. (20) argued that the approach employed to (analytically) explore and validate instruments measuring these factors may be at the core of the issue.

Marsh et al. (21) argued that behavioral observation in psychological research usually takes the form of recorded responses to observed indicators (items on questionnaires), reflecting the overall, unobserved latent factor it is supposed to be measuring. Factor analysis was therefore developed to explore and represent these psychological constructs through constructing latent factors that are seen as the "underlying cause of these behaviors" [(20), p. 1,044]. Although a variety of multivariate factor analysis techniques exist to model and explore the factorial structures of constructs, psychological research has broadly adopted exploratory factor analyses (EFA) and confirmatory factor analysis (CFA) as its methods of choice (20–22).

EFA refers to a set of statistical techniques used to identify or uncover the smallest number of relevant dimensions needed to explain the covariation amongst a set of measured items or variables (23). In other words, EFA aims to identify common factors in data that explains the order and structure amongst measured items (21). EFAs allow for factors to be freely estimated by the available data, and cross-loadings are permitted to achieve a simple and interpretable factorial solution (23). EFAs are not without their criticisms and limitations. EFAs cannot incorporate or control for method effect (24). For example, when two relatively similarly worded items are present in a questionnaire, the covariance between these cannot entirely be explained only by their relationship with the latent construct; a residual correlation would need to be added. Further, within the EFA framework, scores produced by an instrument cannot directly be compared with scores produced by other groups or even over time (23). Direct comparisons are only possible if the item and factor loadings are precisely the same for both groups (which in practice is unlikely). This further implies that factorial equivalence or measurement invariance cannot be estimated or compared (25). A final major limitation is that EFA is data-driven, limiting its usefulness to applied researchers wanting to conduct more complex analysis (21, 26).

In contrast, CFA was developed by Joreskog (26) as a theory-driven approach whereby factor structures rely purely on an a priori specification of unique items onto their respective latent factors. In other words, CFAs aim to explore how well a predefined theoretical model "fits" the data that has been collected. Here, researchers formulate several clear hypotheses about the nature of a construct (e.g., how many factors it comprises of, whether factors are related or not, which items load onto which construct etc.) before data collection or analysis (23, 27). These assumptions are then tested against the data, and different or "alternative" theoretically informed models are sequentially computed and compared to determine which fits the data best. Within CFAs, items are forced to be only related to one specific latent factor, whereby loadings on other factors are constrained to be zero (28). Unlike EFAs, CFAs actively model and incorporate item uniqueness and correct for measurement error (20). Further, CFA models tend to produce more parsimonious models, where latent variables are

easier to understand and interpret (29). CFAs have become more dominant in their use over the last three decades due to the advent of Structural Equation Modeling (SEM) and more powerful computers to process data (21, 28). This allows researchers to model complex data and construct more "accurate" models of behavior given real-world scenarios (21).

Although CFAs within the SEM framework are currently probably the most widely used method to examine the factorial structure of an instrument in psychological research (20, 29), it is not without its limitations. Conceptually, CFA models are often overly simplistic, restrictive and idealistic as it assumes "pure factors," where items only load onto their a priori latent factors (i.e., cross-loadings are constrained to zero) (21, 28, 30). Given that most items on psychological measures tend to measure more than one conceptually related factor, some degree of construct-relevant association between items can and should be expected (31). This naturally leads to several significant, yet small cross-loading items on non-target factors (29). Forcing items to then only load on one a priori latent factor and constraining cross-loadings to zero leads to a more parsimonious model but artificially inflates the associations of items with factors (27, 28). This, in turn, would lead to inflated model fit statistics and inflated measurement quality indicators, which results in positively biased factor correlations; unless all non-target factor loadings are close to zero (21, 27). Simulation studies have shown that even small cross-loadings need to be considered to avoid inflated parameter estimates and biased results (22). Therefore, this positive bias, along with constraining loadings to zero, could undermine the discriminant validity of more restrictive CFA models, as correlations between factor indicators is forced to only go through their main factors (22, 24). In practice, this distorts how the interrelationships between the constructs are interpreted and, therefore, also their meaning (22).

Another issue relates to the goodness-of-fit indices [cf. (32)], which CFA models rely on. These fit indices are usually too restrictive when applied to multi-factor psychological instruments. Therefore, it is almost impossible to achieve "good" data-model fit without significant modifications to the factorial models (33). However, when looking at item level indicators and measures of reliability, these same models that produce "bad fit" can produce reasonable item loadings and high levels of reliability (33–35). Researchers then tend to incorporate more dubious exploratory, data-driven, approaches within the CFA framework to enhance data-model fit, such as correlating residual error terms on items, item parceling, HARKing, or constraining paths to be equal (21). Similarly, various studies have anecdotally shown discrepancies between the reported EFA and CFA results, which cannot solely be due to multiple cross-loadings that were not correctly modeled [cf. (29, 31)]. Therefore, traditional CFA approaches do not seem to fit psychological constructs that well. This poses several challenges for positive psychological research as this positive bias undermines support for (a) the multidimensional view of psychological constructs and instruments assuming to measure such, (b) the discriminant validity of PPAM, (c) the predictive validity of psychometric instruments based on high levels of multi-collinearity, and (d) the practical, diagnostic usefulness of an instrument (22).

More innovative and robust approaches to validating and evaluating psychometric instruments are required to address these analytical challenges. Applying more innovative and flexible approaches to evaluating PPAMs could enhance the discipline's credibility and advance positive psychological science. Recent developments in the field have started to use more flexible approaches to factor analysis such as Bayesian estimation [cf. (36) for a gentle introduction] and even incorporating EFA approaches into CFA models through SEM in order to capitalize upon the strengths of both (20). One of these promising alternatives to overcome the restrictions posed by the aforementioned analytical frameworks is Exploratory Structural Equation Modeling (ESEM) (30).

Exploratory Structural Equation Modeling

ESEM was developed to incorporate the best elements of both CFAs [e.g., predictive relationships between factors (adjusted for measurement error), can produce method factors, correlated item uniqueness, estimate complex error variance structures, and produce bifactor models, estimate measurement invariance, and even be specified into auto-regressive models) and EFAs (e.g., allowing cross-loadings] into the traditional SEM framework (30). Therefore, ESEM provides a compromise between the mechanical iterative approach toward finding optimal factorial solutions through rotations within an EFA and the restrictive a priori theory-driven modeling approach employed within CFA measurement models (20).

Marsh et al. (22) stated that ESEM is fundamentally a confirmatory technique (although it can be used in an exploratory way), which through a target rotation, makes it possible to model data in a confirmatory way by allowing for the presence of cross-loadings between items. Although permitted, cross-loadings (non-target loadings) are constrained to be as close to zero as possible (30). Drawing from CFAs, within the ESEM framework, the researcher has more a priori control over the expected factor structure of an instrument. Further, how ESEM models identify mean structures is typically similar to traditional CFA models where item intercepts are estimated freely and latent factor means are constrained to zero (22, 30). Given that a CFA model is also nested within an ESEM model, both models can directly be compared through traditional model fit indices (22). When an ESEM solution fits the data better than a traditional CFA model, the estimated factor correlation is likely to be substantially less biased than in the CFA model (22).

ESEM incorporates more flexible EFA models into its model estimation by allowing items to cross-load on non-target factors. However, this means that the rotation method employed is critically important as the size and direction of the estimated factor correlations differ depending on the type of rotation (22). Rotation procedures are required for model identification but are employed to simplify the interpretability of the factors which ESEM/EFAs tend to estimate (37). The choice of rotation procedure directly affects the estimated factor correlations and cross-loadings (37). Xiao et al. (29) indicated that the three most popular rotation methods employed in ESEM are the (oblique) geomin- and target rotations, with orthogonal rotations being used for bifactor ESEM models. Asparouhov and Muthén

(30) stated that geomin rotations, where correlations between factors are estimated and incorporated, generally perform well if the estimated model isn't too complex. On the other hand, target rotations allow ESEM model estimations to be used more confirmably but depend on the a priori assumptions made about how cross-loadings are specified (30). Target rotations do not require a researcher to specify "anchor items" with non-targeted factor loadings (22) and provide more control in specifying models (30). This implies that more complex models can be estimated when using target rotation and should be preferred (30). In bifactor ESEM models, the general (G) factor and the specific (S)factors need to be specified as totally independent from one another, and therefore the relationships between factors and variances shared need to be constrained through the orthogonal rotation (37).

ESEM poses several advantages over and above those of the traditional CFA and EFA approaches. Morin et al. (20) and Marsh et al. (22) argued that ESEM is more robust, rigorous, and flexible than most analytical approaches as (a) it can simultaneously estimate both CFA and EFA models, (b) it can estimate less restrictive measurement models that permit cross-loadings which can produce useful-fit indices and parameter estimates, (c) it usually fits the data significantly better than traditional CFA and EFA models, (d) latent factor correlations are less biased and are closer to the true associations and most importantly, (e) these models are usually also more in line with the theoretical conceptualization and considerations of the construct the instruments intend to measure. Van Zyl et al. (38) also showed that ESEM models could potentially compensate for wording effects and cross-cultural differences in the interpretation of items when comparing different national or cultural groups. It is, however, essential to note that ESEM does not necessarily increase the instrument's reliability and that, despite improving model-fit, researchers should always carefully inspect all item-level parameters. Further, various typical simple and more complex CFA model permutations can also be estimated within the ESEM framework, such as first-order-, hierarchical-, and bifactor models (39, 40). As such, ESEM could be used for several purposes such as scale construction (like traditional EFA), refinement and validation (like traditional CFAs), and replication (40).

Although there are several advantages of ESEM, it also has several limitations. Traditional (first-order) ESEM models cannot easily be used in more complex, predictive, or hierarchical models (20, 41). For example, Morin et al. (24) argued that the bootstrapped confidence intervals required to provide support for the indirect effect of a mediator on the relationship between an exogenous and endogenous factor cannot be generated with ESEM models. In other words, one would for instance, not be able to determine how mental health indirectly affects the person-environment-performance relationship. Normal first-order ESEM factors can also not meaningfully be used as indicators for higher-order factors which limits its use in, for example, full latent growth curve models (42). Morin et al. (20) argued that to do so, the higher-order factorial model should be constructed based on the correlations of the first-order factors. However, this provides nothing more than just a simple expression of these inter-factor correlations that do not accurately represent the hierarchical nature of a multidimensional construct. Current estimation procedures for ESEM models, also do not support multilevel- or mixture modeling (39) nor mixture models (42), thus limiting its use in, for example, daily-diary intervention studies. According to Marsh et al. (39) it is also not presently possible to accurately estimate partial factorial invariance. Morin et al. (42) also indicate that latent means can't be constrained in multi-group models, and therefore comparisons between (for example) genders/ages/cultures on mental health is not possible. Marsh et al. (39) also mentioned that within ESEM, multiple sets of (unrelated) factors cannot be estimated simultaneously, as permitting for cross-loadings between these factors (e.g., mental health vs. performance) would undermine the theoretical foundation of both factors. Further, full ESEM models may lack parsimony and that the popular (dubious) approaches to circumvent such used within CFAs (such as item parceling, or using manifest scale scores) could not be used to ensure convergence (22, 39). Another limitation is that using ESEM models within structural models may present convergence and estimation problems (24). For the applied researcher, ESEM models may therefore not be useful above and beyond exploring the factorial validity of an instrument.

To address these issues, and circumvent the limitations of ESEM, *set-ESEM* (39), and *ESEM-within-CFA* (22, 24) was developed. *set-ESEM* allows for the modeling of two or more distinct "sets" of constructs within a single (ESEM) model, where cross-loadings between items are allowed for (first-order) factors that are related to the same construct (or set) but constrained to zero for constructs of different sets (like within a traditional CFA model) (39). These sets could reflect the same construct at different time stamps in longitudinal models or different constructs measured simultaneously within cross-sectional data. For example, if common mental health problems [stress, depression, and anxiety as measured by the DASS-21: (43)] and mental health [emotional-, psychological-, and social well-being as measured by the MHC-SF: (14)] are estimated within set-ESEM, then both "sets" of factors would be modeled simultaneously. Here, the first-order latent factors would be permitted to covary, and cross-loadings between the DASS-21 factors and cross-loadings between the MHC-SF factors would be permitted. However, unlike within a full ESEM model, items from the DASS-21, would not be permitted to cross-load with the MHC-SF and vice versa. Set-ESEM allows for the simultaneous estimation of multiple constructs and finds an optimal balance between CFAs and ESEMs in respect of parsimony, data-model fit, rigor, and well-defined factor estimation (39). set-ESEM therefore also maintains the structural (theoretical) integrity of each set of ESEM models whilst allowing for more flexibility in estimation. Set-ESEM is, however, still a relatively new approach within the ESEM lexicon, and its full practical usefulness needs to be explored. Its therefore beyond the scope of this tutorial to fully reflect upon the technique [interested readers are referred to (39) for a non-technical overview of set-ESEM].

ESEM-within-CFA, on the other hand, assumes that the resulting measurement structure of an ESEM factor model would

remain stable when transformed into a CFA model (39). Within this framework, a first-order ESEM model is re-expressed within a CFA framework by using the (unstandardized) factor loadings of the ESEM model as starting values (42), and factor variances are freely estimated (44). The original ESEM solution and the respecified ESEM-within-CFA solution should produce precisely the same chi-square, degrees of freedom, model-fit statistics, and parameter estimates (standard errors may be slightly inflated, though) (42). By expressing an ESEM model within the ESEM-within-CFA framework, more "traditional" models and analyses can be conducted. For example, hierarchical or "second-order factor" ESEM models of mental health could be constructed where mental health is a function of these three first-order ESEM factors. Morin et al. (24) for example showed that partial mediation could be estimated with the ESEM-within-CFA model and that bias-corrected bootstrapped confidence intervals could be produced. Morin et al. (42) also showed that the ESEM-within-CFA solution could be used to show how factors change over time, longitudinal mediation and could even be used to estimate latent change score models. Further, Howard et al. (45) also showed that an ESEM-within-CFA model could be used in normal structural models to establish the relationships between factors. Therefore, the ESEM-within-CFA framework makes it possible for applied researchers to readily use ESEM models for more complex research questions.

Given these advantages, it's therefore not surprising that the use of ESEM is gaining popularity within the positive psychological sciences (46–49). However, there are several challenges its use poses for positive psychological researchers:

1. The ESEM approach is complex, and its usefulness is challenging to articulate to applied scientists.
2. Given that it is a relatively recent development applied scientists may find it difficult to understand when, where, why, and how to use ESEM and find it challenging to understand what the results mean and what to report. To the best of our knowledge, no best practice guidelines for estimating and reporting ESEM are easily accessible to the average researcher.
3. Finally, there are currently only two software packages that can estimate ESEM models: R Studio and Mplus (50). Mplus is currently the only software package that fully integrates ESEM, whereas R currently only provides partial implementation (37). Estimating ESEM models in either software package requires complex code or syntaxes to run. Further, especially estimating ESEM-within-CFA models is not only complex, but extremely tedious and time-consuming.

These three challenges may significantly hamper researchers to adopt ESEM as an alternative to traditional EFA and CFA modeling strategies. Specifically, the perceived complexity and unfamiliarity with the approach and its estimation procedure may reduce the probability of less experienced researchers exploring or using these alternative factor analytical techniques. Therefore, the purpose of this paper is to provide novice researchers with a practical tutorial on how to apply ESEM with an innovative tool for Mplus. Specifically, we aim to demonstrate the use of ESEM through an illustrative example by using a popular positive psychological instrument: the *Mental Health Continuum-Short Form* [MHC-SF: (14)]. By using the MHC-SF as an example, we aim to provide (a) a brief overview of ESEM (and different ESEM models), (b) guidelines for novice researchers on how to estimate, compare, report and interpret ESEM models, and (c) a step-by-step tutorial on how to run ESEM analyses in Mplus with an easy to use online tool for syntax generation.

The Mental Health Continuum: An ESEM Perspective

Mental health is a foundational component in the positive psychological lexicon and of keen interest to researchers and practitioners alike (1, 51, 52). Therefore, mental health is a popular and familiar framework that applied positive psychological researchers can relate to and thus an interesting concept to use to illustrate ESEM.

Mental health is defined by the World Health Organization [(53), p. 2] as "*a state of wellbeing in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community.*" This definition implies that mental health is a function of (a) overall well-being, (b) effective psychological functioning, and (c) successful integration in and contributions to society (54). These elements form the foundation for a popular approach to mental health in positive psychology called "The Mental Health Continuum" (14). Developed by Keyes (55), this approach defines mental health as "a syndrome of symptoms of positive feelings and positive functioning in life" (p. 207) where an individual is "free of psychopathology and flourishing with high levels of emotional-, psychological-, and social well-being" [(14), p. 539]. From this perspective, Keyes et al. (56) argued that mental health is more than just being free from psychopathology and is an active function of *feeling good* (i.e., *emotional well-being (EWB)*: pursuing pleasure, avoiding pain and experiencing affect balance), *functioning well* (i.e., *psychological well-being (PWB)*: having the capabilities to manage life's challenges and realize one's potential effectively), and *fitting in* (i.e., *social well-being (SWB)*: the extent toward which one optimally functions in, feels accepted by and contributes to their community). Keyes (55) argued that mental health could be described on a continuum between languishing on the lower end and flourishing at the top end of the spectrum. Further, Westerhof and Keyes (54) argued that mental health and mental illness are on separate, yet related continuums where one could (in theory) be both flourishing yet suffering from mental illness. Mental health can therefore be seen as a complete state of well-being whereby individuals have balanced positive/negative experiences, are free to realize their full potential, can play to their strengths to manage daily hassles and can actively contribute to the communities they are embedded in (57, 58).

This mental health approach and definition served as the basis for the *Mental Health Continuum-Short Form* [MHC-SF; (59)], a popular 14-item self-report measure that aims to assess individuals' overall level of emotional-, psychological-, and social well-being. The instrument assesses mental health

as a continuum that ranges from flourishing and moderately mentally healthy to languishing (14). Flourishing is constituted by elevated emotional, psychological, and social functioning levels and languishing by low self-reports on these factors. Mental health is represented by a higher-order factorial model, which assumes that the three first-order factors completely mediate the association between the 14 items and the higher-order well-being factor. This implies that the higher-order factor does not explain any unique variance over and above what is already explained by EWB, PWB, and SWB. Therefore, these three first-order factors confound the variance explained by the higher-order factor, and the variance is uniquely attributable to each of the three first-order factors (20). This factorial model relies heavily on implicit restrictive proportionality constraints whereby the ratio of item variance explained by the first- and higher-order factors is the same for all items associated with the single first-order factor (60). Despite such, this higher-order factorial model is still very prevalent in the literature. The MHC-SF has been adapted, translated, and validated in over 50 countries, thus providing considerable support for its utility and validity (56). Given the instrument's popularity, it has also been subjected to a wide array of structural validity studies, where various factorial permutations have been investigated, ranging from traditional CFAs to more complex ESEMs (58).

From a *traditional CFA perspective*, the MHC-SF has been estimated as a:

1. **Strict uni-dimensional model**, where all items directly load on a single first-order factor model, and more recently:
2. **Three-factor first-order model**, where EWB, PWB, and SWB are estimated as distinct and related (correlated) first-order factors.
3. **Higher-order (second-order) model**, where mental health is estimated as a single, second-order factorial model comprised out of three first-order factors (EWB, SWB, PWB). This CFA model is mathematically equivalent to the previously mentioned model for the MHC-SF.
4. **Bifactor model**, where mental health is seen as a general factor, that is district from the three independent specific factors. A bifactor model provides an alternative to traditional higher-order factorial models as items can simultaneously reflect an overall or "general" factor of mental health (G-factor) and three specific factors (S-factors) reflecting the unique variance shared amongst the items forming each of the three subscales that the G-factor does not explain. Therefore, the G-factor reflects the variance shared by all indicators in the model, where the S-factors represent the shared variance among all the indicators of a specific subscale that's not accounted for by the G-factor. These factors are specified as orthogonal (i.e., being unrelated to each other and therefore unique). This approach aids in solving issues related to high factor correlations and acts as a means to determine the unique contribution of the G-factor and the S-factors to predictive outcomes. Jovanovic (15) found that mental health is better represented by a bifactor model, rather than any of the other theoretical CFA permutations.

More recently, the MHC-SF has also been explored through a variety of ESEM approaches. Joshanloo and Jovanovic (25) argued that the traditional CFA approaches don't adequately represent the multi-dimensionality of the MHC-SF, and that ESEM results in better model fit, provides more accurate parameter estimates and projects more realistic inter-factor correlations. Further, ESEM structures are more closely aligned to the original theoretical conceptualization of mental health as laying on a continuum, where flourishing results from an active interaction of EWB, PWB, and SWB (47). Empirically, employing a CFA approach undermines the continuum conceptualization as it forces factors to be "categorical," instead of allowing for the dynamic interaction required to theoretically constitute a "continuum." Following the a priori factorial structure of the MHC-SF, Lamborn et al. (61), Joshanloo and Jovanovic (25), Joshanloo and Lamers (62), and others indicated that the MHC-SF is better represented by one of the following ESEM models where cross-loadings were permitted but targeted to be close to zero:

1. **Three-factor first-order ESEM model,** where EWB, PWB, and SWB are estimated as distinct and related first order factors. However, within hierarchically organized constructs such as mental health, these first-order ESEM models are more likely to ignore the presence of hierarchical superior constructs as this would instead be expressed through hyper-inflated cross-loadings (42). Hierarchical ESEM models could therefore be estimated.
2. **Higher-order ESEM model (H-ESEM)**, where mental health is estimated (*via* the ESEM-within-CFA framework) as a single, second-order factorial model comprised out of three first-order factors (EWB, SWB, PWB).
3. **Bifactor ESEM model,** where mental health is seen as a general factor, that is district from the three independent specific factors.
4. **ESEM-within-CFA model,** could thus also be specified where the three-factor first-order ESEM model is re-expressed within a CFA framework using the starting values of the original three-factor first-order ESEM. This approach allows the ESEM model to be used in more complex analyses. However, it has not yet been employed with the MHC-SF but will be demonstrated later in this tutorial.

THE PRESENT STUDY

This paper aims to provide an illustrative tutorial on the specification, comparison, reporting, and interpretation of ESEM models in Mplus with the aid of De Beer and Van Zyl's (63) ESEM Code Generator. The ESEM Code Generator assists with generating syntaxes for Mplus estimation based on the basic factor structure of an instrument, limiting the potential for inadvertent errors in manual model specification.

Specifically, this aim translates into two objectives:

1. To provide general guidelines to consider when estimating, interpreting, comparing and reporting ESEM models.

2. To provide a step-by-step guide for estimating ESEM models in Mplus with the ESEM Code Generator (63) and comparing ESEM models with traditional CFA models.

For illustrative purposes, data obtained by the LISS Open Data Project will be used to explore the factor structure of the MHC-SF. Both traditional CFA and ESEM models will be estimated and compared. For the sake of familiarity, the illustration will follow the format of a traditional paper's methods and results section accompanied by ESEM Code Generator and syntax screenshots and guidelines.

MATERIALS AND METHODS

Research Design

The study draws on data from the LISS internet panel of the CentERdata programme (https://www.dataarchive.lissdata.nl). The LISS panel is a functional element of the Measurement and Experimentation in the Social Sciences (MESS) project managed by CentERdata in Tilburg, the Netherlands. The panel gathers longitudinal data from a representative sample of 5,000 random households based on the population register by Statistics Netherlands. For this study, the first measurement of the 2007 Dataset on Mental Health was used ($n = 1{,}806$). Overall, 2,293 individuals were invited to participate in the study with a 78.7% response rate.

Participants

Data were gathered from a random, representative sample of 1,806 respondents from the Netherlands. Data were screened for response quality which led to the removal of two records from the final data set (64).

The final sample consisted of 1,804 participants (cf. **Table 1**). The majority of the participants were married (53.9%) Dutch females (50.7%) who were 65 years or older (21.4%). Most had at least a higher vocational level of education or a university degree (29.7%) and lived in self-owned housing (68.2%).

Measures

The *Dutch version* of the *Mental Health Continuum-Short Form* [MHC-SF; (55)] was used to measure overall mental health and its three components. The instrument consists of 14 self-report items that are rated on a 6-point Likert scale ranging from 1 ("Never") to 6 ("Every Day"). The instrument requests participants to reflect on the past month and indicate to what extent they experienced the three components of mental health: (a) emotional well-being (e.g., "happy"), (b) psychological well-being (e.g., "That you liked most parts of your personality") and social well-being (e.g., "That you had something important to contribute to society"). The instrument showed to be highly reliable in the Dutch context, with Cronbach's alphas ranging from 0.89 to 0.93 on the different subscales (38).

The *Dutch version* of the *Brief Symptom Inventory* [BSI: (65)] was used to measure mental illness. The scale consists of 90 items which measure nine common mental illnesses (Somatization, Obsession-Compulsion. Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism) *via* a five-point Likert type scale ranging from 0 ("Not At All") to 4 ("Extremely"). Each dimension is measured by 10 items. Example questions are: "During the past 7 days, how much were you distressed by feeling easily annoyed or irritated?" and "During the past 7 days, how much were you distressed by feeling lonely?". The overall scale and its sub-dimensions showed to be reliable, with Cronbach Alphas ranging from 0.70 to 0.95 (54). Within the LISS Dataset, only total (factor) scores for each specific factor are provided due to copyright restrictions.

Data Availability and Syntaxes

The data and syntaxes used for this tutorial are available as **Supplementary Material** to this manuscript. The **Supplementary Material** contains: (a) the original dataset in SPSS version 27 format (Mplus.sav), (b) the cleaned dataset used in Mplus (mplus.txt), (c) the ESEM syntaxes generated by the De Beer and Van Zyl (63) ESEM code generator for Mplus, and (d) the syntaxes used to estimate the CFA factor models.

Guidelines for ESEM Estimation

To estimate, compare and report on ESEM, general guidelines were developed based on the best-practices for CFAs (cf. **Table A1** for a summary). These general guidelines divide the procedure into three phases: (a) The Planning Phase, (b) the Data Preparation Phase, and (c) The Data Analysis and Reporting Phase.

The Planning Phase

First, a clear explanation of the instrument, its factorial structure and possible alternative factorial models of such should be described based on theory. When validating a psychometric instrument, clear, theory-informed hypotheses about the instrument's factorial structure or "nature" should be provided. Given that a CFA structure is nested within ESEM, a description of the traditional CFA models' original or expected factorial structure is required. As such, alternative, theory-informed, factorial permutations of the instrument should also be discussed and briefly described. These models may reflect different theoretical propositions underpinning the instrument or contradictions found in previous research (66). If the constructs within a CFA model are expected to be conceptually related, then there is also an expectation that an ESEM model would fit the data. Therefore, both the CFA and ESEM models of the different theoretical models need to be tested against the data to find the best data-model fit. If a global factor can be expected, then bifactor CFA and bifactor ESEM models should be described and later tested (37). In studies where the focus is on establishing relationships between factors or growth over time, the relationships between factors should be clearly described and supported by the literature.

Second, the required sample size should be planned for (67–69). Given that a "relatively large number of parameters need to be estimated in ESEM, smaller sample sizes could lead to decreased precision in model estimation" [(42), p. 3] and present problems with convergence (30), researchers should plan for an appropriate sample size beforehand. Given that CFA models are nested within ESEM models, traditional approaches for sample size estimation for SEM models could also be

TABLE 1 | *Demographic characteristics* of participants ($N = 1,804$).

Characteristics	Category	Frequency (*f*)	Percentage (%)
Gender	Male	890	49.3
	Female	914	50.7
Age	15–24 years	194	10.8
	25–34 years	362	20
	35–44 years	277	15.4
	45–54 years	279	15.5
	55–64 years	306	17
	65 years and older	386	21.4
Marital status	Married	973	53.9
	Separated	5	0.3
	Divorced	172	9.5
	Widow or widower	103	5.7
	Never been married	551	30.5
Level of education	Primary school	80	4.4
	Vmbo (intermediate secondary education, US: junior high school)	469	26
	Havo/vwo (higher secondary education/preparatory university education, US: senior high school)	210	11.7
	Mbo (intermediate vocational education, US: junior college)	392	21.7
	Hbo (higher vocational education, US: college)	392	21.7
	Wo (university)	144	8
	Other	72	4
	Not yet completed any education	44	2.4
	Not yet started any education	1	0.1
Type of living arrangement	Self-owned dwelling	1,229	68.2
	Rental dwelling	562	31.1
	Cost-free dwelling	13	0.7

appropriate to help control for possible convergence problems later. Many different suggestions and rules of thumb for sample size planning have been proposed for SEM and CFA in the literature which researchers could consider [cf. (70–73)]. Wolff et al. [(73), p. 3], did however, suggested three more advanced approaches to estimate the sample size requirements for SEM, whereby the required sample size is estimation based on: (a) the non-centrality parameter (i.e., based on the amount of model misspecification) (74), (b) the model's potential to obtain an acceptable RMSEA value (71), or (c) Monte Carlo simulations (75). The latter, being the most preferred approach [cf. (73) for an easy tutorial with Mplus code for running Monte Carlo simulations to estimate sample size]. However, the actual necessary sample size depends to a large extent on the researcher's goals, and it is up to the researcher to decide which approach to employ.

Data Preparation Phase

Third, the *dataset needs to be screened, cleaned and prepared for analysis*. The dataset needs to be screened for outliers, missing values and an appropriate missing values strategy (e.g., multiple imputations, Full Information Maximum Likelihood estimation, sensitivity analysis, expectation-maximization, etc.) employed before or during analyses. The choice of strategy should be reported and justified. To determine potential multivariate outliers, the Mahalanobis distance estimation method could be used ($p < 0.01$) (76). Outliers and extreme values might need to be removed from the dataset as these could affect model fit and measurement quality (76, 77). Further, data quality checks should be implemented [cf. (64) for a review on possible strategies].

Fourth, the most appropriate software, estimation method, rotation and procedure for the analysis should be decided and reported. For this illustration Mplus 8.6 (50) will be used. Once the software package has been selected, an appropriate estimation method should be decided. By default, Mplus employs the Maximum Likelihood (ML) estimation method. Morin (37) suggests the use of robust ML (MLR) from the start as it compensates for issues pertaining to multivariate normality, however, additional steps should then be implemented for statistical model comparison as chi-squares cannot directly be compared for these estimation methods (e.g., the Satorra-Bentler scaled chi-square difference test which is also implemented in Mplus with the DIFFTEST command for models with different estimators). For bifactor ESEM models with continuous indicators, the MLR estimator is appropriate; for models comprised of ordinal indicators, WLSMV should be used. Once the estimator has been chosen, the most appropriate rotation method for ESEM should be decided and reported. Three rotations are to be considered depending on the purpose of the study: (a) Geomin rotations (with an epsilon value of 0.50) for

more exploratory approaches and to maximally reduce factor correlations, (b) Target rotations for confirmatory approaches, or (c) (Target) Orthogonal rotations are used for bifactor ESEM modeling (37). Finally, the entire competing measurement modeling strategy to be employed should be described.

Data Analysis and Reporting Phase

Fifth, the most appropriate goodness-of-fit indices and indicators of measurement quality should be determined and reported. To determine the best fitting model for the data, both guidelines for goodness-of-fit indices as well as indicators of measurement quality need to be employed. Measurement models need to show both good data-model-fit and have high levels of measurement quality to be retained (33). To determine goodness-of-fit, Hu and Bentler's (55) proposed a number of general fit indices with suggested cut-off scores which are summarized in **Table 2**. It is, however, essential to note that each suggested model fit indicator is subjected to its own limitations and the use thereof needs to be justified. As such, multiple indicators of fit should be used to decide upon the best fitting model for the data (33). The CFI, TLI, and RMSEA should always be reported and used as the primary criterion for both establishing model fit and to discriminate between models. However, recent research has shown that SRMR outperforms RMSEA when the data that is modeled is categorical in nature (78).

After model-fit is established, measurement quality needs to be assessed. Researchers should decide, a priori, on indicators of measurement quality that can range from inspecting the standardized factor loadings (e.g., $\lambda > 0.35$), the item uniqueness (e.g., residual error variances >0.10 but <0.90), levels of tolerance for cross-loadings, and overall R^2 for each item. However, the results should be considered in the context of the study and what they might mean or indicate without being unnecessarily rigid about minor deviations from the aforementioned rules of thumb.

Sixth, *estimate- and report the model fit indicators for various competing CFA models*. In this step, different theoretical factorial models should be estimated and their model fit statistics reported. For the MHC-SF, theory indicates that four types of CFA models are possible: A single factor model, three-factor first-order model, a second-order model and a bifactor model. Here measured items are used as observed indicators for latent factors. For the CFA models, items should only be allowed to load on their a priori theoretical factors and cross-loadings of items should not be permitted. For bifactor models (B-CFA) an orthogonal target rotation should be employed and specified in Mplus under the Analysis Command: ROTATION = TARGET (ORTHOGONAL). Here, a general factor for overall Mental Health (G-factor) should be specified, accompanied by emotional well-being, psychological well-being, and social well-being as specific factors (S-factors). For the G-factor, all observed indicators (measured items) need to be specified to load onto this single factor. For the three S-factors, items related need to be specified to load onto their a priori factorial structures. The orthogonal targeted rotation forces all factors to be uncorrelated. In other words, all covariances between the specific factors and general factor are constrained to be zero. This can also be manually specified in Mplus (e.g., EWB WITH PWB@0;). Further, any potential modifications made to the CFA models to enhance model fit should be reported and justified in text. Only modifications with a strong theoretically informed reason should be permitted. All model fit statistics (mentioned in **Table 2**) for the various models should be tabulated and reported.

Seventh, estimate and report the model fit indicators for various competing ESEM models. Like the previous step, various theoretically informed ESEM models need to be estimated and their model fit statistics reported. For the MHC-SF, theory indicates that three types of ESEM models are possible: a three first-order-, a second-order-, and a bifactor ESEM model. Additionally, an ESEM-within-CFA model could also be estimated based on the first-order ESEM model if more complex analyses are later required. This should be done in Mplus, and a target rotation should be employed. Unlike within the CFA models, cross-loadings between items and non-target factors are permitted; however, these should be constrained to be as close to zero as possible (79) (in Mplus, this is indicated by ∼0; after the specific cross-loadings). Items that load onto their a priori theoretical latent factor should not be constrained. For the bifactor ESEM (B-ESEM) model, a similar approach as mentioned for the B-CFA model should be employed where the MHC-SF is comprised of a single G-Factor and three S-Factors. However, unlike in the B-CFA model, cross-loadings on non-target S-factors are permitted but targeted to be as close to zero as possible. The code for all the ESEM models can be generated with the De Beer and Van Zyl (63) ESEM code Generator for Mplus. This will be explained in the next section. All observed model fit statistics (mentioned in **Table 2**) for the various models should be tabulated and reported. It is suggested that both the CFA and ESEM results be reported in the same table, making model comparisons easier to read.

Eight, to determine the best-fitting model for the data, the competing CFA and ESEM models need to be compared. In this step, the results for both Steps 6 and 7 are compared, based on the criteria specified in **Table 2**, to determine the best-fitting model for the data. Given that the CFA models are embedded within the ESEM models, direct comparisons on model fit can be made. Only models that meet both the measurement quality and goodness-of-fit criteria should be retained for further analyses. Models with the lowest AIC, BIC and aBIC values show better fit and should be favored. For competing nested models, a model shows better fit if both the chi-square difference test between models is significant ($p < 0.05$) and changes in RMSEA/SRMR and TLI/CFI exceed 0.015 and 0.01, respectively (80, 81). It should, however, be noted, that there is considerable debate in the literature with regards to these delta fit indices comparisons [cf. (82)], and that specific focus should also be placed on inspecting and giving substantial consideration to the parameter estimates of the various models and not just goodness-of-fit criteria when selecting the "final" model. The criteria chosen should be specified and justified by the researcher.

Morin et al. (20) further indicated that to retain an ESEM model for further analysis, several conditions need to be additionally met:

TABLE 2 | Model fit statistics.

Fit indices	Cut-off criterion	Sensitive to *N*	Penalty for model complexity
Absolute fit indices			
Chi-Square (χ^2)	• Lowest comparative value between measurement models • Non-Significant Chi-Square ($p > 0.01$) • Significant difference in Chi-Square between Models • For Model Comparison: Retain Model with Lowest Chi-Square	Yes	No
Approximate fit indices			
Root-Means-Square Error of Approximation (RMSEA)	• 0.06–0.08 (Marginally Acceptable); 0.01–0.05 (Excellent) • Not-significant ($p > 0.01$) • 90% Confidence Interval Range should not include Zero • 90% Confidence Interval Range should not overlap between models • For model comparison: Retain Model where ΔRMSEA ≤ 0.015	No	Yes
Standardized Root Mean Square Residual (SRMR)	• 0.06 to 0.08 (Marginally Acceptable); 0.01–0.05 (Excellent) • For model comparison: Retain Model where ΔSRMR ≤ 0.015	Yes	No
Incremental fit indices			
Comparative Fit Index (CFI)	• 0.90 to 0.95 (Marginally Acceptable Fit); 0.96 to 0.99 (Excellent) • For model comparison: Retain Model with Highest CFI value (ΔCFI > 0.01)	No	No
Tucker-Lewis Index (TLI)	• 0.90 to 0.95 (Marginally Acceptable Fit); 0.96 to 0.99 (Excellent) • For model comparison: Retain Model with Highest TLI value (ΔTLI > 0.01)	No	Yes
Akaike Information Criterion (AIC)	• Lowest value in comparative measurement models	Yes	Yes
Consistent AIC (CAIC; calculated as BIC + free parameters	• Lowest value in comparative measurement models	Yes	Yes
Bayes Information Criterion (BIC)	• Lowest value in comparative measurement models	Yes	Yes
Sample-Size Adjusted BIC (aBIC)	• Lowest value in comparative measurement models	Yes	Yes

a. the ESEM model should ideally show better data-model fit than any other CFA model (however, if the factor correlations for the ESEM model are smaller than those of the CFA model, it should still be considered despite showing similar or worse fit). If not, the more parsimonious CFA model should be retained.
b. For normal (not bifactor) models, the ESEM model should show lower factor correlations than those produced by the CFA models.
c. The ESEM model should only show small to medium cross-loadings (<0.50). Should large cross-loadings exist, then there should be a theoretical explanation presented for such. These could potentially be explained by "wording" effects or some practical logic.
d. The estimated latent factors within the ESEM model should be well-defined (i.e., strong loadings, and loadings matching expectations).
e. Should there be multiple medium to large cross-loadings in the ESEM model, it could indicate support for the presence of a larger global factor, and therefore the bifactor ESEM model should be explored.
f. Additional factors to consider for bifactor ESEM models: This model could show a significantly better fit than any of the ESEM or CFA models because of the relatively large number of freed parameters. Therefore, there should be a well-defined G-Factor (where all items load significantly on such), and *reasonably* well-defined S-Factors (cross- and non-significant loadings are permitted). For bifactor models, model fit should not be the only indicator informing a decision to retain. Researchers should also closely inspect the parameter estimates before making final decisions.

Ninth, for the best fitting model(s) the factor correlations should be computed and compared. Morin [(20), p. 1060] argued that "in addition, the model comparison strategies typically advocated for contrasting alternative ESEM and CFA solutions highlight the critical role of the factor correlations, which directly indicate whether the cross-loadings have an impact on improving the factor definition." Therefore, when choosing which model to retain, the factorial inter-correlations between latent factors for all the best fitting models (excluding the bifactor models) should be estimated and considered. This shows the level of unique distinction between factors. The model with the smallest factor correlation is usually retained, however, decisions should be based in the context of the other considerations (model fit, measurement quality, and parameter estimates) mentioned earlier. This step, however, cannot be done for bifactor models as the relationships between the specific and general factors are constrained to zero.

Tenth, report and compare the item level parameters and levels of reliability for the best fitting measurement model(s).

This step aims to investigate the item level parameters and indicators of reliability for the best fitting models to further discriminate between the different models. This step is of particular importance when validating a psychometric instrument. However, item-level parameters should always be inspected, but may not be appropriate to report in studies unrelated to an instrument's validation. When validating an instrument, the standardized factor loadings, standard errors and item uniqueness should always be reported for the best fitting model when the paper's purpose is to validate an instrument. For the CFA models, the corrected item-total correlations (CITC) values represent each item's unique relationship with the overall factor on which it has been specified to load (83). Zijlmans et al. (83) argued that a CITC value bigger than 0.30 indicates that an item accurately represents the overall factor on which it specified. Note, that if a bifactor CFA model is retained, reviewers and editors may request additional information such as the Explained Common Variance (ECV), the H-factor, the Factor Determinacy indicator, the Item level ECV, the Percent of Uncontaminated Correlations (PUC), and the Average Relative Bias Parameters could also be reported as additional indicators of reliability and measurement quality [For a tutorial cf. (84)]. Further, the indicators for reliability should be decided and reported. To determine the level of reliability for the different factorial model, the following could be reported: point-estimate composite reliability [upper-bound; $\rho > 0.80$; (85)] or McDonald's Omega [$\omega > 0.70$; (86)]. For the bifactor CFA models, the explained common variance (ECV) should be reported. A scale is regarded as essentially unidimensional when the general factor explains at least 70% of the total common variance. There are, however, no recommendations yet regarding how cross-loadings should be incorporated for bifactor ESEM models in Omega estimation [(20, 87)]. Morin et al. (20) suggest that these cross-loadings in bifactor ESEM models should, for the time being, be ignored when calculating Omega. It should also be noted, that for ESEM models, Omega cannot entirely reflect the reliability of a construct and should not be used as the only indicator. Rather, Omega should be used as an additional indicator to control for the fallible nature of psychological measurement and supplemented by other metrics of measurement quality.

Further or Additional Analysis

Should there be a need to conduct additional or more complex statistical estimations (e.g., latent growth modeling, invariance testing, multi-group analysis, structural path models, etc.), the ESEM-within-CFA approach should be employed. Here, the best fitting first-order ESEM model is respecified within a CFA structure, where all parameters of the ESEM model parameters are used as starting values for the ESEM-within-CFA model (39, 42). This would afford the opportunity to use the ESEM model as an input in a structural model.

TUTORIAL AND RESULTS

Conceptualization

To determine the factorial validity of the MHC-SF, a competing measurement modeling strategy was employed comparing traditional CFA- with ESEM models. Based on the literature, the following models could be estimated:

0. Model 0: Unidimensional CFA Model of Overall Mental Health (58) (**Figure A1a**)
1. Model 1: Correlated Three First-Order CFA Model comprised of EWB, SWB, and PWB (14) (**Figure A1b**)
2. Model 2: Hierarchical CFA Model compromises a single Second-Order Factor of Mental Health, consisting of three first-order factors (47) (**Figure A1c**)
3. Model 3: Bifactor CFA Model of Overall Mental Health (15, 61) (**Figure A1d**)
4. Model 4: Correlated Three-factor First-Order ESEM Model comprised of EWB, SWB, and PWB (62) (**Figure A1e**)
5. Model 5: Hierarchical ESEM Model compromise of a single Second-Order Factor of Mental Health, made up of three first-order factors (47) (**Figure A1f**)
6. Model 6: Bifactor ESEM Model for Overall Mental Health (46, 61) (**Figure A1g**)
7. Model 7: Correlated Three-Factor First-Order ESEM within CFA Model[1]. Here, mental health is seen as the function of three independent first-order factors (as specified before). However, the starting values from Model 4 are used to constrain the items loadings for each independent factor.

Sample Size Estimation

To estimate the minimum required sample size for the current study, the power and sample size approach of MacCallum et al. (71) was used for testing null hypotheses of not-good fit according to the RMSEA. Considering the previously proposed CFA models of the MHC-SF, ranging from a unidimensional to a bifactor CFA model (with the number of degrees of freedom ranging between 77 and 64), the minimum sample size would range between 249 and 278 to have 90% power to reject the hypothesis of not-close fit (RMSEA $\geq$ 0.05) at a 5% level of significance (88). Given that ESEM models have fewer degrees of freedom, slightly larger sample sizes are required for these models.

Data Screening, Cleaning, and Preparation

The data was screened for potential issues (e.g., outliers, missing values, data quality) and prepared for further analysis. Based on the Mahalanobis' distance, two outliers were removed from the overall dataset [$p < 0.01$; (76, 77)]. No missing values were present in the final dataset. Therefore, the final sample used for the study was $N = 1{,}804$.

Determine the Most Appropriate Software, Estimator, Rotations, and Procedure

CFA and ESEM analyses were conducted using Mplus v 8.6 (50). To explore the factorial validity of the MHC-SF a competing measurement modeling strategy *via* structural

[1]Note that the ESEM-with-CFA is not an alternative factorial model. It's specified as a means to conduct more complex analysis as discussed prior. This model is only specified and compared to demonstrate the full capability of the De Beer and Van Zyl (63) ESEM code generator and to demonstrate that no differences exist between it and Model 4.

equation modeling was used. The maximum likelihood (ML) estimation method was employed, given the relatively normal distribution of the data. For the ESEM models, an oblique target rotation was used and for the bifactor ESEM and bifactor CFA, a target orthogonal rotation was employed.

In order to determine the best-fitting model for the data, we estimated and sequentially compared several CFA (unidimensional, first-order factor, second-order factor, and bifactor) and ESEM (first-order ESEM, hierarchical ESEM, bifactor ESEM, and ESEM-within-CFA) models. The CFA models were specified according to the independent cluster modeling assumptions where items are only permitted to load onto their a priori theoretical factor, and cross-loadings were constrained to zero (79). For the bifactor model (B-CFA), we used a target orthogonal rotation. A general factor (G-factor) was specified, comprising all the items of the MHC-SF Further, three specific factors (S-factors) corresponding to the a priori theoretical dimensions of the MHC-SF were specified. For the ESEM models, we used a target rotation. Here, we specified items to load onto their a priori theoretical constructs where cross-loadings were freed but targeted to be as close to zero as possible (20). With the H-ESEM model, we explored a second-order factor structure where the original ESEM model was re-specified as a CFA model, in line with the ESEM-within-CFA framework. We used the non-standardized loadings from the ESEM model as starting values for the H-ESEM estimation. Factor variances were constrained to one, and one item per latent construct was constrained to be equal to the original ESEM item loading. Then we used the first-order factors to define a higher-order factor to determine the variance and the standardized path coefficient for each individual factor loading onto the overall Mental Health Factor. For the bifactor ESEM (B-ESEM) we followed a strategy similar to its B-CFA counterpart. However, in the B-ESEM, a target rotation was used where cross-loadings were freed and targeted to be as close to zero as possible. For all models, observed items were used as indicators for latent variables. For the ESEM-within-CFA model, mental health is seen as the function of three independent first-order factors, where the starting values from the initial ESEM model are used to constrain the items loadings for each independent factor. Here the original ESEM model is re-expressed within a CFA framework; as such, no rotation is necessary. Using De Beer and Van Zyl's (63) ESEM generator, we generated the Mplus syntaxes for the ESEM, B-ESEM and H-ESEM models.

Determine Appropriate Goodness-of-Fit Indices and Indicators of Measurement Quality

To determine the best fitting measurement model, both goodness-of-fit indices and measurement quality indicators are used to discriminate between models. First, the Hu and Bentler (32) model fit criteria were used to establish data-model fit (cf. **Table 2**). Second, various indicators of measurement quality were used to further inspect and discriminate between models. Here, the standardized factor loadings ($\lambda > 0.35$), the item uniqueness (>0.10 but <0.90), levels of tolerance for cross-loadings, and overall R^2 for each item was inspected. Only models that met both the measurement quality and goodness-of-fit criteria, were retained for further analyses.

Estimate and Report the Model Fit Indicators for Competing CFA Models

Four CFA measurement models were estimated based on the different a priori factorial permutations of the MHC-SF found in the literature (see Point 1 of the Tutorial). In the model estimation, measured items were treated as continuous variables and used as indicators for the latent factors. No items were omitted, error terms were left uncorrelated and item parceling was not allowed. The following models were estimated in Mplus:

Model 0: Unidimensional CFA Model of Overall Mental Health. A unidimensional model for overall Mental Health (labeled "MENTAL") was estimated, where all 14 items (MHC_1 to MHC_14) were specified to load directly on to such. This model acts as the baseline model for analyses.

Model 1: **Correlated Three First-Order CFA Model comprised of EWB, SWB and PWB**. A model was estimated where the MHC-SF is comprised of three first-order factors measured by 14 items: Emotional Well-being (labeled EMOTION comprised of items MHC_1 to MHC_3), Social Well-being (labeled SOCWELL comprised of items MHC_4 to MHC_8) and Psychological Well-being (labeled SOCWELL comprised of items MHC_9 to MHC_14). These factors were left to freely correlate.

Model 2: **A Hierarchical CFA Model compromised of a single Second-Order Factor of Mental Health, made up of three first-order factors**. Here, Mental Health was seen as a second-order factor that is a function of Emotional Well-being (Item: MHC_1 to MHC_3), Social Well-being (Items: MHC_4 to 8) and Psychological Well-being (Items: MHC_9 to MHC_14).

Model 3: Bifactor CFA Model of Overall Mental Health. MHC-SF was estimated to be comprised of a General Factor representing overall mental health (where all 14 items are specified to load onto such directly) which is distinct and independent from its three first-order factors Emotional Well-being (Item: MHC_1 to MHC_3), Social Well-being (Items: MHC_4 to 8), and Psychological Well-being (Items: MHC_9 to MHC_14). Here an Orthogonal Target rotation [ROTATION = TARGET (ORTHOGONAL)] was used and the relationships between specific and general factors were constrained to zero (this is automatically done in Mplus, but can also be manually specified by constraining the relationships between factors in the WITH statement to @0). The first factor loadings for each factor, which are automatically constrained to 1 by Mplus, were permitted to be freely estimated (indicated by the *). The variances for each specific and general factor were constrained to 1 (indicated by @1). The initial results showed that the model couldn't converge, after which the iterations and starting values were increased. However, the results showed that item MHC_10 then produced a negative residual error variance. Kline (77), as well as Wong and Wong (81), indicated that in such cases, the residual error variance of the observed indicator

should be constrained to be positive and slightly bigger than zero (MHC_10@0.03). This allowed the model to converge.

Table 3 provides a summary of the model fit indices for each of the estimated models. The results showed that none of the CFA models completely fitted the data based on the model fit criteria specified in **Table 2**. However, Model 3, the B-CFA model, partially met the goodness-of-fit criteria $\{\chi^2_{(1,802)} = 868.74$; $df = 64$; CFI = 0.94; TLI = 0.89; RMSEA = 0.08 [0.079, 0.088]; SRMR = 0.05$\}$. The parameter estimates showed that this model produced a well-defined general factor (with all items $\lambda > 0.35$; small standard errors <0.04). Further, the specific factors were also relatively well-defined, with item loadings matching expectations for both the Emotional Well-being and Social Well-being subscales. However, items MHC_11, MHC_13, and MHC_14 on the Psychological Well-being Subscale showed non-significant specific factor loadings ($p > 0.01$). Further, items MHC_9, and MHC_12 only produced small, yet significant, factor loadings ($\lambda < 0.39$) on the psychological well-being factor. Under normal circumstances, this model would therefore also be rejected for further consideration. However, for the purposes of this tutorial, Model 3 will be retained for further analyses to show how ESEM and CFA models can be compared.

Estimate and Report the Model Fit Indicators for Competing ESEM Models

Next, a series of ESEM models were estimated based on the a priori CFA factorial structures of the MHC-SF found in the literature (see Point 1 of the Tutorial). Similar to the CFA models, measured items were tread as continuous variables and used as indicators for the latent factors. No items were omitted, and item parceling was not allowed. However, unlike the CFA models, cross-loadings were permitted but targeted to be close to zero. The De Beer and Van Zyl [(63); http://www.surveyhost.co.za/esem/] ESEM syntax generator was used to create the syntaxes needed to run the ESEM, B-ESEM, H-ESEM and ESEM-within-CFA models. For the purposes of this tutorial, we will walk readers through each step of the estimation process, from generating the codes *via* the tool to how it should be presented and interpreted. First, a general overview and step by step guide on using the tool will be presented. Second, the tool will be applied to the MHC-SF dataset to generate the results.

Overview and Purpose of the De Beer and Van Zyl (63) ESEM Code Generator

The purpose of the De Beer and Van Zyl (63) ESEM tool is to aid researchers to generate the Mplus syntaxes needed to run several complex ESEM models. Estimating ESEM models within Mplus is rather complex and could become rather tedious. The tool was developed to intuitively guide researchers to generate the Mplus syntaxes needed to estimate normal ESEM-, bifactor ESEM-, Hierarchical ESEM (H-ESEM), and ESEM-within-CFA models. This tool is based on the ESEM estimation procedure discussed in Asparouhov and Muthén (30) and demonstrated by Howard et al. (45) for bifactor ESEM and Morin and Asparouhov (44) for H-ESEM and ESEM-within-CFA models.

Estimating these ESEM models can be done in four relatively easy steps:

TABLE 3 | Competing CFA and ESEM measurement models.

Model	Type	χ^2	*df*	CFI	TLI	RMSEA	SRMR	AIC	BIC	aBIC	Meets Criteria
Confirmatory factor analytical models											
Model 0	Unidimensional first-order factor model	2345.02	77	0.78	0.74	0.13 [0.128–0.132]	0.07	77662.47	77893.38	77759.95	No
Model 1	Three first-order factor model	1283.34	74	0.88	0.86	0.10 [0.091–0.100]	0.06	76606.79	76854.19	76711.23	No
Model 2	Second-order factor model	1283.34	74	0.88	0.86	0.10 [0.091–0.100]	0.06	76606.79	76854.19	76711.23	No
Model 3	Bifactor model	868.74	64	0.92	0.89	0.08 [0.079–0.088]	0.05	76212.19	76514.57	76339.84	No
Exploratory structural equation models											
Model 4	Three first-order ESEM	634.78	52	0.94	0.90	0.08 [0.073–0.084]	0.03	76002.23	76370.59	76157.73	Yes
Model 5	Higher-order ESEM	983.42	55	0.91	0.85	0.10 [0.091–0.102]	0.08	76344.87	76696.73	76493.40	No
Model 6	Bifactor ESEM	272.29	41	0.98	0.95	0.06 [0.050–0.062]	0.02	75661.74	76090.52	75842.76	Yes
Model 7	ESEM within CFA	634.78	52	0.94	0.90	0.08 [0.073–0.084]	0.03	76002.23	76370.59	76157.73	Yes

χ^2, Chi-square; df, degrees of freedom; TLI, Tucker-Lewis Index; CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation [90%CI]; SRMR, Standardized Root Mean Square Residual; AIC, Akaike Information Criterion; BIC, Bayes Information Criterion; aBIC, Adjusted Bayes Information Criterion.

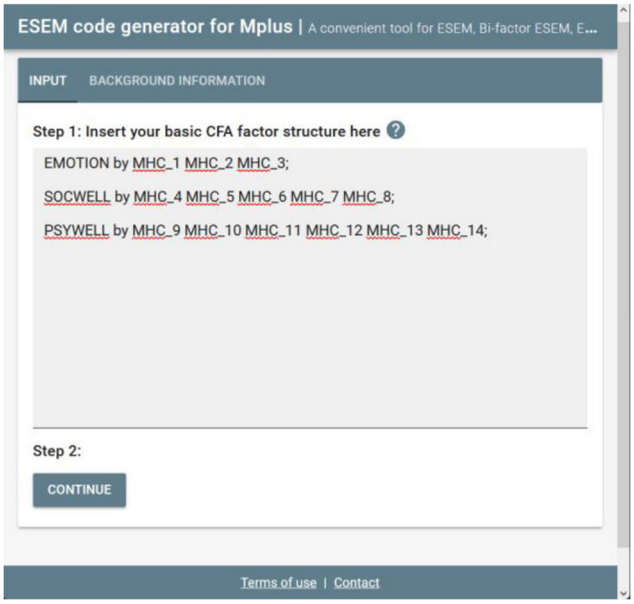

FIGURE 1 | Step 1: specify a CFA model.

1. **STEP 0: Navigate any web browser to** http://www.surveyhost.co.za/esem/
2. **STEP 1: Specify a CFA Model**. Provide the Mplus syntax code for a traditional First-Order CFA Factorial model and Click Continue to generate the syntaxes.
3. **STEP 2: Generate, Copy, and Run the ESEM Syntaxes in Mplus**. Copy the syntax generated for the Regular ESEM or bifactor ESEM solution into Mplus and run these models.
4. **STEP 3 (Optional): Generate, Copy, and Run the Syntax for H-ESEM/ESEM-within-CFA Models**. Upload the Mplus Output produced in STEP 2 to generate the syntaxes for the H-ESEM- and ESEM-within-CFA models and click continue.

Step 1: Specify a CFA Model

Once users have directed their browser to the online tool, they will be requested to specify a basic CFA factorial model. This tool only accepts the syntax commands for Mplus v.6 and above. Specify a basic first-order CFA factorial (measurement) model in the Mplus language on the INPUT command (cf. **Figure 1**). This can be done by using the BY command in Mplus (e.g., EMOTION by Item1 Item2 Item3;). Ensure that all first-order factors are correctly specified, and the command closes with a ';'. Once done, the researcher should click continue to generate the Mplus syntaxes.

Step 2: Generate-, Copy-, and Run the ESEM Syntaxes in Mplus

Upon clicking continue, the user will be redirected to a new page whereby the Original Input is provided as well as the regular- and bifactor ESEM syntaxes generated (see **Figure 2**). The syntax provides the correctly specified ESEM models and brief descriptions of each command for ease of reference (see **Figure 3** as an example). The appropriate syntax required by the researcher should then be copied and pasted in Mplus. Researchers should, however, still specify (1) the name- or location of the dataset next to the FILE IS command, (2) populate the variables names from the dataset under the NAMES ARE command, (3) specify how missing values are labeled under the MISSING ARE ALL (-XX) command, (4) make choices in the ANALYSIS command relating to the estimation method, and rotation type, and (5) any additional outputs required [cf. (50) for an outline of different output commands]. Once these factors have been clarified, the researcher can run both models in Mplus. The output of the ESEM model should be saved, as this will be used as input for the next step, should users want to generate H-ESEM or ESEM-within-CFA models.

Step 3 (Optional): Generate, Copy, and Run the Syntax for H-ESEM or ESEM-Within-CFA Models

Researchers may also be interested in estimating more complex ESEM models such as H-ESEM or ESEM-within-CFA models. These models require that the starting values for both be changed from the defaults. The non-standardized factor loadings from the original ESEM model should be used as starting values for the H-ESEM and ESEM-within-CFA model estimation syntaxes. Factor variances are also constrained to one, and one item per latent construct is constrained to be equal to the original ESEM item loading. Then we used the first-order factors to define a higher-order factor to determine the variance and the standardized path coefficient for each individual factor loading onto a higher-order factor. For the ESEM-within-CFA model, the regular ESEM model is re-expressed as a CFA model. This model employs the unstandardized factor- and cross-loadings estimated from the regular ESEM model as starting values (denoted by the * command). First-order factor variances are again freely estimated, whereas the higher-order factor is constrained to 1 in order to identify the model. Furter, in this model, one item per first-order factor has all its factor loadings constrained to equal that of the original ESEM values denoted by the @ command. Researchers can do this manually, however, this leaves room for error. The tool aids researchers to generate these syntaxes, by requesting that the original Mplus output from the ESEM model generated in Step 3 be uploaded. Researchers should click on the UPLOAD button and direct their explorer to the output file from the original ESEM model (see **Figure 4**)[2].

Once selected, users can select from two cross-loading options: (a) the *Default Loadings* which uses the largest factor loading for each factor and fixes it as the cross-loading for each factor or (b) *Optimized Loadings* whereby the script attempts

[2]It is, however, important to note that in the ESEM within CFA code generator that all referent indicators are fixed (@) in all the factors to the starting values to aid in convergence. Therefore, factor variances are freed (as indicated by the *; e.g., EMOTION*). If these factor variances are manually constrained to 1 (e.g., EMOTION@1;) then the starting value of the referent indicator in its original factor should then be freed (*). Either specification should result in the same degrees of freedom, model fit and parameter estimates.

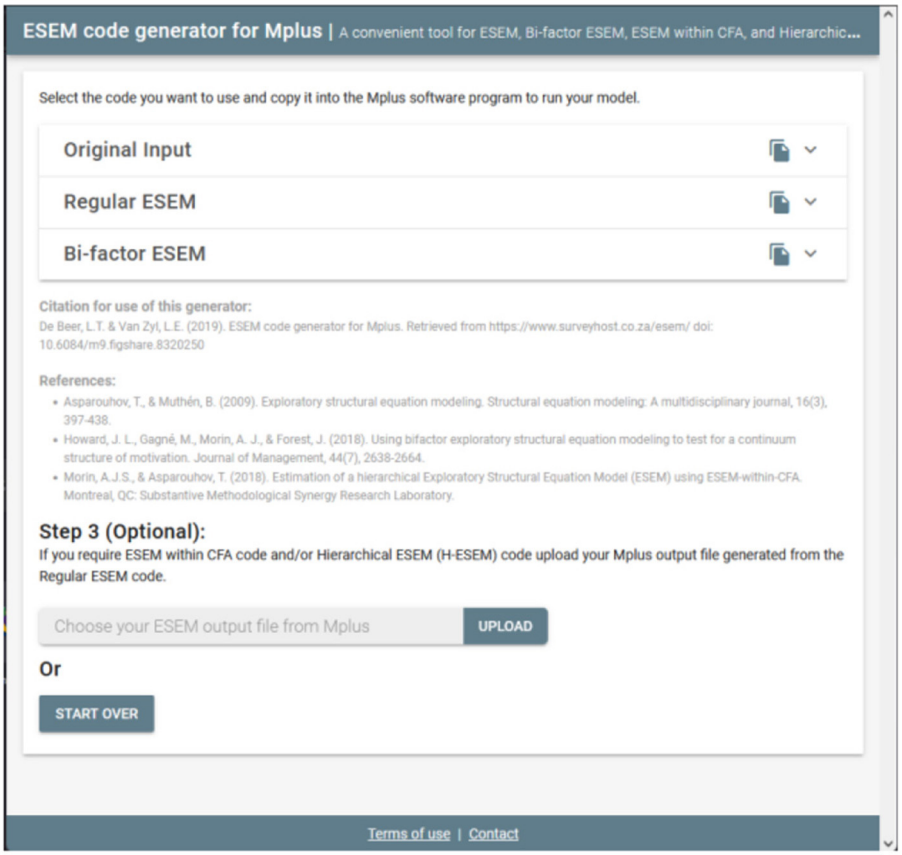

FIGURE 2 | Step 2.1: Generated ESEM and bifactor ESEM Syntaxes.

to find the most optimal loadings from the specific original factor where it also has the smallest cross-loadings to fix in the model. The latter is still an experimental feature and novices are encouraged to use the *Default Loadings* option. Once the cross-loadings option is selected, researchers can click on CONTINUE to generate the new syntaxes (kindly note that no information about the data is captured or stored on the server). This will then redirect the user to a new page where two additional ESEM syntax options have been generated (see **Figure 5**). These syntaxes can then be copied and pasted into Mplus to be run. The tool will be practically applied, and example codes are shown in the next section.

Estimate and Report the Results for Competing ESEM Models

After the four CFA models were estimated, four ESEM models were tested against the data. These ESEM models follow the same theoretical structure of the CFA models, however, cross-loadings were permitted but constrained to be as close to zero as possible. The syntaxes for these ESEM models were generated with the ESEM Code Generator (63). The following ESEM models were estimated in Mplus:

Model 4: Correlated Three-Factor First-Order ESEM Model Comprised of Emotional-, Social-, and Psychological Well-Being

This model assumes that Emotional- (Targeted Items: MHC_1 to MHC_3), Social- (Targeted Items: MHC_4 to 8), and Psychological well-being (Targeted Items: MHC_9 to MHC_14) are separate yet related components of mental health. In this model, items are targeted to load onto their a priori factorial model, but cross-loadings were permitted but targeted to be close to zero. The code generated to run the model in Mplus is presented in **Figure 6**.

Model 5: A Hierarchical ESEM Model Compromised of a Single Second-Order Factor of Mental Health, Made Up of Three First-Order Factors

Mental health was specified as a second-order ESEM model that is a function of EWB (Item: MHC_1 to MHC_3), SWB (Items: MHC_4 to 8), and PWB (Items: MHC_9 to MHC_14). Again, items were specified to load directly onto their a priori first-order factors. Cross loadings were again permitted but constrained to be as close to zero as possible. The ESEM-within-CFA estimation procedure was used to construct the higher-order factorial model. Here the starting values for each item was constrained to be the same as the unstandardized

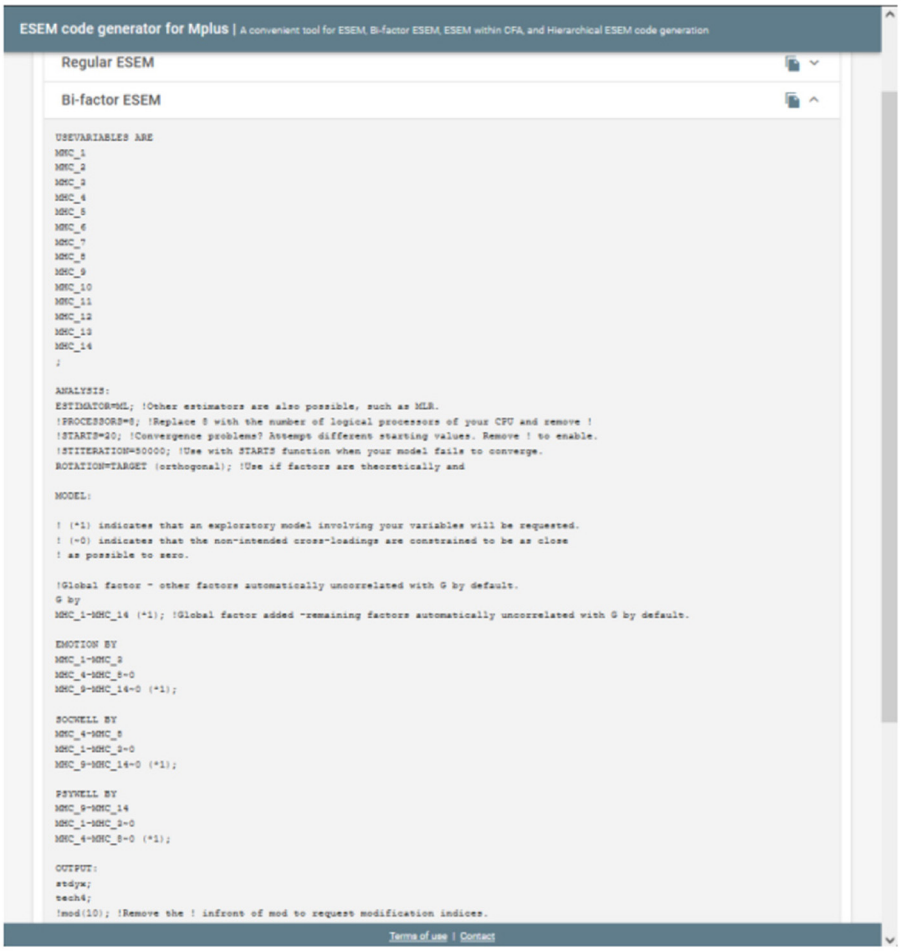

FIGURE 3 | Step 2.2: Example output for the bifactor ESEM Syntax.

factor- and cross-loadings estimated from the Regular ESEM Model (Model 4). The first-order factor variances were freely estimated and that of the higher-order Mental Health Factor was constrained to 1. Further, for the first-order factor, one item per factor is constrained to produce exactly the same loadings and cross-loadings as the ESEM Model 4. Finally, the Higher-Order Factor Mental health (or HFACTOR in the syntax in **Figure 7**) was specified as being comprised of freely estimated first-order factors.

Model 6: Bifactor ESEM Model for Overall Mental Health

Similar to Model 3, the MHC-SF was specified to be comprised of a General Factor representing overall mental health (where all 14 items are specified to directly load onto such) which is distinct and independent from its three first-order factors emotional- (Target Item: MHC_1 to MHC_3), social- (Target Items: MHC_4 to 8), and psychological well-being (Target Items: MHC_9 to MHC_14). Target items were specified to load directly on their a priori factorial models but cross-loadings on the specific factors were permitted but constrained to be as close to zero as possible. **Figure 8** provides a screenshot of the Mplus Syntax.

Model 7. A Correlated Three-Factor ESEM-Within-CFA Model

Here, mental health is seen as the function of three independent first-order factors (as specified before). Within this model, the Regular ESEM Model (Model 4) is re-expressed as a CFA model, where the starting values from Model 5 are used to constrain the items loadings for each independent factor. The variances of the three first-order factors (emotional-, social-, and psychological well-being) are freely estimated. This model is not a separate or different type of ESEM model that should be contrasted/compared. This model is only specified to be used for more complex follow-up analysis. It's only estimated and compared here to demonstrate the tool. The Syntax is presented in **Figure 9**.

The model fit indices of the ESEM models are also captured and summarized in **Table 4**. Unlike the CFA models, the results showed that *Model 4* [$\chi^2_{(1,802)} = 634.78$; $df = 52$; CFI = 0.94; TLI = 0.90; RMSEA = 0.08 [0.073, 0.084]; SRMR = 0.03;

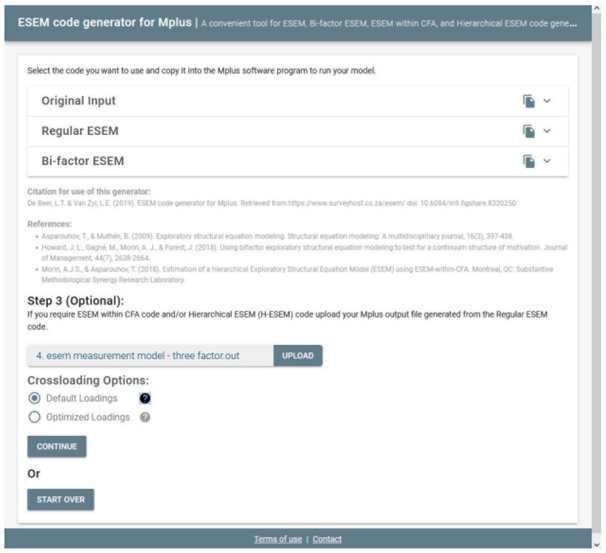

FIGURE 4 | Uploading ESEM output and specifying cross-loadings option.

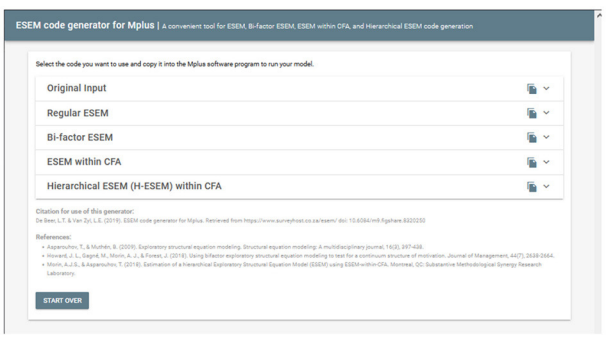

FIGURE 5 | Step 3: H-ESEM or ESEM-within-CFA models codes generated.

AIC = 76002.24; BIC = 76370.59], *Model 6* [$\chi^2_{(1,802)} = 272.285$; $df = 41$; CFI = 0.98; TLI = 0.95; RMSEA = 0.06 [0.050, 0.062]; SRMR = 0.02; AIC = 75661.74; BIC = 76090.56], and *Model 7* [$\chi^2_{(1,802)} = 634.78$; $df = 52$; CFI = 0.94; TLI = 0.90; RMSEA = 0.08 [0.073, 0.084]; SRMR = 0.03; AIC = 76002.24; BIC = 76370.59] all fitted the data. Further Model 4 (the original ESEM model) and Model 7 (the ESEM-within-CFA models) produced, as intended, the same results. Further, by inspecting the parameter estimates, the results showed that all models, except Model 5, showed the expected results with target items loading significantly on their respective factors ($\lambda > 0.35$; small standard errors <0.04). Given that Model 5 did not meet either the model fit or measurement quality criteria, it was disregarded for further analysis.

Compare CFA and ESEM Models to Determine the Best-Fitting Model for the Data

The next step is to contrast and compare the retained CFA and ESEM models. From the previous sections the bifactor

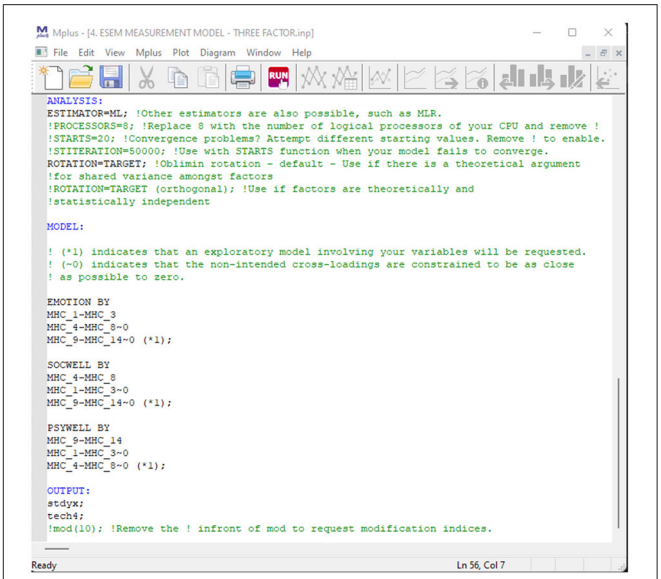

FIGURE 6 | Mplus syntax for a three first-order ESEM model of mental health.

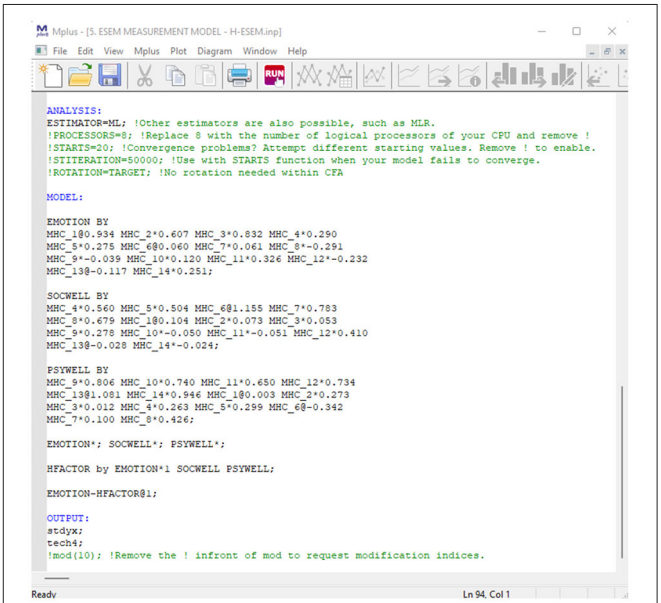

FIGURE 7 | Mplus syntax for a H-ESEM model of mental health.

CFA Model (Model 3) as well as the Three-First Order Factor ESEM (Model 4), bifactor ESEM (Model 6)„ and ESEM-within-CFA (Model 7) were retained. These competing Models are further compared based on their Model Fit Indices, and the results summarized in **Table 4**. The results showed that *Model 6* fitted the data significantly better than *Model 3* ($\Delta\chi^2 = -596.46$; $\Delta df = 27$; ΔCFI = 0.06; ΔTLI = 0.06; ΔRMSEA = −0.03; ΔSRMR = −0.03; ΔAIC: −550.45; ΔBIC: −424.00), and Model 4 and Model 7 ($\Delta\chi^2 = -362.498$; $\Delta df = 39$; ΔCFI = 0.03; ΔTLI = 0.05;

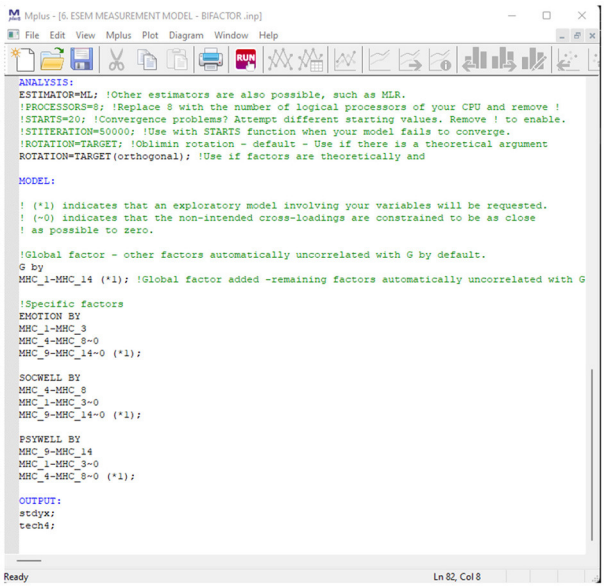

FIGURE 8 | Mplus syntax for a bifactor ESEM model of mental health.

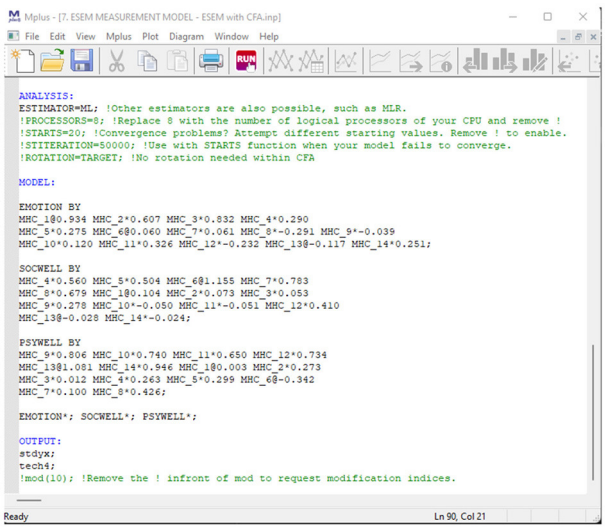

FIGURE 9 | Mplus syntax for the ESEM-within-CFA model of mental health.

ΔRMSEA $= -0.02$; ΔSRMR $= -0.01$; ΔAIC: -340.50; ΔBIC: -280.23). The ESEM Model 6 is therefore retained for further comparisons and the CFA Model 3 just retained for purposes of the tutorial.

Factor Correlations

The factor correlations between factors for the best fitting measurement models should be estimated and compared in the next step. Models where the lowest correlations between factors are shown, show that these models are able to better discriminate between factors. The model with the lowest factor correlations should be retained (20). However, given that a bifactor ESEM and bifactor CFA model fit the data the best in the current sample, inter-factor correlations cannot be computed as relationships are constrained to zero.

Item Level Parameters, Standardized Factor Loadings, and Reliability

In the final step, item level parameters and reliability indicators are reported. For the sake of transparency in this tutorial, item-level descriptive statistics (means, standard deviations, skewness, kurtosis) are also reported. Further, indicators of measurement quality (standardized factor loadings, standard errors, item-level residual variances, and item level R^2) and levels of reliability (omega and CITC) were computed for both the bifactor CFA (Model 3) and bifactor ESEM Models (Model 6).

The results summarized in **Table 5** show that items were relatively normally distributed [Skewness and Kurtosis $< +2; -2$: (89)], that each item was adequately associated with its overall a priori factor [CITC $r > 0.30$: (83)] and that the general and specific factors for both the ESEM and CFA models showed to be reliable.

Further, both Model 3 and Model 6 produced well-defined general factors (with all items $\lambda > 0.35$; small standard errors <0.04). Further, the specific factors were also relatively well-defined with item loadings matching expectations for both the Emotional Well-being and Social Well-being subscales. However, items MHC_11, MHC_13, and MHC_14 on the Psychological Well-being Subscale for both models showed non-significant factor loadings ($p > 0.01$). Further, items MHC_9, and MHC_12 only produced small, yet significant, factor loadings ($\lambda < 0.39$) on the Psychological Well-being factor for both models.

Finally, the level of reliability of the two models and their subscales were computed using Dueber's (84) calculator. The results, summarized in **Table 6**, show that the proportion of the common variance explained by the specific and general factors (ECV), and the overall omega produced similar results. The bifactor ESEM model did, however, produce slightly higher ECV values for the General Mental Health Factor (ECV = 0.70), and the Emotional- (ECV = 0.48) and Social well-being (ECV = 0.42) subscales. Both models produce similar, if not equivalent, levels of reliability with Omega exceeding the suggested cut-off criteria. However, when accounting for the presence of the general factor, the specific factors for neither models produced adequate Omega_{hs} levels ($\text{Omega}_{\text{hs}} < 0.70$).

Taken together, the results show that the bifactor ESEM model fitted the data proportionally better than the bifactor CFA model and produced slightly better parameter estimates. As such, the bifactor ESEM model is retained for potential further analyses.

Further Analysis: Demonstration of ESEM-Within-CFA in a Structural Model

Although the results showed that the bifactor ESEM model should be retained for further analysis, it does not afford the possibility to demonstrate the full usefulness or function of the ESEM-within-CFA framework [cf. (24)]. As stated previously, regular ESEM models can not directly be used within more

TABLE 4 | Comparing CFA and ESEM models.

	Model	Model comparison	$\Delta\chi^2$	Δdf	ΔCFI	ΔTLI	ΔRMSEA	ΔSRMR	ΔAIC	ΔBIC	Differences
Model 3	Bifactor model	M6 vs. M3	−596.46	27	0.056	0.061	−0.027	−0.028	−550.46	−424.01	Yes
		M4/7 vs. M3	−233.96	−12	0.022	0.012	−0.004	−0.016	−209.96	−143.99	Yes
Model 4	Three first-order ESEM	M6 vs. M4	−362.50	39	0.034	0.049	−0.023	−0.012	−340.50	−280.02	Yes
Model 6	Bifactor ESEM	M6 vs. M7	−362.50	390	0.034	0.049	−0.023	−0.012	−340.50	−280.02	Yes
Model 7	ESEM within CFA	M7 vs. M4	0.000	0	0.000	0.000	0.000	0.000	0.00	0.00	No

χ^2, Chi-square; df, degrees of freedom; TLI, Tucker-Lewis Index; CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation [90%CI]; SRMR, Standardized Root Mean Square Residual; AIC, Akaike Information Criterion; BIC, Bayes Information Criterion.

complex estimation procedures (e.g., structural models) as they would produce convergence problems, and therefore the ESEM-within-CFA framework should be used to re-specify an ESEM model within a CFA framework. Within the current tutorial, the ESEM-within-CFA approach was already used to estimate the higher-order ESEM model (Model 5), however, the three first-order factorial ESEM model (Model 4) fitted the data better. Therefore, to demonstrate ESEM-within-CFA function, Model 3 was respecified as an ESEM-within-CFA model to produce Model 7. This model, which should produce exactly the same model fit statistics as the normal ESEM model, could therefore be used for more complex analysis.

Based on Westerhof and Keyes' (54) assertion that mental health and mental illness are on separate, yet related, continuums, the relationship between common mental health problems, and mental health was investigated. As such, a structural path model was estimated based on the three first-order ESEM-within-CFA model of mental health and a traditional CFA model for mental illness as measured by the Brief Symptoms Inventory (cf. **Figure 10**). In this model, the common mental health problems were specified as the exogenous (input) factor and regressed on the three components of mental health (emotional well-being, social well-being, and psychological well-being) as the endogenous (outcome) factors.

The structural model showed acceptable fit: $\chi^2_{(202,\ N\ =\ 1{,}084)} = 1{,}419.42$ ($p < 0.01$; TLI = 0.92; CFI = 0.93; RMSEA = 0.06 [0.055–0.061]; SRMR 0.03; AIC = 116579.93; BIC = 117113.22; aBIC = 116805.05). The results showed that overall common mental health problems explained 35.5% of the variance in Emotional Well-being (β: −0.60; S.E: 0.02; $p < 0.01$), 1.5% in Social Well-being (β: −0.12; S.E: 0.03; $p < 0.01$) and 8.9% in Psychological Well-being (β: 0.30; S.E: 0.03; $p < 0.01$). This implies that higher levels of overall common mental health problems is associated with lower levels of emotional-, psychological-, and social well-being.

CONCLUSION

The purpose of this paper was to illustrate the applicability of ESEM as an alternative to traditional CFA approaches when evaluating the factorial validity of an instrument. By using the MHC-SF as an example, we provided (a) a brief overview of ESEM (and different ESEM models), (b) structured guidelines on how to estimate, compare, report, and interpret ESEM models, and (c) a step-by-step guide on how to produce ESEM syntaxes for Mplus with an innovative online tool. The results of this study highlight the value of ESEM, over and above that of traditional confirmatory factor analytical approaches. The study results also show practical implications for measuring mental health with the MHC-SF, by illustrating that a bifactor ESEM model fits the data significantly better than any other empirical model.

This tutorial demonstrates that restrictive CFA models for the MHC-SF, where items are constrained to only load onto their respective subscales, are insufficient to provide good model fit and adequately describe the data. Specifically, the

TABLE 5 | Item level descriptive statistics, factor loadings, and reliability indicators of Model 3 and Model 6.

Factor	Item	Mean	SD	Skewness	Kurtosis	CITC	Model 3–bifactor CFA model					Model 6–bifactor ESEM model										
							G_{factor}		S_{factor}			G_{factor}			EWB S_{factor}		SWB S_{factor}			PWB S_{factor}		
							λ	S.E.	λ	S.E.	δ	λ	S.E.	R^2	λ	S.E.	λ	S.E.	R^2	λ	S.E.	δ
Emotional well-being																						
	MHC_1	4.41	1.17	−0.78	0.26	0.72	**0.57**	0.02	**0.59**	0.02	0.32	**0.57**	0.02	0.33	**0.62**	0.02	−0.01	0.02	0.00	−0.05	0.02	0.29
	MHC_2	4.85	1.11	−1.11	1.11	0.65	**0.63**	0.02	**0.40**	0.02	0.44	**0.62**	0.02	0.39	**0.41**	0.02	0.02	0.02	0.00	0.08	0.02	0.44
	MHC_3	4.72	1.04	−0.92	0.90	0.72	**0.55**	0.02	**0.64**	0.02	0.29	**0.54**	0.02	0.29	**0.62**	0.02	−0.01	0.02	0.00	0.10	0.02	0.31
Social well-being																						
	MHC_4	3.60	1.46	−0.20	−0.91	0.54	**0.59**	0.02	**0.23**	0.03	0.60	**0.58**	0.02	0.33	0.12	0.02	**0.24**	0.03	0.06	−0.11	0.02	0.58
	MHC_5	3.75	1.62	−0.33	−1.07	0.47	**0.53**	0.02	**0.18**	0.03	0.69	**0.52**	0.02	0.27	0.11	0.02	**0.18**	0.03	0.03	−0.16	0.02	0.67
	MHC_6	2.35	1.29	0.60	−0.61	0.54	**0.38**	0.02	**0.65**	0.04	0.44	**0.38**	0.02	0.15	0.02	0.02	**0.62**	0.03	0.38	−0.10	0.02	0.46
	MHC_7	3.54	1.36	−0.21	−0.83	0.52	**0.48**	0.02	**0.45**	0.03	0.57	**0.47**	0.02	0.22	0.02	0.02	**0.46**	0.03	0.22	0.10	0.02	0.55
	MHC_8	3.40	1.46	−0.13	−1.01	0.46	**0.45**	0.02	**0.36**	0.03	0.67	**0.46**	0.02	0.21	−0.18	0.02	**0.40**	0.03	0.16	0.13	0.03	0.58
Psychological well-being																						
	MHC_9	3.99	1.35	−0.59	−0.43	0.64	**0.69**	0.02	**−0.17**	0.02	0.50	**0.71**	0.02	0.50	−0.07	0.02	0.16	0.02	0.03	**0.23**	0.04	0.42
	MHC_10	4.71	1.16	−1.09	1.02	0.58	**0.65**	0.02	**−0.74**	0.02	0.02	**0.67**	0.03	0.45	0.02	0.02	−0.04	0.02	0.00	**0.47**	0.05	0.33
	MHC_11	4.61	1.23	−0.84	0.16	0.60	**0.67**	0.02	−0.02	0.02	0.55	**0.67**	0.02	0.44	0.17	0.02	−0.08	0.02	0.01	−0.01	0.04	0.52
	MHC_12	3.36	1.58	−0.07	−1.14	0.49	**0.58**	0.02	**0.21**	0.02	0.62	**0.59**	0.02	0.35	−0.19	0.02	0.12	0.02	0.01	**−0.34**	0.05	0.48
	MHC_13	4.08	1.35	−0.53	−0.45	0.67	**0.71**	0.01	0.01	0.02	0.50	**0.73**	0.01	0.54	−0.12	0.02	−0.07	0.02	0.00	−0.06	0.05	0.44
	MHC_14	4.37	1.44	−0.80	−0.27	0.67	**0.76**	0.01	−0.02	0.02	0.42	**0.76**	0.01	0.57	0.09	0.02	−0.08	0.02	0.01	−0.02	0.04	0.42

Bold items, *Significant target loadings ($p < 0.05$); Underlined items indicate cross-loading items; S.E., standard error; CICT, Corrected item total correlation; λ, Standardized factor loadings; S.E., Standard Error; δ, Item Uniqueness.*

TABLE 6 | Reliability estimates and explained common variance.

	Model 3: bifactor CFA			Model 6: bifactor ESEM		
	ECV	Omega	Omega hs	ECV	Omega	Omega hs
General factor	0.68	0.92	–	0.70	0.92	–
Emotional well-being	0.47	0.85	0.39	0.48	0.84	0.40
Social well-being	0.41	0.76	0.28	0.42	0.76	0.29
Psychological well-being	0.19	0.87	0.03	0.12	0.86	0.01

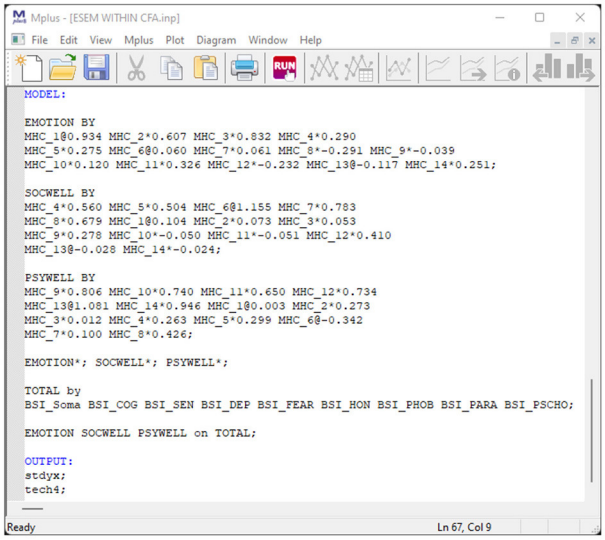

FIGURE 10 | Mplus syntax for the relationship between BSI and ESEM-within-CFA model of mental health.

results showed that neither the original correlated three-factor CFA model for mental health (55), nor a higher-order factorial model (58) could be confirmed. Similarly, the bifactor CFA model proposed by Jovanovic (15), where mental health is seen as a general factor, and emotional-, social-, and psychological well-being are positioned as specific factors, only partially fitted the data. Without specific- and deliberate modifications (e.g., correlating error variances of several items), the data would not adequately represent these models. Therefore, when cross-loadings are constrained, it undermines the measurement model.

In contrast, except for the Hierarchical ESEM Model, all ESEM models seemed to fit the data. Comparatively, the three first-order ESEM models comprised emotional-, psychological-, and social well-being. The bifactor ESEM model fitted the data significantly better than the CFA models. Therefore, it would seem that the less restrictive models, that accounts for small cross-loadings between items, overcame the limitations to the overly restrictive CFA models in terms of both model fit and indicators of measurement quality. These findings are in line with other ESEM studies on the MHC-SF that reported generally better data-model fit and stronger factor loadings compared to CFA models (46, 61, 62). Further, our results are in line with those of Lamborn et al. (61) that showed significant support for a bifactor ESEM model. Here, mental health is better represented by an overall general mental health factor, which is different from, three specific factors of well-being. The general mental health factor accounted for a large proportion of the variance and showed adequate levels of reliability. The three specific factors were also adequately represented by the data, however, these only accounted for a small proportion of the overall variance. This is similar to the findings of Lamborn et al. [(61), p. 15], who also argued that researchers should, therefore, "exercise caution when using and interpreting mean specific subscores" of these bifactor models. The relative strength of the g-factor in this model is not surprising, as studies comparing hedonic and eudaimonic conceptualisations of mental health have shown consistent support for a more general conceptualization of well-being and mental health (90). Our results, therefore, support ever-growing evidence in the literature for a tripartite model of mental health (61).

Therefore, mental health researchers are encouraged to incorporate ESEM into their measurement modeling strategies and structurally compare such to traditional CFA approaches. Given that mental health (and its three components) are structurally and theoretically linked, it requires researchers to apply less restrictive ESEM models, because cross-loadings between factors are inherently expected (62). Failure to employ these measures may lead to "a premature dismissal of central aspects of mental health" [(62), p. 11] and create unwarranted speculations within the literature.

Although this tutorial illustrated the applicability of ESEM when attempting to explore the factorial validity of an instrument, it is not without its limitations. The data employed in this study is derived from self-report measures, is cross-sectional, and participants were remunerated for their responses. This implies that there may be some biased results, and responses may not accurately represent reality. Further, there is currently no accepted means to account for the cross-loadings in ESEM models when estimating the scales' reliability. Therefore, omegas produced may not adequately represent the reliability of ESEM models, and direct comparisons of these with traditional CFA models should be made with caution. Further, the cut-off criteria to establish model fit are still based on CFA principles and the maximum likelihood estimation method. Although researchers are cautioned against stringent reliance on rules of thumb and cut-off scores, they do provide some form of standardization that aids in interpreting results. Therefore, simulation studies are required to determine the relevance and functioning of these goodness-of-fit criteria in relation to ESEM model estimation. Further, given the flexibility of ESEM, it is difficult to demonstrate its full potential and to articulate a full use case with a single dataset clearly. This tutorial provides a gentle introduction to the estimation and exploration of the factorial structure of a single instrument. Future tutorials should aim to incorporate more complex use cases such as auto-regressive modeling, measurement invariance, cross-lagged panel analysis, and the like.

Notwithstanding these limitations, our tutorial provides an illustrative example of approaching and estimating ESEM models with Mplus through an easy to use code generating tool. It also attempted to provide some suggested guidelines for approaching an ESEM related study. We hope that this tutorial and tool will aid researchers in incorporating ESEM into their model estimation approaches and provide more realistic and thorough evidence of their findings.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

LZ and PK conceptualized the tutorial. LZ analyzed the data, drafted the first version of the manuscript, and compiled the **Supplementary Material**. PK provided substantial conceptual input on the manuscript throughout the drafting process. All authors read through and approved the final manuscript before submission.

ACKNOWLEDGMENTS

We would like to extend their sincere appreciation to Prof. Leon T. De Beer for his substantial conceptual contributions to the design of the manuscript and for his insightful comments on earlier drafts of the manuscript. Further, we would like to extend a word of gratitude to all three reviewers of this manuscript who provided substantial and empowering feedback to enhance the quality of the work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.795672/full#supplementary-material

REFERENCES

1. Donaldson SI, Van Zyl LE, Donaldson SI. PERMA+4: a framework for work-related wellbeing, performance and positive organizational psychology 2.0. *Front Psychol*. (2021) 13:1–13.
2. Khademi R, Najafi M. Tracing the historical roots of positive psychology by reference publication year spectroscopy (RPYS): a scientometrics perspective. *Curr Psychol.* (2020) 39:438–44. doi: 10.1007/s12144-018-0044-z
3. Martín-del-Río B, Neipp MC, García-Selva A, Solanes-Puchol A. Positive organisational psychology: a bibliometric review and science mapping analysis. *Int J Environ Res Public Health*. (2021) 18:5222. doi: 10.3390/ijerph18105222
4. Gallagher MW, Lopez SJ. *Positive Psychological Assessment: A Handbook of Models and Measures*. Washington, DC: American Psychological Association (2002).
5. Efendic E, Van Zyl LE. On reproducibility and replicability: Arguing for open science practices and methodological improvements at the South African Journal of Industrial Psychology. *SA J Ind Psychol.* (2019) 45:a1607. doi: 10.4102/sajip.v45i0.1607
6. Earp BD, Trafimow D. Replication, falsification, and the crisis of confidence in social psychology. *Front Psychol*. (2015) 6:621 doi: 10.3389/fpsyg.2015.00621
7. Van Zyl LE. Enhancing scientific credibility: an open science strategy for the South African Journal of Industrial Psychology. *SA J Ind Psychol.* (2019) 45:a1768. doi: 10.4102/sajip.v45i0.1768
8. Wong PTP, Roy S. Critique of positive psychology and positive interventions. In: Brown NJL, Lomas T, Eiroa-Orosa FJ, editors. *The Routledge International Handbook of Critical Positive Psychology*. London: Routledge (2017). doi: 10.4324/9781315659794-12
9. Laher S. Psychological assessment in Africa: The time is now!. *Afr J Psychol Assess.* (2020) 1:1–3. doi: 10.4102/ajopa.v1i0.11
10. Snow NE. Positive psychology, the classification of character strengths and virtues, and issues of measurement. *J Posit Psychol.* (2019) 14:20–31. doi: 10.1080/17439760.2018.1528376
11. Warren MA, Donaldson SI. *Scientific Advances in Positive Psychology*. New York, NY: Praeger (2017).
12. Brown NJL, Sokal AD, Friedman HL. The complex dynamics of wishful thinking: the critical positivity ratio. *Am Psychol.* (2013) 68:801–13. doi: 10.1037/a0032850
13. Brown NJL, Sokal AD, Friedman HL. Positive psychology and romantic scientism. *Am Psychol.* (2014) 69:636–7. doi: 10.1037/a0037390
14. Keyes CLM. Mental illness and/or mental health? investigating axioms of the complete state model of health. *J Consult Clin Psychol.* (2005) 73:539–48. doi: 10.1037/0022-006X.73.3.539
15. Jovanović V. Structural validity of the Mental Health Continuum-Short Form: the bifactor model of emotional, social and psychological wellbeing. *Pers Ind Diff.* (2015) 75:154–9. doi: 10.1016/j.paid.2014.11.026
16. Keyes CL, Wissing M, Potgieter JP, Temane M, Kruger A, Van Rooy S. Evaluation of the mental health continuum–short form (MHC–SF) in setswana-speaking South Africans. *Clin Psychol Psychother.* (2008) 15:181–92. doi: 10.1002/cpp.572
17. Lamers SM, Westerhof GJ, Bohlmeijer ET, ten Klooster PM, Keyes CL. Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF). *J Clin Psychol.* (2011) 67:99–110. doi: 10.1002/jclp.20741
18. Zemojtel-Piotrowska M, Piotrowski JP, Osin EN, Cieciuch J, Adams BG, Ardi R, et al. The mental health continuum-short form: The structure and application for cross-cultural studies–A 38 nation study. *J Clin Psychol.* (2018) 74:1034–52. doi: 10.1002/jclp.22570
19. Keyes CLM, Shmotkin D, Ryff CD. Optimizing well-being: The empirical encounter of two traditions. *J Pers Soc Psychol.* (2002) 82:1007–22. doi: 10.1037/0022-3514.82.6.1007
20. Morin AJS, Myers ND, Lee S. Modern factor analytic techniques: bifactor models, exploratory Exploratory Structural Equation Modeling (ESEM),19 structural equation modeling and bifactor-ESEM. In: Tenenbaum G, Eklund RC, editors. *Handbook of Sport Psychology, 4th Edition, Vol. 2*. New York, NY: Wiley Publishers (2020). p. 1044–73. doi: 10.1002/9781119568124.ch51
21. Marsh HW, Muthén B, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJ, et al. Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. *Struct Equ Model.* (2009) 16:439–76. doi: 10.1080/10705510903008220
22. Marsh HW, Morin AJ, Parker PD, Kaur G. Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annu Rev Clin Psychol.* (2014) 10:85–110. doi: 10.1146/annurev-clinpsy-032813-153700
23. Brown, T. A. (2015). *Confirmatory factor analysis for applied research (2nd ed.)*. New York, NY: Guilford Press.
24. Morin AJS, Arens A, Marsh H. A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Struct Equ Model.* (2016) 23:116–39. doi: 10.1080/10705511.2014.961800
25. Joshanloo M, Jovanović V. The factor structure of the mental health continuum-short form (MHC-SF) in Serbia: an evaluation using exploratory structural equation modeling. *J Ment Health.* (2017) 26:510–5. doi: 10.1080/09638237.2016.1222058
26. Joreskog KG. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika.* (1969) 34:183–202. doi: 10.1007/BF02289343
27. Cao C, Liang X. Sensitivity of fit measures to lack of measurement invariance in exploratory structural equation modeling. *Struct Equat Model.* (2021) 2021:1–11. doi: 10.1080/10705511.2021.1975287
28. Marsh HW. Application of confirmatory factor analysis and structural equation modeling insport/exercise psychology. In: Tenenbaum G, Eklund

RC, editors. *Handbook of Sport Psychology (3rd Ed). New York, NY: Wiley Publishers.* (2007). p. 774–98. doi: 10.1002/9781118270011.ch35
29. Xiao Y, Liu H, Hau K-T. A comparison of CFA, ESEM, and BSEM in test structure analysis. *Struct Equ Model.* (2019) 26:665–77. doi: 10.1080/10705511.2018.1562928
30. Asparouhov T, Muthén B. Exploratory structural equation modeling. *Struct Equat Model.* (2009) 16:397–438. doi: 10.1080/10705510903008204
31. Mai Y, Zhang Z, Wen Z. Comparing exploratory structural equation modeling and existing approaches for multiple regression with latent variables. *Struct Equat Model.* (2018) 25:737–49. doi: 10.1080/10705511.2018.1444993
32. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model.* (1999) 6:1–55. doi: 10.1080/10705519909540118
33. McNeish D, An J, Hancock GR. The thorny relation between measurement quality and fit index cut-offs in latent variable models. *J Pers Assess.* (2018) 100:43–52. doi: 10.1080/00223891.2017.1281286
34. McNeish D, Hancock GR. The effect of measurement quality on targeted structural model fit indices: a comment on Lance, Beck, Fan, and Carter (2016). *Psychol Methods*. (2018) 23:184–90. doi: 10.1037/met0000157
35. Shi D, Lee T, Maydeu-Olivares A. Understanding the model size effect on SEM fit indices. *Educ Psychol Meas.* (2019) 79:310–34. doi: 10.1177/0013164418783530
36. Gucciardi DF, Zyphur MJ. Exploratory structural equation modelling and Bayesian estimation. In: Ntoumanis N, Myers ND, editors. *An Introduction to Intermediate and Advanced Statistical Analyses for Sport and Exercise Scientists.* New Jersey, NY: Wiley (2016). p. 172–94.
37. Morin AJS. Exploratory structural equation modeling. In: Hoyle RH, editor. *Handbook of Structural Equation Modeling, Second Edition.* New York, NY. Guilford Press (2021).
38. Van Zyl LE, Heijenk B, Klibert J, Shankland R, Verger NB, Rothmann S, et al. *Grit across* nations: an investigation into the cross-national equivalence of the Grit-O Scale. *J Happiness Stud.* (2021).
39. Marsh HW, Guo J, Dicke T, Parker PD, Craven RG. Confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM), and set-ESEM: optimal balance between goodness of fit and parsimony. *Multivariate Behav Res.* (2020) 55:102–19. doi: 10.1080/00273171.2019.1602503
40. Tóth-Király I, Morin AJ, Gillet N, Bothe B, Nadon L, Rigó A, et al. Refining the assessment of need supportive and need thwarting interpersonal behaviors using the bifactor exploratory structural equation modeling framework. *Curr Psychol.* (2020) 2020:1–15. doi: 10.1007/s12144-020-00828-8
41. Tóth-Király I, Bõthe B, Rigó A, Orosz G. An illustration of the exploratory structural equation modeling (ESEM) framework on the passion scale. *Front Psychol.* (2017) 8:1968. doi: 10.3389/fpsyg.2017.01968
42. Morin AJS, Marsh HW, Nagengast B. Chapter 10. exploratory structural equation modeling. In: Hancock GR, Mueller RO, editors. *Structural Equation Modeling: A Second Course (2nd ed.).* Charlotte, NC: Information Age Publishing, Inc. (2013).
43. Lovibond SH, Lovibond PF. *Manual for the Depression Anxiety Stress Scales (2nd. Ed.).* Sydney, NSW: Psychology Foundation (1995). doi: 10.1037/t01004-000
44. Morin AJS, Asparouhov T. *Estimation of a Hierarchical Exploratory Structural Equation Model (ESEM) Using ESEM-Within-CFA*. Montreal, QC: Substantive Methodological Synergy Research Laboratory (2018).
45. Howard JL, Gagné M, Morin AJ, Forest J. Using bifactor exploratory structural equation modeling to test for a continuum structure of motivation. *J Manag.* (2018) 44:2638–64. doi: 10.1177/0149206316645653
46. Longo Y, Jovanović V, Sampaio de Carvalho J, Karaś D. The general factor of wellbeing: multinational evidence using bifactor ESEM on the mental health continuum–short form. *Assessment*. (2020) 27:596–606. doi: 10.1177/1073191117748394
47. Rogoza R, Truong Thi KH, Rózycka-Tran J, Piotrowski J, Zemojtel-Piotrowska M. Psychometric properties of the MHC-SF: an integration of the existing measurement approaches. *J Clin Psychol.* (2018) 74:1742–58. doi: 10.1002/jclp.22626
48. Sànchez-Oliva D, Morin AJS, Teixeira PJ, Carraça EV, Palmeira AL, Silva MN. A bifactor-exploratory structural equation modelling representation of the structure of basic psychological needs at work scale. *J Vocat Behav.* (2017) 98:173–87. doi: 10.1016/j.jvb.2016.12.001
49. Van Zyl LE, Olckers C, Roll LC. The psychometric properties of the Grit-O scale within the Twente region in Netherlands: An ICM-CFA vs. ESEM approach. *Front Psychol.* (2020) 11:796. doi: 10.3389/fpsyg.2020.00796
50. Muthén LK, Muthén BO. *Mplus (Version 8.6)* [Statistical software]. Los Angeles, CA: Muthén and Muthén (2021).
51. Van Zyl LE, Rothmann S. *Evidence-Based Positive Psychological Intervention Practices in Multicultural Contexts.* Cham: Springer (2019). doi: 10.1007/978-3-030-20311-5
52. Van Zyl LE, Roll LC, Stander MW, Richter S. Positive psychological coaching definitions and models: a systematic literature review. *Front Psychol.* (2020) 11:793. doi: 10.3389/fpsyg.2020.00793
53. World Health Organization. *Promoting Mental Health: Concepts, Emerging Evidence, Practice: Summary Report*. Geneva: World Health Organization (2004).
54. Westerhof GJ, Keyes CL. Mental illness and mental health: The two continua model across the lifespan. *J Adult Dev.* (2010) 17:110–9. doi: 10.1007/s10804-009-9082-y
55. Keyes CL. The mental health continuum: from languishing to flourishing in life. *J Health Soc Behav.* (2002) 43:207–22.
56. Keyes CL, Yao J, Hybels CF, Milstein G, Proeschold-Bell RJ. Are changes in positive mental health associated with increased likelihood of depression over a two year period? A test of the mental health promotion and protection hypotheses. *J Affect Disord.* (2020) 270:136–42.
57. Richter S, Van Zyl LE, Roll LC, Stander MW. Positive psychological coaching tools: a systematic literature review. *Front Psychiatry.* (2021) 12:667200. doi: 10.3389/fpsyt.2021.667200
58. Van Zyl LE, Olckers C. The mental health continuum-short form in organisational contexts: factorial validity, invariance, and internal consistency. *Eur J Ment Health.* (2019) 14:230–59. doi: 10.5708/EJMH.14.2019.2.2
59. Keyes CL, Grzywacz JG. Health as a complete state: the added value in work performance and healthcare costs. *J Occup Environ Med.* (2005) 47:523–32. doi: 10.1097/01.jom.0000161737.21198.3a
60. Gignac GE. The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intell.* (2016) 55:57–68. doi: 10.1016/j.intell.2016.01.006
61. Lamborn P, Cramer KM, Riberdy A. The structural validity and measurement invariance of the mental health continuum-short form (MHC-SF) in a large Canadian sample. *J Well-Being Asses*. (2016) 2:1–19.
62. Joshanloo M, Lamers SM. Reinvestigation of the factor structure of the MHC-SF in the Netherlands: contributions of exploratory structural equation modeling. *Pers Individ Differ*. (2018) 97:8–12. doi: 10.1016/j.paid.2016.02.089
63. De Beer LT, Van Zyl LE. *ESEM Code Generator for Mplus.* (2019). Retrieved from: https://www.surveyhost.co.za/esem/ (accessed December 8, 2021).
64. Buchanan EM, Scofield JE. Methods to detect low quality data and its implication for psychological research. *Behav Res Methods.* (2018) 50:2586–96. doi: 10.3758/s13428-018-1035-6
65. De Beurs E, Zitman F. De Brief Symptom Inventory (BSI): De betrouwbaarheid en validiteit van een handzaam alternatief. *Maandblad Geestelijke Volksgezondheid.* (2006) 61:120–41.
66. Morrison TG, Morrison MA, McCutcheon JM. Best practice recommendations for using structural equation modelling in psychological research. *Psychol.* (2017) 8:1326–41. doi: 10.4236/psych.2017.89086
67. Guo J, Marsh HW, Parker PD, Dicke T, Lüdtke O, Diallo TM. A systematic evaluation and comparison between exploratory structural equation modeling and Bayesian structural equation modeling. *Struct Equat Model.* (2019) 26:529–56. doi: 10.1080/10705511.2018.1554999
68. Kyriazos TA. Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology.* (2018) 9:2207–30. doi: 10.4236/psych.2018.98126
69. Myers ND, Ntoumanis N, Gunnell KE, Gucciardi DF, Lee S. A review of some emergent quantitative analyses in sport and exercise psychology. *Int Rev Sport Exerc Psychol.* (2018) 11:70–100. doi: 10.1080/1750984X.2017.1317356

70. Hancock GR, Freeman MJ. Power and sample size for the root mean square error of approximation test of not close fit in structural equation modeling. *Educ. Psychol Measur*. (2016) 61:741–58. doi: 10.1177/00131640121971491
71. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychol Methods*. (1996) 1:130–49. doi: 10.1037/1082-989X.1.2.130
72. Maxwell SE, Kelley K, Rausch JR. Sample size planning for statistical power and accuracy in parameter estimation. *Ann Rev Psychol*. (2007) 59:537–63. doi: 10.1146/annurev.psych.59.103006.093735
73. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educ Psychol Meas*. (2013) 73:913–34. doi: 10.1177/0013164413495237
74. Satorra A, Saris WE. Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*. (1985) 50:83–90. doi: 10.1007/BF02294150
75. Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. *Struct Equ Model*. (2002) 9:599–620. doi: 10.1207/S15328007SEM0904_8
76. Hair J Jr, Black W, Babin B, Anderson R. *Multivariate Data Analysis (7th ed.)*. London: Pearson New International Edition (2014).
77. Kline RB. *Principles and Practice of Structural Equation Modeling*. 4th ed. New York, NY: Guilford publications (2015). p. 96–106.
78. Shi D, Maydeu-Olivares A, Rosseel Y. Assessing fit in ordinal factor analysis models: SRMR vs. *RMSEA Struct Equat Model*. (2020) 27:1–15. doi: 10.1080/10705511.2019.1611434
79. Brown TA. *Confirmatory Factor Analysis For Applied Researchers*. New York, NY: Guilford publications (2006).
80. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Model*. (2017) 14:464–504. doi: 10.1080/10705510701301834
81. Wong J, Wong X. (2020). *Structural Equation Modelling: Applications Using Mplus*, 2nd ed. Chichester: Wiley & Sons. p. 536.
82. Murray AL, Johnson W. The limitations of model fit in comparing the bifactor versus higher-order models of human cognitive ability structure. *Intell*. (2013) 41:407–22. doi: 10.1016/j.intell.2013.06.004
83. Zijlmans EA, Tijmstra J, van der Ark LA, Sijtsma K. Item-score reliability as a selection tool in test construction. *Front Psychol*. 9:2298. doi: 10.3389/fpsyg.2018.02298
84. Dueber DM. *Bifactor Indices Calculator: A Microsoft Excel-Based Tool to Calculate Various Indices Relevant to Bifactor CFA Models*. (2017). Available online at: http://sites.education.uky.edu/apslab/resources/ (accessed December 8, 2021).
85. Raykov T. Evaluation of scale reliability for unidimensional measures using latent variable modeling. *Meas Eval Couns Dev*. (2009) 42:223–32. doi: 10.1177/0748175609344096
86. Hayes AF, Coutts JJ. Use omega rather than Cronbach's alpha for estimating reliability. *Commun Methods Measures*. (2020) 14:1–24. doi: 10.1080/19312458.2020.1718629
87. Morin AJS. Exploratory structural equation modeling. In: Hoyle RH, editor. *Handbook of Structural Equation Modeling, 2nd ed*. New York, NY: Guilford publications (In Press).
88. Jak S, Jorgensen TD, Verdam MGE, Oort FJ, Elffers L. Analytical power calculations for structural equation modeling: a tutorial and Shiny app. *Behav Res Methods*. (2021) 53:1385–406. doi: 10.3758/s13428-020-01479-0
89. Kim HY. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod*. (2013) 38:52–4. doi: 10.5395/rde.2013.38.1.52
90. Disabato DJ, Goodman FR, Kashdan TB, Short JL, Jarden A. Different types of well-being? A cross-cultural examination of hedonic and eudaimonic well-being. *Psychol Assess*. (2016) 28:471–82. doi: 10.1037/pas0000209

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

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

Copyright © 2022 van Zyl and ten Klooster. 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.

APPENDIX

TABLE A1 | Summary of ESEM estimation guidelines.

Step	Description	What to consider or report	Reported in manuscript	Where to report
Planning Phase				
1	Develop a clear theoretical model for the instrument based on the available literature	Briefly discuss alternative, theory-informed (CFA) factorial models which the instrument could manifest as. If factors within a CFA model are conceptually related, then there is also an argument for an ESEM model with a similar structure. Should a global factor be expected, a bifactor CFA should be described and an associated bifactor ESEM model mentioned.	Yes	Literature review
		For the validation of psychometric instruments, present clear hypotheses about the (CFA) factorial structure of the instrument and present alternative hypotheses that the ESEM models should, theoretically, provide a better representation of the data. For other studies, the relationships between exo- and endogenous factors should be clearly articulated	Yes	Literature review
2	Plan for the most appropriate sample size	Determine and plan for the most appropriate sample size for the study. This can be done in many ways [c.f. (70, 72, 88)]. Monte Carlo simulations are preferred [c.f. (73)]	No	
Data Preparation Phase				
3	Data cleaning, screening, and preparation	Screen the data for potential issues (e.g. outliers), and prepare it for further analysis. Data quality checks should also be performed [c.f. (64)]	No	
		Decide upon an appropriate missing values strategy (e.g. Multiple imputations, FIML, sensitivity analysis)	Yes	Methods: statistical analysis
4	Determine the most appropriate software, estimation method, rotation and procedure for the analysis	Decide upon and report the software packages (and version number) that will be used for the analysis. ESEM is fully integrated in Mplus, but is currently only partially supported in R.	Yes	Methods: statistical analysis
		If data follows a multivariate normal distribution, employ the Maximum Likelihood Estimator in Mplus. If data is not normally distributed, either transform the data or use more robust estimation methods in Mplus (e.g. MLR, WLSMV). For all models with continuous indicators, the MLR estimator is also appropriate; models comprised of ordinal indicators WLSMV should be used.	Yes	Methods: statistical analysis
		Decide upon the most appropriate rotation method (Geomin / Target / Target Orthogonal). Geomin rotations (with an epsilon value of .5) for more exploratory approaches. Target rotations for confirmatory approaches. (Target) Orthogonal rotations are used for bifactor ESEM modeling.	Yes	Methods: statistical analysis
		Describe the analysis procedure to be employed		
Data Analysis and Reporting Phase				
5	Determine appropriate goodness-of-fit indices, and indicators of measurement quality	Decide upon which goodness-of-fit indices are most appropriate for the analyses (e.g., TLI/CFI, RMSEA, RMSEA Confidence Intervals etc). Report each index as well as the cut-off criteria to be considered. Multiple indicators are to be mentioned (c.f. **Table 2**). CFI/TLI/RMSEA should always be employed as the primary criterion.	Yes	Methods: statistical analysis
		Decide upon a priori indicators of measurement quality to be considered for the study (e.g. Standardized $\lambda > 0.40$; item uniqueness > 0.1 but < 0.9; cross-loading tolerance levels; overall R2). However, the results should be considered in the context of the study and what they might mean or indicate without being rigid about minor deviations from the chosen guidelines.	Yes	Methods: statistical analysis
6	Estimate and report the model fit indicators for competing CFA Models	Multiple measurement models need to be estimated and their model fit statistics reported. Only models with theoretical justification should be estimated. The following models could be estimated: (1) Unidimensional model, (2) correlated first-order factorial models, (3) second-order or 'hierarchal' factorial model, and (4) bifactor models	Yes	Results: competing measurement models

(Continued)

TABLE A1 | Continued

Step	Description	What to consider or report	Reported in manuscript	Where to report
		Report any potential modifications made to enhance model fit	Yes	Results: competing measurement models
		Tabulate all goodness-of-fit indices for all CFA models in a single table and indicate which models meet the pre-defined criteria mentioned and to make the comparison reader-friendly (refer to Point 5).	Yes	Results: competing measurement models
7	Estimate and report the model fit indicators for competing ESEM Models	Multiple ESEM measurement models should be estimated and model fit statistics reported. In principle, the ESEM alternatives to the traditional CFA models estimated in Point 6, should be reported. The following ESEM models could be estimated: (1) correlated first-order factorial ESEM models, (2) second-order or 'hierarchal' factorial ESEM model, (3) bifactor ESEM models and (4) ESEM within CFA models for use in structural models with other factors	Yes	Results: competing measurement models
		Tabulate all goodness-of-fit indices for all ESEM Models into the same table as the CFA models and indicate which models meet the pre-defined criteria mentioned in Point 5.	Yes	Results: competing measurement models
8	Compare CFA and ESEM models to determine the best fitting model for the data	CFA and ESEM models need to be compared against one another with the goodness-of-fit and measurement quality criteria mentioned in Point 5. Models that show comparatively better model fit should be retained for further analysis. It is, however, important to note that model fit should not be the only consideration, but the parameter estimates should also be closely inspected and considered.	Yes	Results: competing measurement models
		To retain ESEM models for further analysis, the following conditions need to be met:	Yes	Results: competing measurement models
		(a) The ESEM model should ideally show better data-model fit than the corresponding CFA model (including the same number of factors defined similarly). If the factor correlations for the ESEM model are smaller than those of the CFA model, then the ESEM model should be retained even if it fits as well as the CFA model.	Yes	Results: competing measurement models
		(b) For correlated factors models, the ESEM model should show reduced factor correlations	Yes	Results: competing measurement models
		(c) The ESEM model should only show small to medium cross-loadings. Should larger cross-loadings exist, then there should be a theoretical explanation presented for such. Perhaps there are 'wording' effects or some logic that researchers can use to explain this.	Yes	Results: competing measurement models
		(d) The estimated latent factors within the ESEM model should be well defined	Yes	Results: competing measurement models
		(e) Should there be multiple medium to large cross-loadings in the ESEM model, it could indicate support for the presence of a larger global factor, and therefore the bifactor ESEM model could be explored.	Yes	Results: competing measurement models
		(f) Additional factors to consider for bifactor models: This model should ideally show better data-model fit than the corresponding CFA and ESEM models, there should be a well-defined G-Factor (where all items load significantly on such), and reasonably well defined S-Factors (cross- and non-significant loadings are permitted). For bifactor models, model fit should not be the only indicator informing a decision to retain. Researchers should also inspect parameter estimates before making final decisions.	Yes	Results: competing measurement models
9	Report factorial correlations	For the final retained measurement model (s), the factor correlations should be reported. This cannot be done for bifactor Models . Smaller factor correlations mean better discrimination between factors. The model with the smallest factor correlation is usually retained, however, decisions should be based in the context of the other considerations (model fit, measurement quality and parameter estimates) mentioned earlier.	Yes	Results: factorial correlations

(Continued)

TABLE A1 | Continued

Step	Description	What to consider or report	Reported in manuscript	Where to report
10	Report and compare item level parameters and reliability	Item level parameters and indicators of measurement quality (standardized factor loadings, standard errors, item-level residual variances), as well as levels of reliability (composite reliability, or omega,), should be tabulated and reported.	Yes	Results: item level parameters
		Note, that if a bifactor CFA model is retained, editors or reviewers may request additional information such as the Explained Common Variance (ECV), the H-factor, the Factor Determinacy indicator, the Item level ECV, the Percent of Uncontaminated Correlations (PUC) and the Average Relative Bias Parameters could also be reported as additional indicators of reliability and measurement quality [For a tutorial cf. (84)].	Yes	Results: item level parameters
		Decide upon appropriate indicators of reliability for both the CFA and ESEM models, such as composite reliability [$\rho > 0.80$; (85)], or Mc Donald's Omega [$\omega > 0.70$; (86)].		
		Report the level of reliability for each (sub) scale of the instrument	Yes	Results: item level parameters
Further or Additional Analysis				
11	For further or additional analysis, the best fitting ESEM model is respecified as a CFA model through the ESEM-within-CFA estimation procedure. This affords the opportunity to use the ESEM-within-CFA model for more complex estimation procedures such as invariance testing, multi-group analysis, latent growth models, structural models and the like.		Yes	Results

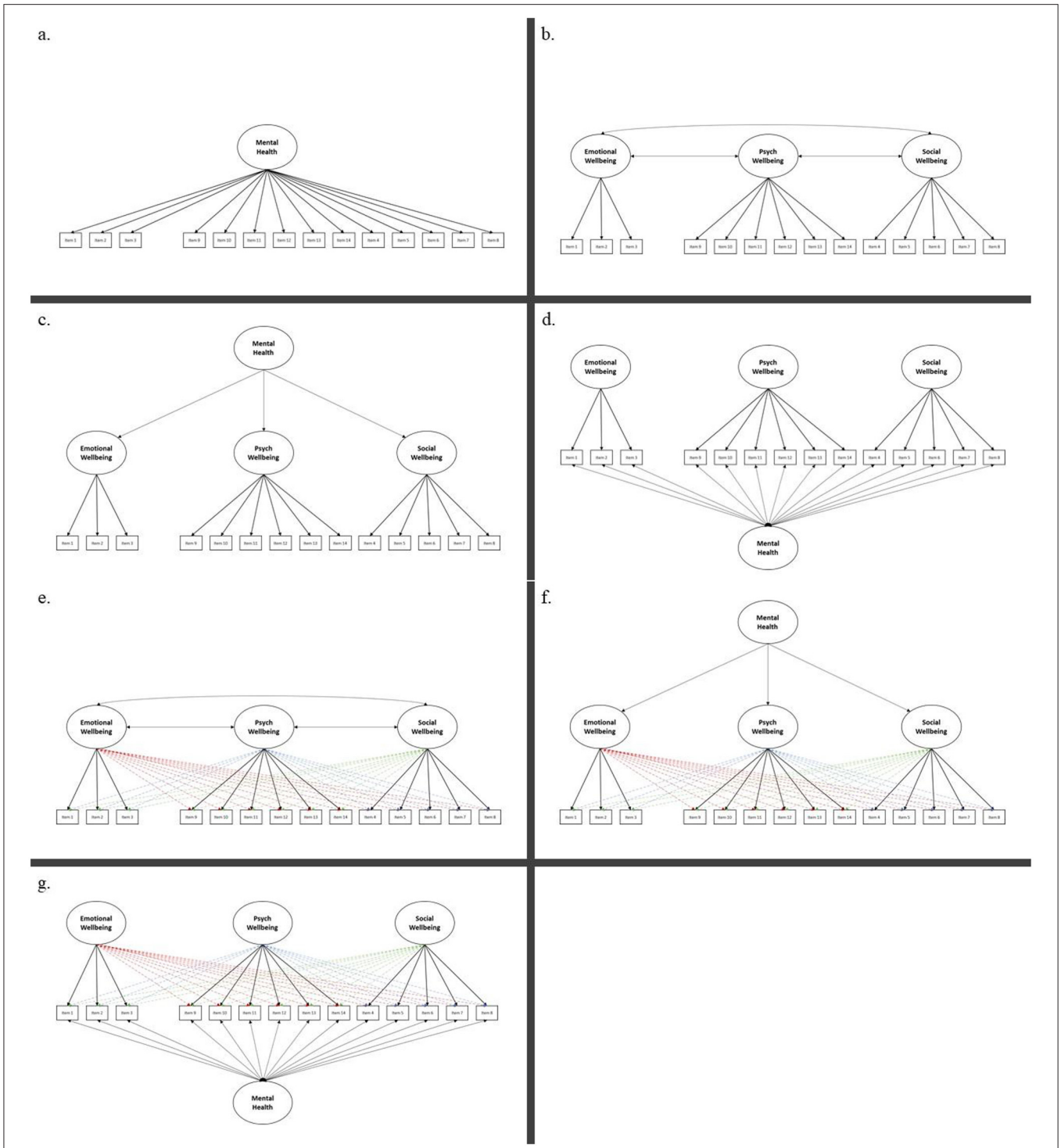

FIGURE A1 | Graphical representations of the MHC-SF's various factorial models.

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 January 2022
doi: 10.3389/fpsyg.2021.795931

Measuring What Counts in Life: The Development and Initial Validation of the Fulfilled Life Scale (FLS)

Doris Baumann and Willibald Ruch*

Department of Psychology, University of Zürich, Zurich, Switzerland

In a recent work, we introduced a theoretical model for fulfillment in life that covers cognitive and affective components and distinguishes different time frames. The present study evaluates this model and describes the construction of the Fulfilled Life Scale (FLS) to assess fulfillment regarding the whole lived life retrospectively. We investigated the scale in two samples (Sample 1: *N* = 282 adults aged 50–93 years; Sample 2: *N* = 406 adults aged 40–85 years). The model of the cognitive component combines three sources of fulfillment (*self, life, impact/legacy*) with three criteria (*wholeness, fit, value*), yielding nine facets. Employing hierarchical factor analysis, we inspected all solutions between one and nine. We identified three optimal factors, which we labeled *unfolded self and life*, *the worthwhile life,* and *positive impact and legacy*. Next, we selected marker items and replicated the factor structure in Sample 2. The three scales were positively intercorrelated and showed good internal consistency in both samples. For the affective component, exploratory and confirmatory factor analyses established a one-factor structure in both samples, and high internal consistency was obtained. Across a range of related constructs, we demonstrated construct and criterion validity. Notably, cognitive and affective fulfillment incrementally predicted a global rating of a fulfilled life and mental well-being, even after controlling for subjective and eudaimonic well-being. Overall, the study proves that the FLS is necessary to capture people's experience of a fulfilled life, which could not be assessed sufficiently with previous well-being measures. Both cognitive and affective fulfillment were able to predict additional variance in mental well-being. Moreover, the study reveals psychometric support for the FLS and presents the first evidence on its validity. Lastly, applications in research and practice are discussed, especially in the context of living and aging well in the second half of life.

Keywords: positive psychological assessment, wellbeing assessment, scale development, validation, positive psychology, life span, positive aging, fulfilled life

OPEN ACCESS

Edited by:
Peter ten Klooster,
University of Twente, Netherlands

Reviewed by:
Judith Glück,
University of Klagenfurt, Austria
Andreas Krafft,
University of St.Gallen, Switzerland

****Correspondence:***
Doris Baumann
d.baumann@psychologie.uzh.ch

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

Received: *15 October 2021*
Accepted: *10 December 2021*
Published: *11 January 2022*

Citation:
Baumann D and Ruch W (2022) Measuring What Counts in Life: The Development and Initial Validation of the Fulfilled Life Scale (FLS). Front. Psychol. 12:795931. doi: 10.3389/fpsyg.2021.795931

INTRODUCTION

If psychologists wish to improve the human condition, it is not enough to help those who suffer. The majority of "normal" people also need examples and advice to reach a richer and more fulfilling existence (Seligman and Csikszentmihalyi, 2000, p. 10).

From the inception of the field, positive psychology has emphasized fulfillment in life (FiL) as a central topic. Though the term has appeared regularly in the literature, virtually no related research has taken place. This gap in the research can be attributed to the lack of a theoretical

conceptualization and the absence of an instrument for its assessment (Baumann and Ruch, 2021). Consequently, we proposed a definition along with a theoretical model for FiL that distinguishes several time frames, from fulfillment in an activity to perceiving one's life as fulfilled. The present study serves as a next step, having the objective to test this model and develop a measure that will assess a fulfilled life. We decided to focus on a fulfilled life in retrospect, as taking the whole lived life into account for an evaluation seems to have the greatest relevance, more than a fulfilling activity or having had a fulfilling life at a particular life stage. What counts is that individuals can look back and arrive at the conclusion that their lives were fulfilled. The availability of such a scale and the resulting findings will contribute to a deeper understanding of the construct, facilitate building and expanding its nomological network, and may further stimulate research on FiL. In turn, insights into a fulfilled life may have important implications for practice. Because life satisfaction does not seem to be the sole criterion of how individuals evaluate their lives and how well they age (Westerhof et al., 2001), our scale measuring a fulfilled life can provide valuable insights from a different angle.

Conceptualization of a Fulfilled Life

We defined FiL as "a cognitive-affective experience referring to a sense of wholeness, fit, and value toward the self, one's life, and one's impact" (Baumann and Ruch, 2021, p. 6). Accordingly, a fulfilled life refers to the positive appraisal of the person one has become, how one has led one's life, and the impact one has made. The FiL model represents the cognitive component as a 3 × 3 matrix (see **Table 1**) that combines the three criteria for fulfillment *(wholeness, fit, and value)* with the three sources of fulfillment *(self, life, and impact/legacy)*. The sources constitute the three main strands from which individuals derive fulfillment when referring to life as a whole. The combination of the criteria and sources yields nine facets that are presented in the FiL model as outlined in **Table 1**: (a) *realized uniqueness*, (b) *a life lived fully,* (c) *the making of a positive difference*, (d) *authentic pursuits,* (e) *a life true to oneself*, (f) *a contribution reflecting the self*, (g) *worthwhile involvements*, (h) *a life that was worthwhile, and* (i) *a life that mattered to others*. **Table 1** displays a brief description of the nine facets (for more details, see Baumann and Ruch, 2021).

This arrangement of criteria and sources provides a systematized way to represent prior thinking. The experience of fulfillment requires certain qualities (the criteria) to be present to a sufficient extent. As delineated in our theoretical article (Baumann and Ruch, 2021), the first criterion of wholeness designates the extent to which one could become a whole person, live life fully, and positively influence others' lives. The criterion of fit refers to a sense of congruence and alignment and involves the perception that one was true to the self, lived a life that suited one deeply, and has been able to make a contribution that is reflective of what one holds dear. Value as a third qualitative requirement for fulfillment relates to the perception that one has invested one's own capacities well and lived a worthwhile and meaningful life; moreover, one's life will have had value and mattered to others. Affective fulfillment was considered to consists of low-arousal positive affect comprising such feelings as inner contentment, gratefulness, harmony with oneself and one's life, or inner peace (Baumann and Ruch, 2021). In addition, a fulfilled life is characterized by the absence of intense negative affective experiences, such as feelings of emptiness, deep regret, or disappointment.

TABLE 1 | The fulfillment in life (FiL) model depicting the cognitive-evaluative component.

Criteria for fulfillment	Sources of fulfillment		
	Self	**Life**	**Impact/Legacy**
Wholeness Sense of wholeness and completeness	*(A) Realized Uniqueness:* Fulfillment from having been able to become more fully oneself	*(B) A Life Lived Fully:* Fulfillment from realizing life goals and having lived life consciously	*(C) The Making of a Positive Difference:* Fulfillment from having been able to make a positive contribution and to leave something of value
Fit Sense of congruence and alignment	*(D) Authentic Pursuits:* Fulfillment from having had the courage to be true to oneself	*(E) A Life True to Oneself:* Fulfillment from having led a life that felt right	*(F) A Contribution Reflecting the Self:* Fulfillment from having been able to combine own values, talents, and interests while making a positive contribution
Value Sense of meaningfulness, significance, worthwhileness	*(G) Worthwhile Involvements:* Fulfillment from having used one's resources and potentialities sensibly	*(H) A Life that was Worthwhile:* Fulfillment from perceiving one's life as worthwhile and meaningful	*(I) A Life that Mattered to others:* Fulfillment from a sense that one's life mattered and made a positive difference to others

Rows represent criteria for fulfillment, columns represent sources of fulfillment, and the cells represent the major content of the nine cognitive facets of a fulfilled life.

Measuring a Fulfilled Life

A fulfilled life might be measured in different ways. Fulfillment is subjective in nature, which must be acknowledged in measurement. Even though developing a checklist of factors that are empirically confirmed to contribute to fulfillment is possible, the total score will not suffice unless each of the items is subjectively weighted. Peer evaluation might be hampered by a variety of biases but, at the same time, be useful as a validation criterion. A global subjective evaluation on an anchored rating scale might serve as a useful initial indicator, but there is no substitute for a genuine measurement. The anchoring may be established by having participants first describe their most apt example of a fulfilled life and then stipulate how close their own life is to their self-defined ideal. This latter approach is the one we chose for validation purposes.

Our preference has been to determine the contents of the evaluation by directly asking the questions that cover our model. Having economic measures for each of the nine facets from **Table 1** as an intermediary state is certainly of interest, but most importantly is a measure of its essence (in other words, the

TABLE 2 | Phases of scale development.

Phases	Aims	Sample	Data analysis
Phase 1: Substantive validity	• Development of initial item pool		
	• Expert review		
	• Cognitive pretest		
Phase 2: Structural validity I	• Data collection in a sample of the target group	Sample 1	• Descriptive statistics
			• Hierarchical factor analysis
	• Psychometric evaluation of items		• Exploratory factor analysis
	• Item preselection		• Parallel analysis
	• Examination of factor structure		• Minimum average partial test
	• Creation of provisional scales		• Reliability coefficients
Phase 3: Structural validity II	• Assessment of factor structure	Sample 2	• Exploratory factor analysis
	• Test of similarity of factors		• Confirmatory factor analysis
	• Examination of reliability in a new sample		• Tucker's Phi coefficients
			• Reliability coefficients
Phase 4: External validity	• Assessment of convergent validity	Sample 1 and 2	• Correlation analysis
	• Evaluation of discriminant validity	Sample 1	• Hierarchical regression analysis
	• Assessment of concurrent validity	Sample 1 and 2	
	• Assessment of incremental validity	Sample 1 and 2	• *t*-tests
	• Testing known groups validity	Sample 3	

factors underlying these facets). The results will reveal whether total scores for the cognitive and affective domains may be derived as well.

Another consideration relates to determining the best target audience for the scale being developed. We have set the target group to encompass middle-aged and older adults, as global life evaluations are more common with age (Westerhof et al., 2001); moreover, taking stock of one's life requires a certain number of experiences. Some of what truly holds value or is significant might be recognizable only in retrospect, and what seems essential at one moment might lose its value at a later point in time. The development of the instrument reflects the following objectives: We determined that it should (a) be multidimensional to capture the nine cognitive facets and an affective component, (b) possess good psychometric properties, and (c) contain comprehensive content while also taking brevity into account. For scale construction, we have followed the standards as outlined by Simms (2008). The scale development process involved four phases (see **Table 2**): (a) developing items, employing expert review and cognitive pretesting; (b) evaluating the psychometric properties of the individual items, examining the factor structure, creating initial scales, and establishing reliability; (c) assessing factor structure, the similarity of factors, and the reliability of the final scale in a new sample; and (d) investigating convergent and discriminant validity and testing concurrent and incremental validity. This approach helped us address the content and structure of a fulfilled life, determine whether the proposition of a new construct and measure is justified, and identify how the construct is located in its nomological network.

The Present Study

The main objective of this study involved developing and validating the Fulfilled Life Scale (FLS). In addition to a pilot form encompassing the nine facets, we intended to create an economical version, optimally comprising 20–30 items. Our pursuit of this aim began with an examination of the underlying factors. We expected that all cognitive facets would positively intercorrelate and that we would find a multi-dimensional factor structure beyond the global level. Our plan included establishing validity by investigating the relationship between a fulfilled life and similar constructs and the criteria it predicts, drawing on our theoretical article (Baumann and Ruch, 2021). In the first place, the participants' global rating of how close one's own life comes to a maximally fulfilled life serves as the prime validity criterion. A total score or components should highly correlate with a layperson's view of fulfillment. Next, it is vital to show that fulfillment overlaps with concepts like life satisfaction, subjective well-being, and eudaimonic well-being without being redundant (i.e., containing additional unique variance). Two types of results will underscore the usefulness of fulfillment as a new concept. In one of these, reliable variance in the fulfillment measures should not be fully explained by the existing concepts alone or together (e.g., hedonic and eudaimonic well-being do not fully account for fulfillment). According to the other, fulfillment should exhibit incremental validity when predicting important life outcomes (i.e., should be predictive over and above traditional variables). In the present study, we explore whether affective and cognitive fulfillment can incrementally predict a global rating of a fulfilled life after controlling for established well-being measures, such as subjective and eudaimonic well-being. Furthermore, we suggest that a fulfilled life can predict relevant key variables for aging well, including prospective life satisfaction, mental well-being, and self-perceptions of aging. Establishing these relationships will be essential.

In particular, we expect that the orientations to engagement, meaning, and accomplishment relate to general fulfillment. Nevertheless, we also believe that positive relationships are more relevant in terms of leaving a legacy. We assume that

a favorable psychosocial development should play a significant role in attaining a fulfilled life. Generativity and ego integrity are development tasks in the second half of life (Erikson, 1985), and their successful resolution leads to a mature, well-rounded personality. We expect both concepts to be related to a fulfilled life. As research has suggested that perceiving one's job as a calling is fulfilling (Wrzesniewski et al., 1997; Hall and Chandler, 2005), we expect that persons arriving at a fulfilled life are more likely to report a calling. Furthermore, we intend to demonstrate known-groups validity by showing that selected individuals pursuing a calling (named calling exemplars) differ from a general sample regarding their level of fulfillment. Finally, we will report the results of comparisons with sociodemographic and contextual variables.

MATERIALS AND METHODS

Phase 1: Substantive Validity

For the initial item pool generation, we used a rational-theoretical approach based on the theoretical model of FiL (Baumann and Ruch, 2021). The study of specialist and non-specialist literature about the different facets of a fulfilled life yielded rich sources for developing an item pool to ensure good content validity. Such an approach should ensure that the construct also occurs in everyday life and is wholly and accurately covered. We developed items to assess the nine cognitive facets and the affective component described in the introduction and illustrated in **Table 1**. The first author wrote 101 items to measure the cognitive component and 12 items for assessing the affective component. All statements referred to one's life lived so far in retrospect. The cognitive items were positively worded, while the affective statements also comprised six negatively phrased items to be recoded. Items corresponding to the affective experience consisted of positive low-arousal feelings, such as deep inner contentment, inner peace, and negative feelings, including disappointment, emptiness, or deep regret when looking back on one's life. The number of items was purposely overinclusive to be able to select those with the best psychometric properties and content coverage. We chose a 6-point Likert scale to ensure sufficient response variance and to avoid overwhelming older participants with a too large response format. Eight experts in the field of positive psychology (Ph.D. students, senior researchers, and a professor) who were also proficient with the concept of FiL, rated the content validity (the extent to which the items reflect the content domain) and item quality (comprehensibility, conciseness, and redundancy). We improved the wording of a few items, and selected the best nine items from each of the nine facets of the cognitive component. Construct validity was further improved by using a think-aloud procedure, a cognitive interviewing method, to identify and correct sources of response error in the survey questions. Employing this method in the early stage of test construction can prevent problems in the areas of comprehension, recall, or decision processes, enabling lay people to provide valuable information about the construct. After conducting three face-to-face cognitive interviews with persons from our target group, we modified the wording of four items. The provisional scale for further analysis comprised 81 cognitive and 12 affective items.

Phase 2: Structural Validity I

Participants

Sample 1 (development sample) consisted of N = 282 German-speaking participants aged 50–93 (M = 60.93, SD = 8.64, 79.4% women). Two participants did not indicate their age. Of the sample, 50.7% were married or in a registered partnership, 25.2% were divorced, 15.2% were single and never married, 5.7% were widowed, and 3.2% were separated. Approximately half of the sample (53.2%) had attained a university degree as their highest level of education, while 20.6% held a professional education diploma, 11.7% had a general education (e.g., baccalaureate), 13.5% had a vocational education and training, and 1.1% had a compulsory school qualification (9 years of education).

Sample 2 (replication sample) included N = 406 German-speaking participants aged 40–85 (M = 58.81, SD = 10.66, 78.8% women). One participant did not indicate her age. Of this sample, 50.7% were married or in a registered partnership, 20.0% were single and never married, 20.9% were divorced, 5.7% were widowed, and 2.7% were separated. About half of the sample (49.5%) had attained a university degree as their highest level of education, while 25.1% held a professional education diploma, 12.3% had a general education (e.g., baccalaureate), 11.8% had a vocational education, and 1.2% had a compulsory school qualification (9 years of education).

Sample 3 (calling exemplars) consisted of N = 39 German-speaking participants aged 41–89 (M = 57.92; SD = 9.63; 79.5% men). Of these, 53.8% were married, 28.2% were divorced, 12.8 were single or never married, and 5.1% were widowed. This sample was highly educated: 71.8% held a university degree, 23.1% held a professional education diploma, 2.6% held a baccalaureate, and 2.6% had a compulsory school qualification.

The three samples included only participants who provided complete and valid responses. Participants were excluded when they did not meet the inclusion criteria or showed obvious or irregular response patterns (Sample 1: n = 6; Sample 2: n = 13; Sample 3: n = 4). Other outliers were kept and were in the range of what is expected in a normal distribution.

Procedure

We collected data for all samples in the German language, employing online surveys as part of a larger research project. Participants were recruited through voluntary organizations and by means of online advertising. In a first wave, we collected Sample 1, and in a second wave, Sample 2 together with Sample 3. For the Sample 3, we recruited professionals who had been interviewed and portrayed in books on the topic of profession and calling (see, for example, Morgenthaler, 2010). They received a personal invitation to participate in the study, and we built a separate data collector for them. The inclusion criterion for the Sample 1 was a minimum age of 50, while Samples 2 and 3 involved a minimum age of 40 years. The participants received no renumeration. Upon completion of the surveys, participants could download brochures with suggestions on how

TABLE 3 | Instruments used for this study.

Instrument	Authors	Content	Sample item	Number of items	Response format	Use in sample	α in this study
Fulfilled Life item pool	Baumann and Ruch	Cognitive and affective experience of a fulfilled life	"I have led my life in a way that has deeply suited me."	81 and 12	6-point Likert scale (1 = does not apply at all to 6 = applies completely)	Sample 1	0.87–0.93 (Cognitive facets), 0.96 (Affective Experience)
Fulfilling Life Rating – present[a]	Baumann and Ruch (developed for this study)	Fulfilling life (comprising life at the current life stage)	"Compared to your given example, how fulfilling is your life at the current life stage?"	1	11-point Likert scale (0 = not at all fulfilling to 10 = entirely fulfilling)	Sample 1 and 2	–
Fulfilled Life Rating – retrospect[a]	Baumann and Ruch (developed for this study)	Fulfilled life (comprising the whole lived life)	"Compared to your given example, how fulfilled is your life lived so far in retrospect?"	1	11-point Likert scale (0 = not at all fulfilled to 10 = entirely fulfilled)	Sample 1 and 2	–
Fulfilled Life Rating – if life ended tomorrow	Baumann and Ruch (developed for this study)	Fulfilled life (comprising the whole lived life)	"If my life were to end tomorrow, I could say with full conviction that my life was . . ."	1	6-point Likert scale (1 = not fulfilled at all to 6 = completely fulfilled)	Sample 2	–
Orientations to Happiness questionnaire (OTH)[b]	Peterson et al. (2005; in the German adaptation by Ruch et al., 2010)	Orientations to well-being: pleasure, engagement, and meaning	"I seek out situations that challenge my skills and abilities."	15	5-point Likert scale (1 = very much unlike me to 5 = very much like me)	Sample 1	0.72 (Pleasure), 0.68 (Engagement), 0.79 (Meaning)
The short scales for assessing positive relationships and accomplishment	Gander et al., 2017	PERMA dimensions of positive relationships and accomplishment	"A good life means to me that I can share it with others."	10	5-point Likert scale (1 = very much unlike me to 5 = very much like me	Sample 1	0.74 (Positive Relationships), 0.80 (Accomplishment)
Temporal Satisfaction With Life Scale (TSWLS)	Pavot et al. (1998; in the German adaptation by Trautwein, 2004)	Life satisfaction in the past, present, and future	"My life in the past was ideal for me."	12	7-point Likert scale (1 = strongly disagree to 7 = strongly agree)	Sample 1 and 2	0.88 (Past Life Satisfaction), 0.89–0.92 (Present Life Satisfaction), 0.90–0.92 (Future Life Satisfaction)
Positive and Negative Affect Schedule (PANAS)	Watson et al. (1988; in the German adaptation by Krohne et al., 1996)	Intensity of positive and negative affect	"Active"	20	5-point Likert scale (1 = very slightly or not at all to 5 = extremely)	Sample 1	0.88 (Positive Affect), 0.87 (Negative Affect)
Lie Scale of the short form of the Eysenck Personality Questionnaire-Revised (EPQ-R)	Eysenck and Eysenck (1991; in the German adaptation by Ruch, 1999)	Social desirability	"Have you ever cheated at a game?"	12	"Yes" or "no" questions	Sample 1	0.72
The Questionnaire for Eudaimonic Well-Being (QEWB)	Waterman et al. (2010; in a German version translated following the standard translation process)	Eudaimonic functioning	"I find a lot of the things I do are personally expressive for me."	21	5-point Likert scale (0 = strongly disagree to 4 = strongly agree)	Sample 2	0.83
Loyola Generativity Scale (LGS)	McAdams and de St. Aubin (1992; in the German translation reported by Hofer et al., 2008)	Generative concern	"I try to pass along the knowledge I have gained through my experiences."	20	4-point Likert Scale (0 = not at all to 3 = extremely)	Sample 2	0.86

(Continued)

TABLE 3 | (Continued)

Instrument	Authors	Content	Sample item	Number of items	Response format	Use in sample	α in this study
Ego Integrity Scale (RHEIS)	Ryff and Heincke (1983; in the German translation reported by Busch et al., 2018)	Erikson's conceptualization of psychological maturity in late adulthood	"In general, I would say I have few regrets about my past life."	16	6-point Likert Scale (0 = strongly disagree to 5 = strongly agree)	Sample 2	0.90
Serenity subscale of the Positive and Negative Affect Schedule – Extended (PANAS-X)	Watson and Clark (1994; in the German version by Grühn et al., 2010)	Affective state of serenity	"At ease."	3	5-point Likert Scale (1 = very slightly or not at all to 5 = extremely)	Sample 2	0.88
Warwick-Edinburgh Mental Well-Being Scale (WEMWBS)	Tennant et al. (2007; in the German translation)	Positive mental health during the last 2 weeks	"I've been feeling useful."	14	5-point Likert scale (1 = none of the time to 5 = all of the time)	Sample 2	0.90
Attitude Toward Own Aging subscale of the Philadelphia Geriatric Center Morale Scale	Lawton (1975; in the German translation reported by Beyer et al., 2015)	Self-perceptions of aging	"Things keep getting worse as I get older."	5	4-point Likert Scale (1 = does not apply at all to 4 = fully applies)	Sample 2	0.72
The Presence subscale of the Brief Calling Scale (BCS)	Dik et al. (2012; in the German translation by Hirschi, 2011)	A sense of a calling in one's work	"I have a calling to a particular kind of work."[c]	2	5-point Likert scale (1 = not at all true of me to 5 = totally true of me)	Sample 2	0.89[d]

[a] *Those ratings were anchored in a short description provided by each participant, in which participants presented their most telling example of a fulfilling and a fulfilled life in a few sentences for each, respectively.*
[b] *This scale was used together with the short scales to assess positive relationships and accomplishment (Gander et al., 2017) to assess all PERMA dimensions of Seligman's (2011) well-being theory. For this purpose, each OTH scale was reduced by one item.*
[c] *In order to not restrict a calling to paid employment and in consideration that our sample comprises retirees, we added the term activity to both items.*
[d] *Spearman–Brown coefficient.*

to promote their mental well-being and healthy aging, and create a fulfilling life.

Instruments

Table 3 presents all instruments used for this study. Sociodemographic and relevant contextual questions were asked about age, gender, marital status, being a parent, educational level, employment status, financial status, self-rated health, spirituality in daily life, childhood experience, and volunteering.

DATA ANALYSIS AND RESULTS

Data Analysis

For all statistical analyses, we used IBM SPSS Statistics, version 25.0. Descriptive statistics (mean, standard deviation, minimum, maximum, skewness, kurtosis, corrected item-total correlations) and reliability coefficients (Cronbach's alpha) were calculated. To investigate the structure of the FLS and determine the number of components aside from the general factor of a fulfilled life, we employed principal component analysis. Since we expected the factors to be correlated, we performed oblique rotation. The number of factors was determined through the use of parallel analysis (Horn, 1965) and the minimum average partial test (MAP; Velicer, 1976), using SPSS syntax written by O'Connor (2000), while also by examining all possible factor solutions between one and nine (according to our nine facet-based scales) through hierarchical factor analysis (Goldberg, 2006). This analysis technique reveals the unfolding of the factors by correlating the factor scores of each level (beginning with the first unrotated principal component [FUPC]) with those of the next level. The procedure facilitates examining whether factors are reorganized or remain stable across different levels. In addition to these extraction criteria, we selected a factor solution that accounted for a substantial proportion of variance and – most importantly – that was plausibly interpretable.

Results

Fulfilled Life Cognitive Experience

Preliminary Results

We used Sample 1 to compute total scores for the nine cognitive facets by averaging the assigned items. All nine facets were strongly correlated, which indicated a general factor of a fulfilled life. Cronbach's alpha ranged between 0.87 and 0.93, and the corrected item-total correlations ranged from 0.28 to 0.84. For economic reasons and as eight facets had between one and three items with a lower corrected-item-total correlation as its correlation with other facets, we selected the best six items per facet (54 items in total) for further analyses and

TABLE 4 | Descriptive statistics of fulfilled life facets.

	M	*SD*	*S*	*K*	α	citc	FUPC
(A) Realized uniqueness	4.58	0.73	–0.69	0.57	0.82	0.47–0.65	0.81
(B) A life lived fully	4.49	0.80	–0.83	1.12	0.88	0.63–0.76	0.86
(C) The making of a positive difference	4.75	0.72	–0.67	1.57	0.89	0.63–0.77	0.82
(D) Authentic pursuits	4.56	0.79	–0.85	1.29	0.87	0.61–0.73	0.86
(E) A life true to oneself	4.62	0.82	–0.96	1.34	0.89	0.68–0.75	0.87
(F) A contribution reflecting the self	4.55	0.84	–0.50	0.31	0.92	0.58–0.84	0.74
(G) Worthwhile involvements	4.68	0.75	–0.89	1.83	0.90	0.70–0.74	0.90
(H) A life that was worthwhile	4.78	0.79	–1.33	3.00	0.87	0.43–0.82	0.84
(I) A life that mattered to others	4.64	0.81	–0.76	1.58	0.90	0.54–0.81	0.74

Sample 1 = $N_{Development}$ = 282. Six items per facet.
α, Cronbach's alpha; citc, corrected item total correlation range; FUPC, first unrotated principal component.

TABLE 5 | Zero-order correlations fulfilled life facets.

	1	2	3	4	5	6	7	8	9
1 (A) Realized uniqueness	–								
2 (B) A life lived fully	0.71	–							
3 (C) The making of a positive difference	0.62	0.61	–						
4 (D) Authentic pursuits	0.71	0.79	0.59	–					
5 (E) A life true to oneself	0.67	0.78	0.60	0.83	–				
6 (F) A contribution reflecting the self	0.49	0.50	0.74	0.50	0.55	–			
7 (G) Worthwhile involvements	0.67	0.76	0.64	0.76	0.79	0.60	–		
8 (H) A life that was worthwhile	0.65	0.69	0.60	0.68	0.68	0.50	0.79	–	
9 (I) A life that mattered to others	0.52	0.51	0.74	0.43	0.49	0.67	0.62	0.61	–

Sample 1 = $N_{Development}$ = 282. Six items per facet. All correlations $p < 0.001$.

to form a raw version with nine scales. Descriptive statistics, reliability (Cronbach's alpha) coefficients (**Table 4**), and zero-order correlations were calculated for these nine raw version scales (**Table 5**).

Table 4 reveals that all scales were negatively skewed, suggesting that the respondents favorably rated their lives as fulfilling. All scales had good internal consistency, with alpha coefficients ranging from 0.82 to 0.92.

Table 5 demonstrates that all raw version scales were positively intercorrelated, indicating a g-factor of cognitive fulfillment. The facets of the cells A, B, D, and E were highly intercorrelated, representing the sources self and life in relation to the criteria completeness/wholeness and congruence. The high correlation between cells D and E, which had different sources but the same criterion, might suggest that the criterion had greater importance than the sources. The same applies to the correlation between G and H. In addition, G and H were also strongly related to the cells B, D, and E, sharing the same sources. Higher correlations could also be found for the cells C, F, and I, which represented the facets of the column impact/legacy and were related due to their common source. Thus, the pattern of the 3 × 3 matrix seemed to exist within the data and reflected the characteristics of the model. Nevertheless, further examination was required to find a structure.

We submitted the pool of 54 items to principal component analysis to investigate the underlying factor structure. The Kaiser–Meyer–Olkin measure verified the sampling adequacy for the analysis, KMO = 0.96, which exceeded the minimum criteria of 0.5 (Hutcheson and Sofroniou, 1999), and Bartlett's test of sphericity suggested that the data were suitable for factor analysis, $\chi^2(1431) = 11{,}990.29$, $p < 0.001$. Eight factors exceeded unity (the first ten eigenvalues were 23.69, 4.24, 2.19, 1.62, 1.51, 1.20, 1.11, 1.01, 0.93, 0.85). The parallel analysis yielded three components with eigenvalues exceeding the randomly generated values (95% CI, 1000 random correlation matrices; see O'Connor, 2000), and MAP suggested a seven-factor solution. In summary, the different tests did not suggest the extraction of the same number of factors.

Hierarchical Factor Analysis

In seeking to arrive at the best factor-solution at the item level, we examined all solutions for between one and nine factors by employing hierarchical factor analysis (Goldberg, 2006). Using Sample 1, we performed principal component analyses with varimax rotation. **Figure 1** illustrates the resulting hierarchical structure.

In the overview, the factors are designated by their size (denoted by the number to the right of the factor; e.g., 3/1 is the largest and 3/3 the smallest factor at that level), which is the amount of explained variance by that factor. The figure further shows the correlations between the factor scores and those at adjacent levels. In addition, we correlated the factors at each level with the nine facets (average scores) to examine the correspondence between the facets and the derived factors.

All items loaded on the first unrotated principal component, which represented the general factor of a Fulfilled Life. At the second level, the first unrotated principal component was divided into one factor (2/1), labeled Unfolded Self and Life, and a smaller second factor (2/2), deemed Positive Impact and Legacy. The first factor was strongly loaded by items of the six facets that referred to the columns Self and Life of the model, and the second factor was mainly loaded by the three Impact/Legacy cells and partly by G and H. At the third level, the items of the facets G and H that had previously loaded on both factors now formed a new factor: The Worthwhile Life. This three-factor solution offered a rearrangement of the nine facets. The factor Unfolded Self and Life (3/1) covered mainly the facet of Self in terms of the criterion congruence and also the two facets Life regarding the criteria of completeness and congruence. The factor Positive Impact and Legacy (3/2) consisted of high loadings from items of all three facets concerning Impact/Legacy. Whereas, The Worthwhile Life

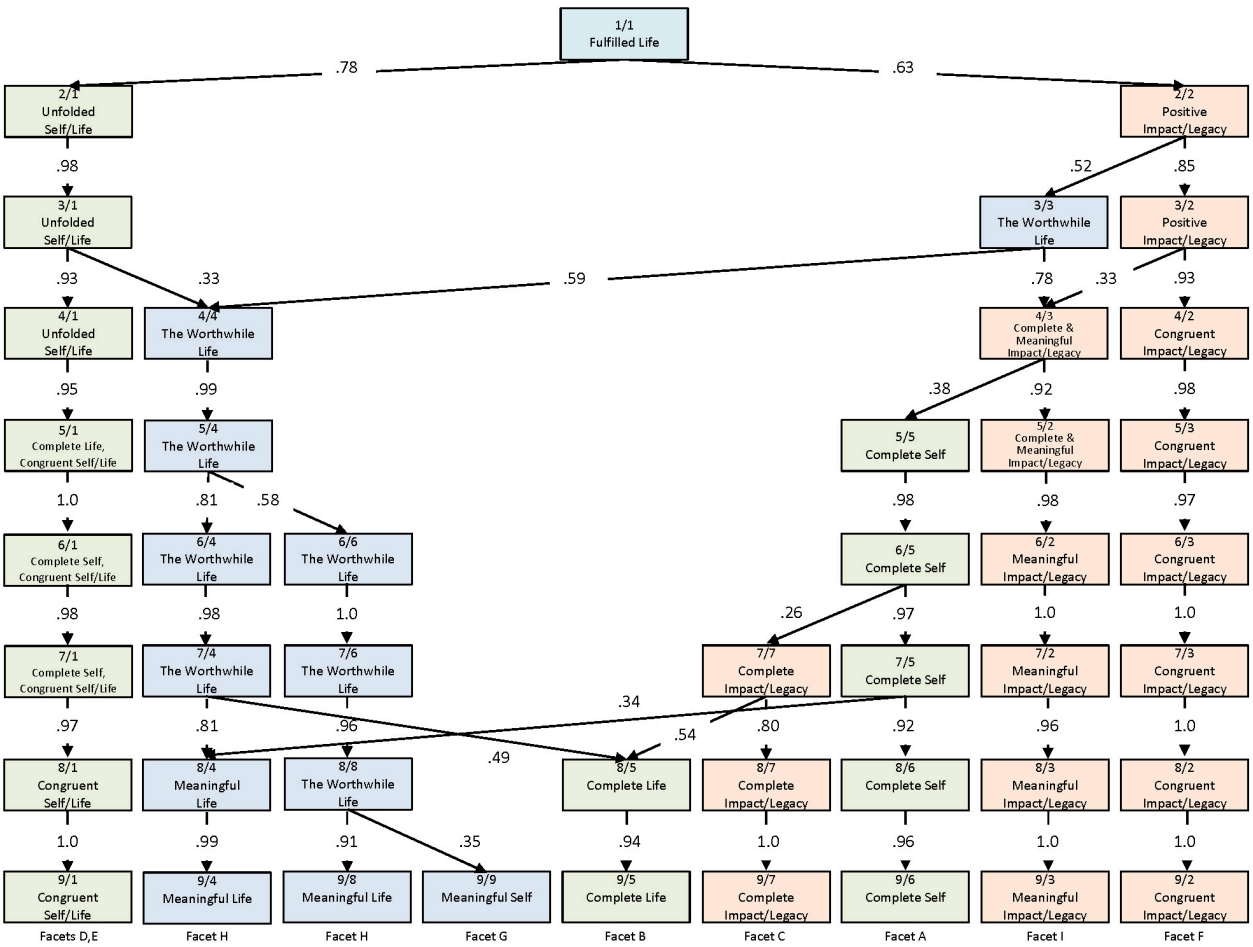

FIGURE 1 | Hierarchical factor analysis of the fulfilled life item set of the cognitive component. Hierarchical representation of the emergence of the first nine varimax-rotated principal components derived from the 54 cognitive items of the FLS. Correlations between the factors of adjacent levels are represented using arrows (only correlations > | 0.30| are shown). The color of the squares indicates the assignment of the factors from level 3 onward.

(3/3) contained high loadings from items of the facets in terms of the criterion meaningfulness. The evolution of the factors revealed, especially with respect to the sources Self and Life, a separation between the two criteria completeness and congruence and the criterion of meaningfulness.

In the next six steps, the factors divided further. It should be mentioned before continuing the discussion that at level nine, only six of the nine factors represented a facet from the model; one factor combined two facets, and two factors represented the same single facet. Thus, the evident redundancy among the facets suggested that a lower number might be more suitable. More specifically, the factor Unfolded Self and Life divided into two factors after level four, building a new factor (5/5) together from the division of the factor (4/3). At the eighth level, three factors emerged, and these remained the same also at the ninth level. The three factors (9/1, 9/5, 9/6) had high item loadings from the facets Self and Life relating to the criteria completeness and congruence. However, those factors did not correspond completely to the facets since two facets were combined in one factor (9/1). Nevertheless, these results support the notion of sources and criteria in the model, as for the factor Congruent Self/Life (9/1), the criteria seem to have a greater effect on the factor than the sources.

Positive Impact and Legacy divided into two factors at the fourth level, into three at the seventh level, and remained unchanged all the way through the ninth level. These three resulting factors represent the columns impact and legacy exactly, comprising the three facets representing the source and one of the three criteria.

At the fourth level, The Worthwhile Life was mainly represented by items from the Meaningful Self and Life facets. However, it split up into two factors at the sixth level and into three at the ninth level, whereby two factors represented the Meaningful Life and one factor the Meaningful Self, as already mentioned.

In summary, the pattern of the nine facets became evident and presented a justification of the 3 × 3 matrix, but not all of them could be separated. Though a solution with a higher number of factors could account for more variance, a three-factor solution appeared to be the most interpretable and relevant for research

TABLE 6 | Pattern matrix of the three-factor principal component analyses with oblimin rotation.

Items	Items in English	Sample 1				Sample 2			
		F1	F2	F3	h^2	F1	F2	F3	h^2
Factor 1: Unfolded Self and Life									
(2) Ich konnte meine Einzigartigkeit zeigen. (A)	I could show my uniqueness.	**0.60**	0.22	0.06	0.46	**0.74**	0.25	0.21	0.61
(5) Ich konnte mein wahres Können im Leben zeigen. (A)	I was able to show my true ability in life.	**0.61**	0.24	0.06	0.58	**0.77**	0.11	0.04	0.65
(8) Ich habe meine Chancen im Leben genutzt. (B)	I took advantage of my opportunities in life.	**0.46**	0.13	0.30	0.51	**0.67**	0.01	0.14	0.59
(12) Ich konnte eigene Träume verwirklichen. (B)	I could realize my own dreams.	**0.85**	0.02	0.03	0.71	**0.69**	0.10	0.25	0.66
(20) Ich habe den Mut gehabt, so zu sein, wie ich wirklich bin. (D)	I have had the courage to be as I really am.	**0.65**	0.11	0.19	0.51	**0.76**	0.01	0.04	0.60
(21) Ich habe meinen Leidenschaften nachgehen können. (D)	I have been able to pursue my passions.	**0.81**	0.02	0.02	0.66	**0.83**	0.06	0.04	0.69
(26) Ich konnte im Leben das tun, wofür ich am besten geeignet war. (E)	I could do in life that which I was best suited for.	**0.66**	0.14	0.16	0.66	**0.82**	0.06	0.00	0.72
(29) Ich habe mein Leben so geführt, wie es mir zutiefst entsprochen hat. (E)	I have led my life in a way that has deeply suited me.	**0.65**	0.01	0.32	0.70	**0.74**	0.10	0.21	0.70
Factor 2: Positive Impact/Legacy									
(13) Ich habe Möglichkeiten genutzt, um zum Wohlergehen anderer beizutragen. (C)	I have used opportunities to contribute to others' well-being.	0.05	**0.82**	0.09	0.64	0.02	**0.82**	0.01	0.69
(16) Ich konnte mit meinem Leben eine positive Spur bei Menschen in meinem Umfeld hinterlassen. (C)	I was able to leave a positive mark with my life on people in my environment.	0.11	**0.63**	0.17	0.62	0.17	**0.62**	0.11	0.60
(18) Ich konnte einen positiven Beitrag zum Wohle anderer Menschen leisten. (C)	I could make a positive contribution to other people's welfare.	0.09	**0.76**	0.01	0.65	0.08	**0.79**	0.04	0.72
(34) Es war mir ein Anliegen, etwas zum Gelingen unserer Gesellschaft beizutragen. (F)	It was important to me to contribute something to the success of our society.	0.11	**0.74**	0.09	0.55	0.00	**0.84**	0.15	0.63
(36) Ich habe meine Fähigkeiten genutzt, um einen Beitrag für das Allgemeinwohl zu leisten. (F)	I have used my abilities to make a contribution to the common good.	0.20	**0.80**	0.12	0.70	0.11	**0.84**	0.06	0.75
(51) Ich habe andere Menschen in ihrer Entwicklung massgeblich unterstützt. (I)	I have significantly supported other people in their development.	0.17	**0.63**	0.35	0.63	0.07	**0.80**	0.11	0.67
(52) Ich konnte zum Gelingen des Lebens anderer Menschen beitragen. (I)	I could contribute to the success of other people's lives.	0.27	**0.64**	0.37	0.64	0.05	**0.81**	0.12	0.71
(54) Ich habe für einen Zweck gelebt, der über mein Leben hinausgeht. (I)	I have lived for a purpose that goes beyond my life.	0.00	**0.51**	0.27	0.47	0.05	**0.60**	0.26	0.53
Factor 3: The Worthwhile Life									
(39) Die Anstrengungen im Leben haben sich gelohnt. (G)	The efforts in life have been worthwhile.	0.13	0.08	**0.67**	0.61	0.09	0.04	**0.75**	0.67
(40) Ich habe die Gewissheit, dass ich für die richtigen Dinge gelebt habe. (G)	I have the certainty that I have lived for the right things.	0.23	0.08	**0.64**	0.66	0.23	0.16	**0.58**	0.68
(43) Ich habe etwas Wertvolles mit meinem Leben gemacht. (H)	I have done something valuable with my life.	0.12	0.29	**0.53**	0.61	0.17	0.33	**0.54**	0.73
(44) Ich kann auf ein gut gelebtes Leben zurückblicken. (H)	I can look back on a life well lived.	0.41	0.12	**0.62**	0.68	0.28	0.04	**0.69**	0.74
(45) Auch für die schwierigen Zeiten im Leben habe ich Bedeutung und Sinn erkennen können. (H)	Even in the difficult times in life, I have been able to recognize meaning and purpose.	0.10	0.09	**0.74**	0.55	0.10	0.01	**0.76**	0.51
(46) Mein Leben hat sich gelohnt. (H)	My life has been worthwhile.	0.10	0.01	**0.83**	0.79	0.14	0.01	**0.81**	0.80
(47) Ich habe mein Leben als sinnvoll erfahren. (H)	I have experienced my life as meaningful.	0.18	0.00	**0.81**	0.82	0.20	0.05	**0.73**	0.78
(48) Ich habe erkannt, worauf es im Leben wirklich ankommt. (H)	I have realized what really matters in life.	0.01	0.00	**0.55**	0.30	0.04	0.20	**0.60**	0.47

Sample 1 = $N_{Development}$ = 282. Sample 2 = $N_{Replication}$ = 406.
*Bold loadings indicate the factor on which the item was retained. The letter in brackets indicates the original facet of the model (Find the meaning of the letters in **Table 1**). English items were translated from the original German items employing a translation/back translation procedure. The translation has not been validated.*

and application. Selecting a four-factor solution in which Positive Impact and Legacy was divided into two factors might not have led to additional value, as the one factor already represented a well-rounded concept. The advantage of having conducted a hierarchical factor analysis compared to solely performing EFAs lay in the ability to separate the two modes of the model (sources and criteria) and gain a more comprehensive understanding of the structural representation of the nine cognitive facets.

Exploratory Factor Analysis

Based on the decision to retain three factors, we performed a principal component analysis with oblique (oblimin direct with delta = 0) rotation on Sample 1. The three factors explained 55.78% of the variance. For economic reasons and to create provisional scales with an equal number of items, we selected the best eight items per factor by considering factor loadings, construct representation, and avoidance of content overlap among the selected items. We subjected the reduced item set (24 items) to new principal component analysis. Three eigenvalues exceeded unity (the first five eigenvalues were 10.79, 2.32, 1.57, 0.99, 0.85). Again, oblique (oblimin direct with delta = 0) rotation was performed. The three factors explained 61.19% of the variance, and factor loadings ranged from 0.46 to 0.85. **Table 6** exhibits the new pattern matrix and the extracted communalities. Unfolded Self and Life correlated with Positive Impact/Legacy at $r = 0.35$ and with The Worthwhile Life at $r = 0.45$. Positive Impact/Legacy correlated with the Worthwhile Life at $r = 0.49$.

Fulfilled Life Affective Experience

Exploratory Factor Analysis

In this process, we first submitted the 12 items for affective fulfillment to principal component analysis using Sample 1. The negatively worded items were recoded beforehand. To verify the sampling adequacy for the analysis, we used the Kaiser–Meyer–Olkin measure, KMO = 0.95, which was highly acceptable. Bartlett's test suggested that the data were suitable for factor analysis, $\chi^2(66) = 3{,}278.31$, $p < 0.001$. Two eigenvalues exceeded unity; the first four eigenvalues were 8.27, 1.04, 0.51, 0.47. Parallel analysis indicated a one-factor solution and the MAP test a two-factor solution. For economic reasons and due to the high internal consistency of the Fulfilled Life Affective Experience scale with Cronbach's alpha of 0.96, we selected the best eight items (five positively worded items and three recoded items) to create a one-dimensional scale. Due to its highly positively skewed distribution, the item "I feel empty" was slightly changed to "I feel rather empty." To evaluate the suggested one-factor structure of the eight items, we conducted a principal component analysis. The PCA revealed a one-factor structure (the first three eigenvalues were 5.46, 0.80, 0.49). The first factor explained 68.30% of the variance. Parallel analysis and MAP equally suggested a one-factor solution. The factor loadings ranged from 0.67 to 0.90 (see **Table 7**).

Descriptive Statistics

Total scores for all scales were computed by averaging the assigned items. Descriptive statistics, internal consistencies of the FLS scales, and their correlations with age and gender in Sample 1 are presented in **Table 8**. The scales were negatively skewed, suggesting that the respondents appraised their lives, in general, and the worthwhileness of their lives, in particular, as fulfilling. Skewness ranged from –0.63 to –1.38, and kurtosis ranged from 0.90 to 3.37. The internal consistencies of the scales were high ($\alpha = 0.89$–0.94). Corrected item-total correlations (not included in **Table 8**) varied from 0.44 to 0.85 for the cognitive items and from 0.59 to 0.85 for the affective items. Results revealed only for affective fulfillment and age a small positive correlation ($r = 0.15$, $p < 0.05$). No significant association emerged between the scales and gender. Meanwhile, the FLSs were strongly intercorrelated (see **Table 9**). The Fulfilled Life Scale (FLS) and scoring information are available in the **Supplementary Material** of this article.

TABLE 7 | Factor loadings of fulfilled life affective items.

Items	Items in English	Sample 1	Sample 2
		Factor loadings	Factor loadings
(1) verspüre ich eine tiefe innere Zufriedenheit.	I feel deep inner contentment.	0.87	0.85
(2) fühle ich mich im Einklang mit mir und dem gelebten Leben.	I feel in harmony with myself and the lived life.	0.89	0.87
(3) habe ich einen inneren Frieden.	I have inner peace.	0.88	0.87
(5) empfinde ich grosse Dankbarkeit.	I feel great gratitude.	0.71	0.81
(6) fühle ich mich erfüllt.	I feel fulfilled.	0.90	0.89
(7) empfinde ich tiefe Reue. (R)	I feel deep regret. (R)	0.67	0.55
(11) fühle ich mich enttäuscht. (R)	I feel disappointed. (R)	0.82	0.81
(12) fühle ich mich eher leer. (R)	I feel rather empty. (R)	0.84	0.82

Sample 1 = $N_{Development}$ = 282. Sample 2 = $N_{Replication}$ = 406. Reverse-scored items are denoted with (R).

Phase 3: Structural Validity II

Data Analysis

All statistical analyses used IBM SPSS Statistics (Version 25), except for confirmatory factor analyses, which were conducted with the lavaan package for R. The analyses were based on Sample 2. The following goodness-of-fit indices were applied to evaluate the CFA model: values ≥ 0.90 in the comparative fit index (CFI) and Tucker–Lewis index (TLI; Hu and Bentler, 1999), value ≤ 0.08 in the root mean square error of approximation (RMSEA; Browne and Cudeck, 1992), and value ≤ 0.08 in the standardized root mean residual (SRMR; Hu and Bentler, 1999).

Results

Fulfilled Life Cognitive Experience

The cognitive items of the final version of the FLS were first subjected to principal component analysis. The KMO index was 0.96, and the Bartlett's test of sphericity suggested that the data were adequate for factor analysis, $\chi^2(276) = 7{,}147.21$, $p < 0.001$. Three factors exceeded unity (the first five eigenvalues were 11.60, 2.76, 1.51, 0.77, 0.67). We extracted three factors and performed an oblique (oblimin direct with delta = 0) rotation. The three factors explained 66.13% of the variance. The intercorrelations of the factors were: $r = 0.44$ for Unfolding of the Self and Life and Positive Impact and Legacy, $r = 0.57$

for Unfolding the Self and Life and The Worthwhile Life, and $r = 0.42$ for Positive Impact and Legacy and The Worthwhile Life. The new pattern matrix, displayed in **Table 6**, shows that all items had their highest loading on the intended factor. The factor loadings ranged from 0.54 to 0.84, with cross-loading differences > 0.20. Tucker's Phi coefficients indicated that the extracted factors were similar across the two samples: Unfolded Self/Life: $\varphi = 0.97$, Positive Impact/Legacy: $\varphi = 0.98$, The Worthwhile Life: $\varphi = 0.96$. Lastly, the three-factor model was evaluated by confirmatory factor analysis with robust estimator: model fit: $\chi^2(249) = 636.92$, $p < 0.001$; all other fit indices indicated an acceptable fit to the data (CFI = 0.931, TLI = 0.924, RMSEA = 0.069, SRMR = 0.060). The robustness of the three-factor structure determined from the EFA of Sample 1 was supported. Information on item statistics is provided in **Supplementary Table A1**.

Fulfilled Life Affective Experience

We first subjected the items of the final version to principal component analysis. The KMO index was 0.94, and Bartlett's test of sphericity indicated that the data were adequate for factor analysis, $\chi^2(28) = 2{,}319.21$, $p < 0.001$. One factor exceeded unity (the first three eigenvalues were 5.31, 0.83, 0.49). The one-factor explained 66.36% of the variance. The factor loadings ranged from 0.55 to 0.89 (see **Table 7**). Tucker's Phi coefficient indicated that the factor was similar across the two samples ($\varphi = 1.00$). To determine the model quality, we conducted confirmatory factor analysis with a robust estimator. Model fit was $\chi^2(20) = 73.46$, $p < 0.001$; all other fit indices indicated an adequate fit to the data (CFI = 0.969, TLI = 0.957, RMSEA = 0.093, SRMR = 0.040). Information on item statistics is presented in the **Supplementary Table A**.

Descriptive Statistics

Table 8 presents descriptive statistics, internal consistencies of the FLS scales, and their correlations with age and gender. The table shows that the means were above the scale's midpoint and that the standard deviations were higher in the replications sample. Skewness ranged from –0.64 to –1.15, and kurtosis ranged from 0.38 to 1.66. The internal consistencies of the scales were high ($\alpha = 0.91$–0.95). Small positive correlations between all scales (except Unfolded Self and Life) and age were found. Women reported slightly higher affective fulfillment than men. Overall, the scales were strongly intercorrelated (see **Table 9**).

Phase 4: External Validity

For the evaluation of construct and criterion validitiy, analyses were conducted on Sample 1 and 2 separately. We partly used different instruments in the samples. Results of convergent, discriminant, and concurrent validity can be found in **Table 10**, while results for incremental validity are displayed in **Table 11**.

Convergent and Discriminant Validity

The three global fulfilled life ratings were positively associated with all FLS dimensions. Numerically higher correlations were found for the two ratings from a retrospective view compared to the one assessing the present. While affective fulfillment yielded the strongest relationships with all three ratings, Positive Impact and Legacy had the lowest correlation.

Retrospective life satisfaction positively correlated with all FLS dimensions and yielded the lowest correlation for the Positive Impact and Legacy. The correlation pattern of the PERMA dimensions and the FLSs showed the strongest relationships to engagement, meaning, and accomplishment. Eudaimonic functioning was positively related to all fulfilled life dimensions. Generativity yielded the largest correlation for Positive Impact and Legacy, while ego integrity showed the largest correlation with The Worthwhile Life and affective fulfillment. Serenity was positively related to all FLS dimensions, with the highest correlation for affective fulfillment, as expected. Large positive correlations between positive affect and the

TABLE 8 | Descriptive statistics, internal consistencies of the fulfilled life scales and correlations of the fulfilled life scales with age and gender.

Scale	Descriptive statistics and internal consistencies								Correlations	
	N items	*Min*	*Max*	*M*	*SD*	*S*	*K*	α	Age	Gender
Sample 1										
USL	8	1.38	6.00	4.34	0.80	–0.74	0.90	0.90	0.06	−0.11
PIL	8	1.38	6.00	4.64	0.73	–0.63	1.01	0.89	0.10	0.01
TWL	8	1.00	6.00	4.76	0.78	–1.38	3.37	0.90	0.11	0.04
FLCE	24	1.75	6.00	4.58	0.67	–0.87	1.38	0.94	0.10	−0.03
FLAE	8	1.38	6.00	4.77	0.91	–1.25	1.58	0.93	0.15*	0.01
Sample 2										
USL	8	1.25	6.00	4.28	0.90	–0.64	0.38	0.92	0.10	0.01
PIL	8	1.00	6.00	4.59	0.85	–0.87	1.28	0.91	0.15**	0.09
TWL	8	1.13	6.00	4.71	0.84	–1.08	1.66	0.92	0.16**	0.06
FLCE	24	1.75	5.92	4.53	0.74	–0.75	0.75	0.95	0.15**	0.06
FLAE	8	1.00	6.00	4.73	0.94	–1.15	1.16	0.92	0.20***	0.11*

Sample 1 = $N_{Development}$ = 280–282. Sample 2 = $N_{Replication}$ = 405–406. Scale range: 1–6.
S, Skewness; K, Kurtosis; α, Cronbach's alpha. Male = 1, female = 2. FLCE, Fulfilled Life Cognitive Experience; USL, Unfolded Self and Life; PIL, Positive Impact and Legacy; TWL, The Worthwhile Life; FLAE, Fulfilled Life Affective Experience.
$^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

TABLE 9 | Pearson correlations of the fulfilled life scales.

	1	2	3	4	5
Sample 1					
(1) USL	–				
(2) PIL	0.54	–			
(3) TWL	0.70	0.65	–		
(4) FLCE	0.87	0.83	0.90	–	
(5) FLAE	0.69	0.48	0.80	0.76	–
Sample 2					
(1) USL	–				
(2) PIL	0.52	–			
(3) TWL	0.72	0.60	–		
(4) FLCE	0.87	0.82	0.90	–	
(5) FLAE	0.71	0.46	0.85	0.78	–

Sample 1 = $N_{Development}$ = 282. Sample 2 = $N_{Replication}$ = 406.
USL, Unfolded Self and Life; PIL, Positive Impact and Legacy; TWL, The Worthwhile Life; FLCE, Fulfilled Life Cognitive Experience; FLAE, Fulfilled Life Affective Experience.
All correlations $p < 0.001$.

fulfilled life dimensions were found. The FLS subscales had no correlations with social desirability and small to moderate negative correlations with negative affect.

Concurrent and Incremental Validity

All fulfillment dimensions were related to current and prospective life satisfaction, mental well-being, self-perceptions of aging, and calling. In most cases, we found the highest correlations for The Worthwhile Life and affective fulfillment.

The first hierarchical regression analysis based on Sample 1 revealed that, after controlling for age and gender, cognitive and affective fulfillment could significantly predict variance in a global fulfilled life rating above and beyond subjective well-being and PERMA. When trying to predict the global fulfilled life rating – in retrospect, we entered age and gender in Step 1, which explained 4% of the variance. Next, we added life satisfaction and positive and negative affect (i.e., the variables defining subjective well-being) in Step 2, which increased the prediction by 35%. In Step 3, the PERMA dimensions caused a further increment of 2%. Entering cognitive and affective fulfillment in Step 4 yielded an additional 18% of variance. Cognitive and affective fulfillment turned out to be the strongest predictors ($\beta = 0.37$, $p < 0.001$, for both). This outcome means that both subjective and eudemonic (PERMA) well-being do not fully account for global subjective fulfillment, but measured fulfillment is also needed.

To examine whether measured fulfillment might have incremental validity over and above established predictors when predicting important life outcomes, we performed an additional hierarchical regression analysis on Sample 2. The results demonstrated that affective and cognitive fulfillment incrementally predicted mental well-being beyond the temporal life satisfaction scales and eudaimonic well-being and after controlling for age and gender. In Step 1, age and gender were entered first, explaining 2% of the variance; in Step 2, the temporal life satisfaction scales increased the prediction by 44%. In Step 3, the eudaimonic well-being caused a further increment by 7%, and in Step 4, cognitive and affective fulfillment contributed an additional 3% of the variance.

Known-Groups Validity

We performed *t*-tests to compare the general sample (Sample 2) with a sample of calling exemplars (Sample 3) regarding the level of the presence of a calling and the experience of a fulfilled life. As expected, the exemplars were significantly more likely to report the presence of a calling ($M = 8.46$, $SD = 1.62$) than persons from the general sample ($M = 7.42$, $SD = 2.18$). This difference, −1.04, BCa 95% CI [−1.527, −0.378], was significant $t(52.244) = 3.71$, $p = 0.001$ (the *t*-test is reported with equal variance not assumed) and represented a medium-sized effect, Hedges' $g = 0.49$. The groups also differed regarding the level of a fulfilled life on all dimensions (see **Table 12**), with the calling exemplar reporting significantly greater fulfillment than the general sample. The Unfolded Self and Life yielded the greatest difference, representing a large effect.

Relationships With Sociodemographic and Contextual Characteristics

We examined the relationships between the fulfilled life dimensions and the global rating, along with various sociodemographic and contextual characteristics in the merged sample (see **Table 13**). Generally, the partial correlations (controlled for age and gender) were meaningful, in the right direction, and small in size. **Table 13** displays age effects for the global rating and all fulfilled life dimensions. No significant relationship was found for gender. Better educated individuals reported higher levels in the global rating and all dimensions with the numerically highest coefficient for the Unfolded Self and Life and the lowest coefficient for The Worthwhile Life. In contrast, whether individuals were employed or retired was not associated with fulfillment. Persons in a financially better position reported higher levels of fulfillment in all dimensions, except for Positive Impact/Legacy. Volunteers, in comparison to non-volunteers, had higher scores on the dimensions of Positive Impact and Legacy and cognitive fulfillment. Being married went along with higher levels of fulfillment in the global rating and all dimensions, except in the Unfolded Self and Life and Positive Impact and Legacy. Furthermore, we found significant correlations for parenthood and the global rating, Positive Impact and Legacy, The Worthwhile Life, and cognitive fulfillment. Conversely, the number of children was solely associated with Positive Impact and Legacy. There was a slight positive correlation between being spiritual in daily life and fulfillment in all dimensions, except for the global rating and Unfolded Self and Life. Individuals with better self-evaluated health reported higher levels of fulfillment in the global rating and all dimensions. Finally, a good childhood experience was positively associated with all fulfillment dimensions, except Positive Impact/Legacy.

A Closer Look at Age

We expected higher fulfillment scores with increasing age and a potential leveling off at the highest age group. Because they consecutively compare an age group with the average of the

TABLE 10 | Descriptive statistics, intercorrelations, convergent, discriminant, and concurrent validity of the FLS subscales.

Measures	Sample 1							Sample 2						
	M	*SD*	USL	PIL	TWL	FLCE	FLAE	*M*	*SD*	USL	PIL	TWL	FLCE	FLAE
Convergent validity														
Fulfilling life rating – present	7.41	1.94	0.52***	0.39***	0.61***	0.59***	0.71***	–	–	–	–	–	–	–
Fulfilled life rating – retrospect	7.59	1.77	0.66***	0.46***	0.64***	0.68***	0.70***	7.64	1.78	0.64***	0.36***	0.70***	0.66***	0.74***
Fulfilled life rating – if life ended tomorrow	–	–	–	–	–	–	–	4.77	0.91	0.64***	0.38***	0.76***	0.69***	0.80***
Life satisfaction past	4.68	1.27	0.53***	0.24***	0.47***	0.48***	0.49***	4.55	1.36	0.52***	0.12*	0.39***	0.40***	0.46***
PERMA														
Pleasure	3.06	0.70	0.31***	0.20 ∗ *	0.28***	0.31***	0.22***	–	–	–	–	–	–	–
Engagement	3.24	0.68	0.43***	0.39***	0.42***	0.48***	0.34***	–	–	–	–	–	–	–
Positive relationships	3.37	0.72	0.17**	0.28***	0.18**	0.24***	0.09	–	–	–	–	–	–	–
Meaning	3.10	0.83	0.33***	0.58***	0.51***	0.54***	0.36***	–	–	–	–	–	–	–
Accomplishment	3.39	0.74	0.42***	0.38***	0.41***	0.47***	0.26***	–	–	–	–	–	–	–
EWB	–	–	–	–	–	–	–	62.08	9.07	0.58***	0.51***	0.68***	0.69***	0.62***
Generativity	–	–	–	–	–	–	–	36.32	8.79	0.51***	0.76***	0.56***	0.70***	0.45***
Ego integrity	–	–	–	–	–	–	–	56.68	12.90	0.67***	0.36***	0.71***	0.67***	0.77***
Positive affect	34.86	6.18	0.54***	0.50***	0.57***	0.62***	0.57***	–	–	–	–	–	–	–
Serenity	–	–	–	–	–	–	–	10.30	2.38	0.47***	0.26***	0.55***	0.50***	0.59***
Discriminant validity														
Social desirability	2.26	2.07	0.06	0.01	0.02	0.03	0.02	–	–	–	–	–	–	–
NA	15.95	5.59	−0.30***	−0.14∗	−0.31***	−0.28***	−0.48***	–	–	–	–	–	–	–
Concurrent validity														
Life satisfaction present	5.50	1.16	0.51***	0.36***	0.55***	0.55***	0.65***	5.37	1.27	0.52***	0.28***	0.66***	0.57***	0.68***
Life satisfaction future	5.38	1.00	0.28***	0.28***	0.37***	0.36***	0.34***	5.46	1.11	0.36***	0.29***	0.58***	0.47***	0.56***
Mental well-being	–	–	–	–	–	–	–	54.69	7.18	0.58***	0.39***	0.67***	0.63***	0.67***
Self-perceptions of aging	–	–	–	–	–	–	–	15.47	2.75	0.37***	0.30***	0.52***	0.46***	0.54***
Calling	–	–	–	–	–	–	–	7.42	2.18	0.45***	0.46***	0.45***	0.53***	0.39***

Sample 1 = $N_{Development}$ = 282, Sample 2 = $N_{Replication}$ = 406
USL, Unfolded Self and Life; PIL, Positive Impact and Legacy; TWL, The Worthwhile Life; FLCE, Fulfilled Life Cognitive Experience; FLAE, Fulfilled Life Affective Experience; EWB, Eudaimonic Well-being; NA, Negative Affect.
**p < 0.05; **p < 0.01; ***p < 0.001.*

consecutive groups, Helmert contrasts are well suited to detect such a satiation point. Thus, the merged sample was divided into four groups (covering a decade each), and ANOVAS with subsequent Helmert contrasts were computed for the different fulfillment measures. There was a significant age-group effect on the Unfolded Self and Life, $F(3,681) = 3.03$, $p = 0.029$, $\eta^2 = 0.01$, Helmert contrasts significant at level 1 ($p = 0.014$); Positive Impact Legacy, $F(3,681) = 6.75$, $p < 0.001$, $\eta^2 = 0.03$, Helmert contrasts significant at level 1 ($p = 0.002$) and level 2 ($p = 0.002$); The Worthwhile Life, $F(3,681) = 5.15$, $p = 0.002$, $\eta^2 = 0.02$, Helmert contrasts significant at level 1 ($p = 0.002$) and level 2 ($p = 0.008$); cognitive fulfillment, $F(3,681) = 6.23$, $p < 0.001$, $\eta^2 = 0.03$, Helmert contrasts significant at level 1 ($p = 0.001$) and level 2 ($p = 0.006$); and affective fulfillment, $F(3,681) = 9.43$, $p < 0.001$, $\eta^2 = 0.04$, Helmert contrasts significant at level 1 ($p < 0.001$) and level 2 ($p = 0.002$). Hence, in this cross-sectional study, fulfillment increased gradually with age until the age group of 60–69 years, after which there was no further increase or decrease (i.e., the level of fulfillment stayed the same). The inspection of scatterplots revealed that scores of younger adults ranged between low and high levels, whereas with age, there was a tendency toward higher minimal scores.

DISCUSSION

The main objective of this study was to develop a reliable and valid measure of a fulfilled life based on our recently proposed theoretical conceptualization (Baumann and Ruch, 2021). The FLS provides an instrument that allows researchers to assess and study the phenomenon of a fulfilled life at different levels. First, it permits to investigate fulfillment at the level of the facets. The assumption that they can be distinguished from each other was supported, as the alpha coefficients were high (even when utilizing only six items); moreover the level of intercorrelation shows that each is characterized by reliable, unique variance. However, the expected pattern of intercorrelations could only be partially substantiated. The three sources (Self, Life, Legacy) and the three criteria (wholeness, fit, value) did not contribute equally and additively to the variance, which eventually led to a revision of the model. The major discovery is that

TABLE 11 | Hierarchical regression predicting the fulfilled life rating – retrospect and mental well-being.

Variable	Predictor	Step 1				Step 2				Step 3				Step 4			
		B	**95% CI**	***SE B***	**β**	***B***	**95% CI**	***SE B***	**β**	***B***	**95% CI**	***SE B***	**β**	***B***	**95% CI**	***SE B***	**β**
Fulfilled life rating – retrospect	Constant	6.01	[4.52, 7.47]	0.75		2.10	[0.35, 4.18]	0.92		1.04	[–1.18, 3.53]	1.12		–1.92	[–3.87,0.18]	0.91	
	Age	0.04	[0.01,0.06]	0.01	0.18**	0.02	[0.00,0.03]	0.01	0.08	0.02	[0.01,0.04]	0.01	0.12*	0.02	[0.00,0.03]	0.01	0.09*
	Gender[a]	–0.36	[–0.81,0.12]	0.24	–0.08	–0.31	[–0.63,0.02]	0.17	–0.07	–0.29	[–0.61,0.04]	0.18	–0.07	–0.30	[–0.62,0.04]	0.16	–0.07
	Life satisfaction present					0.44	[0.17,0.69]	0.12	0.29***	0.43	[0.20,0.65]	0.11	0.28***	0.08	[–0.13,0.31]	0.12	0.05
	Positive affect					0.09	[0.06,0.13]	0.02	0.33***	0.09	[0.05,0.13]	0.02	0.33***	0.04	[0.01,0.08]	0.02	0.16*
	Negative affect					–0.04	[–0.08,–0.01]	0.02	–0.12*	–0.05	[–0.08,–0.02]	0.02	–0.15**	0.00	[–0.03,0.03]	0.02	0.01
	Pleasure									–0.01	[–0.30,0.29]	0.14	–0.01	0.00	[–0.23,0.24]	0.12	0.00
	Engagement									–0.32	[–0.67,0.03]	0.18	–0.12	–0.30	[–0.62,0.00]	0.16	–0.12*
	Positive relationships									0.32	[0.04,0.60]	0.15	0.13*	0.26	[0.03,0.49]	0.13	0.11*
	Meaning									0.00	[–0.28,0.28]	0.13	0.00	–0.32	[–0.54, –0.10]	0.12	–0.15**
	Accomplishment									0.22	[–0.13,0.56]	0.19	0.09	0.04	[–0.25,0.29]	0.15	0.02
	FLCE													0.98	[0.38,1.54]	0.28	0.37***
	FLAE													0.72	[0.34,1.12]	0.23	0.37***
	R^2				0.04				0.39				0.41				0.59
	ΔR								0.35***				0.02				0.18***
Mental well-being	Constant	49.60	[44.62, 54.41]	2.51		26.18	[20.73,31.49]	2.69		19.29	[13.93,24.32]	2.64		20.55	[15.63,25.73]	2.64	
	Age	0.08	[0.02,0.16]	0.04	0.12*	0.06	[0.01,0.12]	0.03	0.09*	0.03	[–0.02,0.08]	0.03	0.04	0.01	[–0.04,0.06]	0.02	0.02
	Gender[a]	0.15	[–1.51,1.78]	0.87	0.01	–0.28	[–1.43,0.85]	0.63	–0.02	–0.45	[–1.57,0.62]	0.58	–0.03	–0.77	[–1.87,0.27]	0.57	–0.04
	Life satisfaction past					0.59	[0.12,1.06]	0.24	0.11**	0.52	[0.07,0.98]	0.22	0.10**	0.09	[–0.38,0.56]	0.25	0.02
	Life satisfaction present					2.27	[1.69,2.80]	0.30	0.40***	1.89	[1.32,2.42]	0.30	0.34***	1.34	[0.70,1.94]	0.34	0.24***
	Life satisfaction future					1.93	[1.19,2.74]	0.36	0.30***	1.32	[0.63,2.05]	0.36	0.20***	1.00	[0.31,1.76]	0.35	0.16**
	Eudaimonic well-being									0.24	[0.18,0.31]	0.03	0.31***	0.13	[0.05,0.20]	0.04	0.16**
	FLCE													1.62	[0.40,2.84]	0.58	0.17**
	FLAE													1.37	[0.37,2.50]	0.54	0.18**
	R^2				0.02				0.46				0.53				0.56
	ΔR								0.44***				0.07***				0.03***

Hierarchical regression predicting the fulfilled life rating – retrospect using Sample 1 = $N_{Development}$ = 280, hierarchical regression predicting mental well-being using Sample 2 = $N_{Replication}$ = 404.
FLCE, Fulfilled Life Cognitive Experience; FLAE, Fulfilled Life Affective Experience.
[a]Male = 1, female = 2. CI = confidence interval. Confidence interval and standard errors based on 1000 bootstrap samples.
**p < 0.05; **p < 0.01; ***p < 0.001.*

TABLE 12 | Independent samples *t*-test comparing a general and calling exemplar group.

Scale	General		Calling exemplar		*t*	*df*	*p*	BCa 95% CI	Hedges' *g*
	M	*SD*	*M*	*SD*					
USL	4.28	0.90	5.01	0.54	–7.434	60.343	<0.001	[–0.909, –0.536]	0.83
PIL	4.60	0.85	4.91	0.59	–3.059	54.522	0.003	[–0.521, –0.096]	0.38
TWL	4.71	0.84	5.09	0.52	–4.037	58.886	<0.001	[–0.561, –0.188]	0.46
FLCE	4.53	0.74	5.00	0.43	–6.056	62.556	<0.001	[–0.630, –0.312]	0.65
FLAE	4.73	0.94	5.10	0.61	–3.383	56.884	0.001	[–0.584, –0.154]	0.40

Sample 2 = $N_{Replication}$ = 406 (general sample) and Sample 3 = N = 36 (calling exemplar sample).
USL, Unfolded Self and Life; PIL, Positive Impact and Legacy; TWL, The Worthwhile Life; FLCE, Fulfilled Life Cognitive Experience; FLAE, Fulfilled Life Affective Experience. BCa 95% CI for mean difference. t-tests are reported with equal variance not assumed.

fulfillment regarding the self and one's life is empirically more intertwined than expected, at least, in terms of the criteria of wholeness and fit. Conceptually, one can postulate a reciprocal relationship between developing the self and having led a fulfilled life, which might have facilitated the intertwining. Both sources are sufficiently different from impact/legacy, where the three criteria intercorrelated so highly that this factor could be extracted. The other surprising deviation is that the value of self and life intercorrelated so highly that they formed a separate factor. Thus, the 3 × 3 bimodal arrangement has been collapsed into a unimodal separation of three components that now form the new structural model: specifically, fulfillment through life and self-actualization, fulfillment through legacy and impact, and fulfillment by experiencing worthwhileness vis a fulfilled life and having fulfilled one's potential. As the factor structure is replicable, the best items could be selected for the measurement and they yield high Cronbach's alpha values. The three cognitive components yield a differential correlation pattern with predictors and provide a more nuanced understanding of each subconstruct. Additionally, they may be aggregated and are supplemented by a unidimensional affective component. The high correlation of affective fulfillment with The Worthwhile Life dimension is also noteworthy. It seems in line with theoretical reasoning suggesting that feelings of fulfillment, defined as an affective component, accompany the perception of meaning (Reker and Wong, 1988, 2012). Furthermore, evidence resulting from a multi-method approach has revealed strong associations between positive affect and meaning in life (King et al., 2006). Because the effects can be bidirectional, positive affect serves as a source of information for global life evaluations (e.g., Schwarz, 2001) and, as experimental evidence shows, as an enhancer of meaning, and vice versa, in that the perception of meaning is conducive to positive affect (King et al., 2006). A content-valid rating of fulfillment correlated with all FLS subscales, confirming the validity. Deriving fulfillment from legacy and impact had the comparatively lowest predictive power, and the affective fulfillment scale correlated highest with rated fulfillment. Thus, a valid fulfillment profile can now be studied (at facet and scale levels), which was not feasible before. We also intended to address whether the proposition of a new construct and measure is justified after all, and how the construct is located in its nomological network. In the following sections, we will answer these questions and discuss our findings.

TABLE 13 | Partial correlations between fulfilled life rating, fulfilled life scales, sociodemographic and contextual variables.

	Fulfilled Life Rating (retrospect)	USL	PIL	TWL	FLCE	FLAE
Age	0.20***	0.09*	0.14***	0.14***	0.14***	0.18***
Gender[a]	−0.02	−0.04	0.06	0.05	0.03	0.07
Education	0.10**	0.24***	0.19***	0.09*	0.20***	0.14***
Employment status[b]	−0.02	−0.01	−0.03	−0.04	−0.03	−0.05
Financial status	0.19***	0.20***	0.02	0.11**	0.13***	0.20***
Volunteering[c]	0.03	0.03	0.19***	0.06	0.11**	0.05
Marital status[d]	0.11**	0.07	0.04	0.10**	0.08*	0.12**
Parenthood[e]	0.11**	0.05	0.15***	0.13**	0.13**	0.06
Number of children	0.02	0.02	0.11*	0.06	0.07	0.06
Religion/spirituality	0.07	0.03	0.19***	0.21***	0.17***	0.16***
Self-rated health	0.19***	0.22***	0.12**	0.23***	0.22***	0.27***
Childhood	0.26***	0.24***	0.01	0.20***	0.20***	0.22***

$N_{MergedSamples}$ = 379–688. All correlations are controlled for age and gender, except age is controlled for gender and gender is controlled for age.
USL, Unfolded Self and Life; PIL, Positive Impact and Legacy; TWL, The Worthwhile Life; FLCE, Fulfilled Life Cognitive Experience; FLAE, Fulfilled Life Affective Experience.
Childhood was assessed with the question: "How happy was your childhood?"
[a]Male = 1, female = 2.
[b]0 = employed full- or part-time, self-employed, 1 = retired.
[c]0 = no, 1 = yes.
[d]0 = single/never married, separated, divorced, widowed, 1 = married, in a registered partnership.
[e]0 = no, 1 = yes.
**p < 0.05; **p < 0.01; ***p < 0.001.*

A Fulfilled Life's Nomological Network

Related Constructs

Regarding convergent and concurrent validity, significant correlations between the FLS and other constructs were achieved. The FLS subscales had small to large correlations with the temporal life satisfaction scales, with the weakest relationships found for Positive Impact and Legacy. This finding implies that the aspect of fulfillment, which includes making a positive difference to others, is scarcely covered by life satisfaction and indicates a limited overlap between the constructs. Depending on the time perspective regarding life satisfaction, the pattern of associations varied. Concerning past life satisfaction, the

highest association was found for the Unfolded Self and Life. Hence, favorable past life circumstances might especially impact the extent to which a person would have the ability to realize the self and lead life fully. On the other hand, contributing to others might depend less on excellent living conditions while permitting persons to gain fulfillment regardless of their circumstances. Indeed, a few participants reported deficient levels of retrospective life satisfaction and, at the same time, very high levels of a fulfilled life. Present life satisfaction was most strongly associated with affective fulfillment. In contrast, an examination of future life satisfaction found the highest correlation for the Worthwhile Life, suggesting that looking back on a fulfilled life can inspire hope and confidence.

The small to large strength of the correlations with PERMA suggest that the FLS subscales overlap with some orientations toward well-being yet are distinct. Higher associations were found with an orientation to engagement, accomplishment, and meaning than pleasure and positive relationships. Further results demonstrate that eudaimonic functioning plays a role in attaining a fulfilled life, showing a strong relationship with The Worthwhile Life, in particular. Thus, these findings indicate that FLS covers well-being constructs but is also conceptually and empirically distinct.

Additional evidence for convergent validity came from the finding that a favorable psychosocial development, including attaining generativity and ego integrity, was associated with all fulfilled life dimensions, as was theoretically expected.

Concurrent validity was supported by significant positive correlations between the FLS subscales, mental well-being, self-perceptions of aging, and a calling. Fulfilled individuals reported more positive views of their aging and higher levels of mental well-being. A calling was positively related to all FLS dimensions, and the calling exemplars significantly differed from a general sample. These outcomes confirm the theoretical assumptions that individuals experience a calling as fulfilling. Significantly, a calling is not restricted to age or employment status. In the general sample, retired individuals also yielded high means in a calling.

In summarizing this discussion, the bottom line seems to relate to how fulfilled the sample was in actuality, according to the participants' responses. The fact that the rating scales were anchored (running from 0 = *not at all fulfilled* to 10 = *entirely fulfilled*) allows that the average fulfillment is 2.5 points away from maximum, and the different age groups are between 3.0 and 1.5 away from being entirely fulfilled (i.e., 10). When rescaling the FLS scales to the same metric, measured fulfillment is roughly one point below rated fulfillment, showing a similar span regarding the age groups. We do not have absolute scales, nevertheless, future studies will produce similar scores if the identical scale is used and can gradually build a reference or a norm to which new studies can compare. Ideally, a representative sample will be collected. When taking the present research in isolation, the majority of people were located above the scale's midpoint. The question arises whether this can be expected for each sample or whether the current study simply attracted more fulfilled people than a representative sample. Therefore, it will be essential to pay attention to the level of fulfillment in the sample and report it. Eventually, it will also be interesting to compare the mean levels with the means of related constructs (e.g., a particular sample might not be thriving so much but be very fulfilled).

Sociodemographic and Contextual Variables

Several results related to sociodemographic variables are worth highlighting, although correlations were small in size. We found age effects for all fulfilled dimensions, with the numerically highest effect on affective fulfillment. This increase in affective fulfillment with age mirrors previous findings on positive emotions across the life span (e.g., Carstensen et al., 2011). The findings revealed an increase in fulfillment until the age of 70, where the level was maintained. Future studies shall take a closer look at high age, employing sufficiently large samples to cover smaller age spans. It seems that over the lifespan, the potential exists to evolve as a person, realize life goals, and make a personally significant contribution. At an advanced age, the reached level of fulfillment might be kept, and people might enjoy the harvest of previous decades.

Further results of this study seem to indicate that education and financial well-being offer persons more opportunities to unfold and lead a life that suits them well. In contrast, we found no associations between financial status and Positive Impact and Legacy. This outcome might imply that persons with fewer financial resources could still make a meaningful contribution and derive fulfillment. Volunteers reported greater fulfillment regarding Positive Impact and Legacy and general cognitive fulfillment than non-volunteers. Through volunteering, individuals gain a sense of mattering and find a meaningful engagement. We found a small positive effect for married in contrast to unmarried persons for The Worthwhile Life, cognitive fulfillment, and affective fulfillment. Marriage seems to provide a sense of significance and meaningfulness (Schnell, 2009), enriching life and contributing to emotional well-being. Parents reported higher levels of fulfillment regarding Positive Impact and Legacy, The Worthwhile Life, and general cognitive fulfillment. Raising children may enable persons to satisfy their need to be generative and allow them to derive meaning and fulfillment by helping their offspring unfold their potential, and become responsible adults. However, this cross-sectional study cannot rule out an alternative causality. Individuals with better subjective health also reported higher levels of fulfillment in all dimensions. This relationship might consist of an interplay between the variables, in that healthier individuals may benefit from greater vitality to engage with life, and the other way around, where fulfilled individuals reap health benefits. Additionally, spirituality is related to all fulfilled life dimensions, except for the Unfolded Self and Life. This finding seems reasonable, as the relation to the transcendent is partly reflected in the Worthwhile Life and the Positive Impact and Legacy dimensions. A positive childhood experience is also related to a fulfilled life, which might imply that a good start is not only critical but may very well pay dividends later. As positive psychology also focuses on enabling positive institutions and communities, it would be worthwhile to help couples maintain successful relationships, strengthen families, and further efforts to promote positive education.

Proving the Value of This Construct and Measure of a Fulfilled Life

Our findings show that a fulfilled life is a distinct construct and that the FLS is both needed and valuable. To begin with, we found evidence for construct validity. The medium to high correlations between the fulfilled life dimensions and the global ratings demonstrate that the FLS is able to capture participants' subjective evaluation of a fulfilled life. Next, our study could demonstrate that our scale could predict global fulfillment above existent well-being measures when asking individuals how fulfilled their lives had been. Notably, according to our study findings, a global assessment of a fulfilled life was not sufficiently predicted by hedonic and eudaimonic well-being but required both cognitive and affective fulfillment of the FLS to provide significant increments to the prediction. One reasonable conclusion that can be drawn is that the experience of fulfillment goes beyond hedonic and eudaimonic well-being. These findings also confirm that assessing fulfillment requires both strands in terms of cognitive and affective appraisals. In addition, our measure is necessary to adequately assess the construct. Though the global ratings are economical and can be considered a good criterion for a general orientation, our findings show that they do not cover all aspects of a fulfilled life or do so only insufficiently. The FLS as a multidimensional scale assesses a fulfilled life more differentiated as a global rating, which does not cover the legacy aspect sufficiently. Lastly, our results confirm that the FLS has incremental value beyond established well-being measures. Specifically, the FLS significantly predicted mental well-being above and beyond hedonic and eudaimonic well-being. All these findings justify the importance and additional value of the FLS.

Implications

Our findings have various implications. A prerequisite for ultimately being able to look back on a fulfilled life might involve reconsidering one's life periodically and setting the right course. Accordingly, the FLS could be used in psychological counseling on aging or career and life planning, as the assessment helps individuals to reconsider their lives in a differentiated way and reflect on the kind of person they wish to become, the life they want to lead, and the legacy they hope to leave. The result may lead to deeper self-knowledge, greater awareness of what is essential in life, and recognition of various possibilities to shape one's life. Completing the scale could provide new impulses and inspire persons to be more courageous, particularly addressing the view that individuals might regret missed chances in later life, especially in areas they value most (Roese and Summerville, 2005). The results of the FLS could indicate where to refocus or initiate change. For instance, low scores on the Unfolded Self and Life subscale might suggest that a person should pursue more projects that personally matter, use their strengths to a greater extent, or lead a truer life. The FLS can support people in using their remaining lifetime to live a truly fulfilling life and look back on a life that was well lived according to what they hold most dear.

Further implications pertain to the societal level. Supporting individuals in the second half of life to facilitate living a fulfilling life necessitates revising classic structures in the labor market and education system need revision to permit persons of advanced age to participate in society, realize their potential, find meaningful engagement, and contribute. Appreciation of the gained years due to increasing life expectancy and a more positive view on aging would encourage individuals to consciously use the additional life years and society to capitalize on the human potential of older adults. Therefore, supporting a change in societal attitudes and promoting opportunities for older people would lead to more fulfilled seniors and benefit society as a whole (McNaught, 1994).

Limitations and Future Research

The studies have several limitations that require some consideration. For example, the samples were convenience samples. Most of the participants were highly educated, and Samples 1 and 2 consisted mainly of women, which may limit the generalizability of the findings. Moreover, as the participants were not paid, we cannot rule out that that some participants volunteered to take part due to their specific interest in the topic. Besides, conducting the study using online surveys could have excluded older individuals, in particular, who were not proficient in using electronic devices. We must point out that while the replication we achieved indicates the stability of a previously found factor structure, it does not provide evidence of model accuracy. Another task for future studies will be establishing test–retest reliability. In this regard, we expect high stability along with an element of malleability. Future studies will also have to continue to explore the role of age. In particular, further research is needed to expand upon the finding that the increase of fulfillment levels off at the age of 70. The high intercorrelation of the subscale The Worthwhile Life with affective fulfillment also needs further investigation. Our studies constitute a first step in empirical research on the promising topic of a fulfilled life. More research is needed to build a more comprehensive understanding of the construct, its correlates, and consequences. For instance, it might be of interest to relate the FLS to the recently proposed concept of psychological richness (Oishi and Westgate, 2021). Alternatively, further research could test the cross-cultural applicability of the FLS. As our scale refers to a fulfilled life in retrospect, future work might construct a scale that assesses a fulfilling life at present or fulfillment in activities. It is necessary to acknowledge that creating a fulfilling life also depends on the environment, including such elements as freedom, security, or institutional quality. Lastly, future research could also examine whether the specific components of a Fulfilled Life could be trained.

CONCLUSION

The availability of a reliable and valid measure can pave the way for future research on fulfillment and therefore make an essential contribution to the field of positive psychology and in the area of aging well. The measure holds great practical relevance, too, as it allows individuals to take stock of their lives, gain valuable

information, and initiate modifications toward a more fulfilling life. Our results provide first insights into the understanding of a fulfilled life, and further research will be required to achieve a more comprehensive knowledge. Finally, our findings indicate that to arrive at a fulfilled life does count and that a fulfilled life is worth measuring.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

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

AUTHOR CONTRIBUTIONS

DB and WR: conception and design of the work, interpretation of data analysis, and final approval of the published version. DB: data collection, data analysis, and drafting of the manuscript. WR: critical revision of the manuscript. Both authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors are grateful to Alexander Georg Stahlmann for his comments on an earlier version of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.795931/full#supplementary-material

REFERENCES

Baumann, D., and Ruch, W. (2021). *The Unveiling of a Treasure and the Proposal of a Conceptualization: Fulfillment in Life*. Zurich: University of Zurich.

Beyer, A. K., Wolff, J. K., Warner, L. M., Schüz, B., and Wurm, S. (2015). The role of physical activity in the relationship between self-perceptions of ageing and self-rated health in older adults. *Psychol. Health* 30, 671–685. doi: 10.1080/08870446.2015.1014370

Browne, M. W., and Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociol. Methods Res.* 21, 230–258. doi: 10.1177/0049124192021002005

Busch, H., Hofer, J., Šolcová, I. P., and Tavel, P. (2018). Generativity affects fear of death through ego integrity in German, Czech, and Cameroonian older adults. *Arch. Gerontol. Geriatr.* 77, 89–95. doi: 10.1016/j.archger.2018.04.001

Carstensen, L. L., Turan, B., Scheibe, S., Ram, N., Ersner-Hershfield, H., Samanez-Larkin, G. R., et al. (2011). Emotional experience improves with age: evidence based on over 10 years of experience sampling. *Psychol. Aging* 26, 21–33. doi: 10.1037/a0021285

Dik, B. J., Eldridge, B. M., Steger, M. F., and Duffy, R. D. (2012). Development and validation of the Calling and Vocation Questionnaire (CVQ) and Brief Calling Scale (BCS). *J. Career Assess.* 20, 242–263. doi: 10.1177/1069072711434410

Erikson, E. (1985). *Childhood and Society*, 2nd Edn. New York, NY: Norton.

Eysenck, H. J., and Eysenck, S. B. G. (1991). *Manual of the Eysenck Personality Scales (EPQ adults)*. London: Hodder and Stoughton.

Gander, F., Proyer, R. T., and Ruch, W. (2017). The subjective assessment of accomplishment and positive relationships: initial validation and correlative and experimental evidence for their association with well- being. *J. Happin. Stud.* 18, 743–764. doi: 10.1007/s10902-016-9751-z

Goldberg, L. R. (2006). Doing it all bass-ackwards: the development of hierarchical factor structures from the top down. *J. Res. Pers.* 40, 347–358. doi: 10.1016/j.jrp.2006.01.001

Grühn, D., Kotter-Grühn, D., and Röcke, C. (2010). Discrete affects across the adult lifespan: evidence for multidimensionality and multidirectionality of affective experiences in young, middle-aged and older adults. *J. Res. Pers.* 44, 492–500. doi: 10.1016/j.jrp.2010.06.003

Hall, D. T., and Chandler, D. E. (2005). Psychological success: when the career is a calling. *J. Organ. Behav.* 26, 155–176. doi: 10.1002/job.301

Hirschi, A. (2011). Callings in career: a typological approach to essential and optional components. *J. Vocation. Behav.* 79, 60–73. doi: 10.1016/j.jvb.2010.11.002

Hofer, J., Busch, H., Chasiotis, A., Kärtner, J., and Campos, D. (2008). Concern for generativity and its relation to implicit pro-social power motivation, generative goals, and satisfaction with life: a cross-cultural investigation. *J. Pers.* 76, 1–30. doi: 10.1111/j.1467-6494.2007.00478.x

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* 30, 179–185. doi: 10.1007/BF02289447

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

Hutcheson, G., and Sofroniou, N. (1999). *The Multivariate Social Scientist*. Thousand Oaks, CA: Sage. doi: 10.4135/9780857028075

King, L. A., Hicks, J. A., Krull, J. L., and Del Gaiso, A. K. (2006). Positive affect and the experience of meaning in life. *J. Pers. Soc. Psychol.* 90, 179–196. doi: 10.1037/0022-3514.90.1.179

Krohne, H. W., Egloff, B., Kohlmann, C. W., and Tausch, A. (1996). Untersuchungen mit einer deutschen Version der "Positive and Negative Affect Schedule" (PANAS) [Investigations with a German version of the positive and negative affect schedule (PANAS)]. *Diagnostica* 42, 139–156. doi: 10.6102/zis146

Lawton, M. P. (1975). The Philadelphia geriatric center morale scale: a revision. *J. Gerontol.* 30, 85–89. doi: 10.1093/geronj/30.1.85

McAdams, D. P., and de St. Aubin, E. (1992). A theory of generativity and its assessment through self-report, behavioral acts, and narrative themes in autobiography. *J. Pers. Soc. Psychol.* 62, 1003–1015. doi: 10.1037/0022-3514.62.6.1003

McNaught, W. (1994). "Realizing the potential: some examples," in *Age and Structural Lag: Society's Failure to Provide Meaningful Opportunities in Work, Family and Leisure*, eds M. White Riley, R. L. Kahn, and A. Foner (Hoboken, NJ: John Wiley), 219–235.

Morgenthaler, M. (2010). *Beruf und Berufung: Interviews [Profession and Calling: Interviews]*. Bern: Zytglogge Verlag.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behav. Res. Methods Instr. Comput.* 32, 396–402. doi: 10.3758/BF03200807

Oishi, S., and Westgate, E. C. (2021). A psychologically rich life: beyond happiness and meaning. *Psychol. Rev.* [Epub ahead of print]. doi: 10.1037/rev0000317

Pavot, W., Diener, E., and Suh, E. (1998). The temporal satisfaction with life scale. *J. Pers. Assess.* 70, 340–354. doi: 10.1207/s15327752jpa7002_11

Peterson, C., Park, N., and Seligman, M. E. P. (2005). Orientations to happiness and life satisfaction: the full life versus the empty life. *J. Happin. Stud.* 6, 25–41. doi: 10.1007/s10902-004-1278-z

Reker, G. T., and Wong, P. T. (1988). "Aging and the individual process: toward a theory of personal meaning," in *Emergent Theories of Aging*, eds J. E. Birren and V. L. Bengston (Berlin: Springer), 214–246.

Reker, G. T., and Wong, P. T. (2012). "Personal meaning in life and psychosocial adaptation in the later years," in *The Human Quest for Meaning: Theories, Research, and Applications*, ed. P. T. P. Wong (Milton Park: Routledge), 433–456.

Roese, N. J., and Summerville, A. (2005). What we regret most and why. *Pers. Soc. Psychol. Bull.* 31, 1273–1285. doi: 10.1177/0146167205274693

Ruch, W. (1999). Die revidierte Fassung des Eysenck Personality Questionnaire und die Konstruktion des deutschen EPQ-R bzw. EPQ-RK [The Eysenck Personality Questionnaire-Revised and the construction of German standard and short versions (EPQ-R and EPQ-RK)]. *Z. Diff. Diagn. Psychol.* 20, 1–24. doi: 10.1024//0170-1789.20.1.1

Ruch, W., Harzer, C., Proyer, R. T., Park, N., and Peterson, C. (2010). Ways to happiness in German-speaking countries: the adaptation of the German version of the Orientations to Happiness Questionnaire in paper- pencil and internet samples. *Eur. J. Psychol. Assess.* 26, 227–234. doi: 10.1027/1015-5759/a000030

Ryff, C. D., and Heincke, S. G. (1983). Subjective organization of personality in adulthood and aging. *J. Pers. Soc. Psychol.* 44, 807–816. doi: 10.1037/0022-3514.44.4.807

Schnell, T. (2009). The Sources of Meaning and Meaning in Life Questionnaire (SoMe): relations to demographics and well-being. *J. Positive Psychol.* 4, 483–499. doi: 10.1080/17439760903271074

Schwarz, N. (2001). "Feelings as information: implications for affective influences on information processing," in *Theories of Mood and Cognition: A User's Handbook*, eds L. L. Martin and G. L. Clore (Mahwah, NJ: Erlbaum), 159–176.

Seligman, M. E. P. (2011). *Flourish*. Mumbai: Free Press.

Seligman, M. E. P., and Csikszentmihalyi, M. (2000). Positive psychology: an introduction. *Am. Psychol.* 55, 5–14. doi: 10.1037/0003-066X.55.1.5

Simms, L. J. (2008). Classical and modern methods of psychological scale construction. *Soc. Pers. Psychol. Compass* 2, 414–433. doi: 10.1111/j.1751-9004.2007.00044.x

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., et al. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. *Health Qual. Life Outcomes* 5:63. doi: 10.1186/1477-7525-5-63

Trautwein, U. (2004). Die temporalen Facetten der Lebenszufriedenheit [The temporal facets with life satisfaction]. *Diagnostica* 50, 182–192. doi: 10.1026/0012-1924.50.4.182

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika* 41, 321–327. doi: 10.1007/BF02293557

Waterman, A. S., Schwartz, S. J., Zamboanga, B. L., Ravert, R. D., Williams, M. K., Bede Agocha, V., et al. (2010). The questionnaire for eudaimonic well-being: psychometric properties, demographic comparisons, and evidence of validity. *J. Positive Psychol.* 5, 41–61. doi: 10.1080/17439760903435208

Watson, D., and Clark, L. A. (1994). *Manual for the Positive and Negative Affect Schedule – Expanded Form*. Iowa: University of Iowa. doi: 10.17077/48vt-m4t2

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037//0022-3514.54.6.1063

Westerhof, G. J., Dittmann-Kohli, F., and Thissen, T. (2001). Beyond life satisfaction: lay conceptions of well-being among middle-aged and elderly adults. *Soc. Indic. Res.* 56, 179–203. doi: 10.1023/A:1012455124295

Wrzesniewski, A., McCauley, C., Rozin, P., and Schwartz, B. (1997). Jobs, careers, and callings: people's relations to their work. *J. Res. Pers.* 31, 21–33. doi: 10.1006/jrpe.1997.2162

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

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

Copyright © 2022 Baumann and Ruch. 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.

frontiers in Public Health

ORIGINAL RESEARCH
published: 13 January 2022
doi: 10.3389/fpubh.2021.767534

Psychometric Properties of the Diener Satisfaction With Life Scale With Five Response Options Applied to the Colombian Population

Begoña Espejo[1†], *Marta Martín-Carbonell*[2*†], *Irene Checa*[1†], *Yadid Paternina*[2], *Martha Fernández-Daza*[2], *Juan D. Higuita*[3], *Angela Albarracín*[4] *and Ara Cerquera*[4]

[1] *Department of Behavioral Sciences Methodology, School of Psychology, Universidad de Valencia, Valencia, Spain,* [2] *Psychology Program, Universidad Cooperativa de Colombia, Santa Marta, Colombia,* [3] *Psychology Program, Universidad Cooperativa de Colombia, Medellín, Colombia,* [4] *Psychology Program, Universidad Pontificia Bolivariana de Bucaramanga, Bucaramanga, Colombia*

OPEN ACCESS

Edited by:
Peter ten Klooster,
University of Twente, Netherlands

Reviewed by:
Esra Aki,
Hacettepe University, Turkey
Claudia Charry,
Santo Tomás University, Colombia
Alexander Moreno,
Université de Montréal, Canada

***Correspondence:**
Marta Martín-Carbonell
martha.martinc@campusucc.edu.co

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Public Mental Health, a section of the journal Frontiers in Public Health

Received: *31 August 2021*
Accepted: *08 December 2021*
Published: *13 January 2022*

Citation:
Espejo B, Martín-Carbonell M, Checa I, Paternina Y, Fernández-Daza M, Higuita JD, Albarracín A and Cerquera A (2022) Psychometric Properties of the Diener Satisfaction With Life Scale With Five Response Options Applied to the Colombian Population. Front. Public Health 9:767534. doi: 10.3389/fpubh.2021.767534

Introduction: The Satisfaction with Life Scale (SWLS), developed by Diener, Emmons, Larsen, and Griffin in 1985, comprises five items with seven response options in terms of agreement–disagreement. Recently, there has been a suggestion to reduce the response options of the SWLS to optimize its applicability in different cultural contexts.

Objective: The study aims to assess the psychometric properties of the SWLS with five response options in the Colombian population. Specifically, we studied the dimensionality, invariance by gender and age (among a group of adolescents and emerging adults under 25 years and a group of adults of intermediate age and established adulthood under 59 years), convergent validity (with optimism), and divergent (with pessimism) and concurrent validity with other measures of well-being (flourishing, positive, and negative affects).

Methodology: This project was a cross-sectional study using a non-probabilistic sample of the general population. Participants were included if they identified themselves as Colombian and were at least 18 years of age. The final sample comprised 1,255 participants. The average age was 25.62 years (SD = 8.60) ranging from 18 to 67 years of age, and 35.8% of the participants were men. In addition to SWLS, we used the Flourishing Scale (FS), Life Orientation Test-Revised (LOT-R), and Scale of Positive and Negative Experience (SPANE).

Results: Cronbach's alpha coefficient (0.842), composite reliability (0.851), and average variance extracted (0.537) showed very good values. CFA was conducted to test the one-dimensional structure of FS, showing excellent goodness of fit [$\chi^2_{(5)} = 15.774$, $p < 0.001$, CFI = 0.992, TLI = 0.985, RMSEA = 0.042, 90% RMSEA CI (0.020, 0.066), and SRMR = 0.016]. The correlations calculated among life satisfaction (SWLS) with flourishing (FS), positive and negative affects (SPANE), optimism, and pessimism (LOT-R) were statistically significant and as expected. Configural, metric, and scalar invariance across gender and age were confirmed. Percentiles were provided for the total score and for age.

Conclusions: The SWLS with five response options has adequate psychometric properties in the Colombian population, and the use of this version (with 5 response options) is recommended due to its greater applicability.

Keywords: satisfaction with life scale, response options, psychometric properties, confirmatory factor analysis, well-being assessment, measurement invariance, structural equation modeling

INTRODUCTION

In recent decades, politicians and governments have shown an increased interest in evaluating well-being (1). Among other reasons, this is due to the accumulated evidence about its impact on health (2), in academic performance (3) and in labor (4), as well as its value to inform government decisions and evaluate programs aimed at promoting mental health and quality of life in risk groups (5). Research shows that the concept of subjective well-being is multidimensional (6). Among the constitutive components of well-being, satisfaction with life has been identified as a distinct construct that involves a cognitive and global assessment of quality of life as a whole. It has also been conceptualized as the self-assessment of an individual's quality of life according to the comparison between their current state and their standard of what is desirable (7).

According to various authors (5, 8), the most widely used measure for this research is the Satisfaction with Life Scale (SWLS), originally developed by Diener et al. (9). These authors suggest that the scale allows access to the positive side of the individual experience and that it emphasizes self-assessment itself, because the person can establish the basis of their evaluation by choosing the domains that they will take into account when assessing their life, regardless of their emotional state (10). The SWLS is based on the theory of global satisfaction originally proposed by Sumner (11), who conceptualized global satisfaction as a positive attitude toward life itself. This implies that it is an evaluation of all areas and stages of life, which includes both the affective and cognitive aspects, according to the person's expectations.

The SWLS consists of five items with seven response options in terms of agreement–disagreement on a Likert scale ranging from 1 to 7. While the authors of the scale (10) did not provide normative data, they proposed cut-off points that correspond directly to the seven response categories: 31–35, very satisfied; 26–30, satisfied; 21–25, slightly satisfied; 20, neutral; 15–19, slightly dissatisfied; 10–14, dissatisfied; and 5–9, very dissatisfied.

There is sufficient evidence of the validity of the satisfaction with life construct and the SWLS scale, verified, first by their ability to detect differences associated with objective conditions and different life circumstances (12); second, by their correlations with measures not based on self-reporting (13); third, by their association with genetic and physiological variables (14); fourth, by significant changes in scores associated with major life events; and finally by the predictive value of suicidal behaviors (15, 16). SWLS scores have also been shown to positively correlate with health variables and negatively with emotional symptoms, negative thoughts, and coping strategies such as experiential avoidance (17).

Since it was introduced, the SWLS has been used in hundreds of studies (13, 15). It has been validated in numerous languages, such as French (18), German (19), Portuguese (20), Turkish (21), Chinese (22), and even sign language (23). Its psychometric properties have been explored with a wide variety of populations, including adolescents, the elderly, and patients with different health problems (24).

However, modifications to the SWLS response options have recently been recommended in order to optimize its applicability in different cultural contexts (25). This goes in line with previous studies that have shown that offering too many response options could be problematic for people with a low cultural level (26) or generate confusion and boredom in respondents who may find it difficult to distinguish subtle differences between categories (27).

Since the use of Likert scales is intended to adequately represent a construct or continuous latent variable, the question of the optimal number of response alternatives and the effect of categorizing continuous variables becomes particularly relevant. Some authors showed in a study that no psychometric advantages were apparent as of six-response options (28), but this is a subject still under discussion (29). On the other hand, and as a result of a systematic review of published literature on this issue (30), it was concluded that it is best to use five response options. In general, according to studies of the International Test Commission, when data are one-dimensional the best fit is achieved when working with four to six categories (31). In addition, the easier it is for the user to respond to this type of measure, the greater its applicability is, allowing its use to be extended to people with limited comprehension or communication skills, who are often excluded from studies, such as the visually or hearing impaired, people with a low education or those who are illiterate, and people with cognitive problems (26). On the other hand, an invariance study was carried out specifically for the SWLS with Italian and African populations (25). It found that the scale may not be sensitive when it comes to detecting low levels of satisfaction with life, so the authors recommended using fewer response options, especially for the South African population.

There are several adaptations of the SWLS in Spanish: one can be found in public domain on the website of Ed Diener (https://eddiener.com/), the main author of the original scale. It was translated by José A. Reyes-Torres. However, we also found several reports using back translation (32–36). These adaptations differ both in the wording of the items and in their order of presentation, as well as in the number of options and the text of the responses.

Several studies in Ibero-American populations have used the SWLS with five instead of the seven options from the original scale, e.g., in Spain (32, 33, 35, 37), in Chile (38), in Peru (39), in Costa Rica (40) and in Puerto Rico (41). In Mexico it was conducted a validation study of the scale with three-response options (disagree, intermediate, and agree) in a national sample of 13,220 adults above the age of 50 years, finding adequate internal consistency, criterion validity, and confirmation of the one-factor structure (42). In Colombia, different variants of the SWLS have been used in different populations and contexts (43–45). This motivated another authors (46) to conduct an initial validation study with a sample of 121 University students using a version with seven response options (32), but drafted in terms of satisfaction-dissatisfaction (from very dissatisfied to very satisfied), which limits its applicability in cross-cultural studies.

A recent research (47) have recently studied the psychometric properties of Atienza's version (36) in a Colombian sample but with seven response options, instead of the five whose translation was validated by Atienza et al. They have not clarified the origin of the translation used for these seven-answer options. This goes against the recommendations of the International Test Commission for adapting instruments that have been developed in other contexts (31, 48). For the version they studied, Ruiz et al. reported adequate internal consistency and corroborated that the unifactorial model had a very good fit, significant correlations, went in the expected direction with other measures of well-being, and showed metric and scalar gender invariance and with a Spanish sample. However, they did not report normative data that could facilitate its application in contexts where individual differences need to be established.

For these reasons, the present study aimed to evaluate the psychometric properties of Diener's SWLS in the Spanish version by Atienza et al. (35, 36) with Colombian population and using five response options. In addition to being easier to answer for our population, it has been and is widely used in Ibero-American contexts. We investigated whether the one-dimensionality of the instrument with five response options was maintained, as well as the gender invariance. We also studied the age invariance of the SWLS in Colombians, which had not been done previously and constitutes a topic of interest given then reports on the absence of age invariance in other samples (24). We were interested in corroborating the concurrent validity of the SWLS with measures of optimism and pessimism because of existing evidence regarding their relationship (49–51) and their convergent validity with other measures related to well-being, such as prosperity and positive and negative affects (51–53). Finally, another objective was to obtain information to interpret the scores of the Colombian population.

METHODS

Participants

The present project was a cross-sectional study using a non-probabilistic sample of the general population. Participation in the survey was completely anonymous and voluntary, and no participant received any type of financial compensation for it. Participants were included if they identified themselves as Colombian and were at least 18 years old. The final sample comprised of 1,255 participants. The average age was 25.62 years ($SD = 8.60$) ranging from 18 to 67 years, and 64.5% of the participants were female. People who had completed University or graduate studies formed the majority (42.9%) along with those who had completed high school (41.2%), 12.9% had completed secondary school, and only 2.7% of the sample had completed or partially attended primary education; 75.5% were single, 22% were married or had an intimate partner, and 2.5% were divorced or widowed. On the contrary, 43.9% were full-time students and 26.1% were in school and had sporadic or part-time jobs; regarding working status, 23.7% were employed or self-employed, 4.9% were unemployed, 1% were inactive, and 0.4% were retired.

Measures

Satisfaction With Life Scale

This is an instrument designed to measure global cognitive judgment of satisfaction with one's life (9). For the present study, we used the Spanish adaptation with five response options by Atienza et al. (35, 36).

Flourishing Scale

This is an eight-item instrument describing important aspects of human functioning including positive relationships, feelings of competence, and having meaning and purpose in life (54). The instrument uses a seven-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). Total scores range from 8 to 56 with high scores indicating respondents viewing themselves in positive terms in important areas of functioning. This instrument was validated in a general sample of Spanish adults and showed an internal consistency of 0.85 (55). The present study uses the version validated in a Colombian sample, based on the previous Spanish version (56), with an internal consistency of 0.916 in the sample from this study.

Life Orientation Test-Revised

This questionnaire has been used to measure optimism and pessimism (57). The scale is comprised of ten items, four control items, three pessimism items, and three optimism items. Each item of the LOT-R is answered on a five-point Likert scale that ranges from 1 (strongly disagree) to 5 (strongly agree). The scores range from 0 to 12. Higher scores in both subscales indicate high optimism or high pessimism, respectively. This scale was validated in a sample of Colombian adults and showed good psychometric properties (58, 59). Internal consistency in this sample is 0.693 for the Optimism subscale, and 0.636 for the Pessimism subscale. Although some authors question Cronbach's alpha values lower than 0.70, this consideration should not be taken as a "golden rule," especially due to the reduced number of items on the LOT subscales, since an alpha that is too high could lead one to think that, in reality, the three items measure the same indicator of the construct (60).

Scale of Positive and Negative Experience

This scale allows us to learn how the person evaluates the frequency with which they experience positive and negative

feelings, as well as the balance of affections. To this effect, 12 adjectives organized in 2 subscales of 6 items each are used: 6 positive (SPANE-P) and 6 negative experiences (SPANE-N), measuring 3 general and 3 specific emotions in each subscale. The instrument uses a five-point Likert scale ranging from 1 (very rarely or never) to 5 (very often or always). Total scores range from 6 to 30, with high scores indicating a high positive or high negative affect. SPANE-P and SPANE-N can be subtracted to obtain a balanced measure (SPANE-B) that ranges from −24 to 24. In this study, the adapted version was used for a general sample of Colombians (61). The internal consistency in this sample was 0.811 for SPANE-P and 0.799 for SPANE-N.

Procedures

Data from a larger study aimed at validating well-being scales in the Colombian population were used. Following the recommendations of Muñiz et al. (48), an initial qualitative pilot study was conducted. The pilot study's participants were selected using purposeful theoretical sampling serially until obtaining data saturation. In total, 14 people were included based on their willingness to collaborate and after ensuring they were Colombian adults (9 women and 5 men), with different education levels (8 people with a University education, 3 with high school diplomas, and 3 with primary school education), and between the ages of 18 and 81. The scale was responded to using paper and pencil and in an online version. The analysis of the participants' responses revealed that the wording of the items in the version for Spaniards was appropriate for the Colombian context and that the participants correctly understood the items in both versions (paper and pencil, and online). All participants stated that they understood the response options and had no difficulties in choosing the one they considered appropriate in both application formats.

Participants were recruited by different means (email, social networks, and also face-to-face). Data were collected online, with LimeSurvey, an open-source survey tool. When accessing the survey, an explanation of the study was presented, and participants had to read and accept an online informed consent before answering the survey. The study was approved by the Ethics Committee of the Cooperative University of Colombia, which guarantees that data collection complied with the Colombian Law of Data, ensuring confidentiality and anonymity.

Data Analysis

Before studying the construct validity by means of a confirmatory factor analysis, the distribution of frequencies and percentages of the sociodemographic variables was investigated. The means and asymmetry coefficients of the items were checked, as well as the magnitude of the inter-item correlations (using Pearson's correlation coefficient). In addition, item-total corrected correlations were calculated for each item. A confirmatory factor analysis was performed to test the one-factor structure. The parameters were calculated using maximum likelihood robust estimation (MLR). While the nature of the data is ordinal, some studies suggest the use of MLR when the distribution of the data does not fit the normal curve and if there are five or more response options (62–64). In these situations, it can be assumed that the data is continuously distributed (65). The solution offered presents very little variability in the parameters (64), less biased standard errors, and good estimates of correlations between factors (66).

In order to study the fit of the data to the model, the Comparative Fit index (CFI), the Tucker-Lewis index (TLI), the root-mean-square error of approximation (RMSEA), and the standard-root-mean residual (SRMR) were used. Values of 0.90 for the CFI and the TLI, as well as values between 0.06 and 0.08 for the RMSEA and SRMR, indicate an acceptable model fit. Values above 0.95 for the CFI and the TLI and values below 0.05 for the RMSEA and SRMR indicate a good fit to the model (67–69). The factor measurement reliability of the SWLS was evaluated with Cronbach's alpha and with the Composite Reliability Index (CRI) (less biased than alpha) (70). The CRI is identical to the Omega coefficient (71) but more adequate when standardized factor loadings are used (72). In addition, the average variance extracted (AVE) (73) was calculated to evaluate the level of variance captured by the factor.

The measurement invariance by gender and age was evaluated by calculating three nested models that impose successive restrictions: configural, metric and scalar. Configural invariance test identical factor structures (i.e., the same number of factors and items and the same patterns of free and fixed loadings), metric invariance test equality of factor loadings, and scalar invariance test equality of factor loadings and thresholds. A configural model was first tested as a baseline model. In this model, all factor loadings and thresholds were estimated freely across groups. Unlike in models with continuous indicators, in models with categorical indicators with delta parameterization, metric invariance cannot be tested separately from scalar invariance (74, 75). Thus, a scalar invariance model was tested where equality constraints were simultaneously imposed on factor loadings and thresholds. Measurement invariance was examined by comparing the fit indices of the configural model and those of the scalar model. We used the cutoff criteria conventionally used. When sample size is adequate (total $N >$ 300) and sample sizes are equal across the groups, a change of $\geq$-0.010 in CFI, supplemented by a change of $\geq$0.015 in RMSEA or a change of $\geq$0.030 in SRMR would indicate non-invariance (76). To study the measurement invariance according to age, the sample was divided into two groups: a group of adolescents and emerging adults (up to 25 years old) and a group of established adults (between 26 and 59 years old), in line with works by other authors (77). In **Table 1** are shown the descriptive statistics for these two groups.

To study the convergent validity of satisfaction with life with other dimensions of well-being, Pearson correlations were calculated between the total scores of this scale and those of FS, SPANE (P and N), Optimism, and Pessimism. Pearson correlations from 0.20 to 0.39 were interpreted as weak; from 0.40 to 0.59 as moderate; and from 0.60 to 0.79 as strong. Above 0.80, the correlation was considered very strong. Finally, descriptive statistics and percentiles were provided for each age group and gender.

In order to carry out the confirmatory factor analysis and the invariance study, the statistical program Mplus 8.6 was used

TABLE 1 | Sociodemographic characteristics for the "emerging adults" group and for the "adults" group.

		Emerging adults		Adults	
		Mean	**Standard deviation**	**Mean**	**Standard deviation**
Age					
		21.07	2.12	34.86	9.37
		N	**%**	***N***	**%**
Gender					
	Male	267	59.47	182	40.53
	Female	574	71.22	232	28.78
Personal situation					
	Single	746	78.69	202	21.31
	Married or cohabiting	91	32.97	185	67.03
	Divorced	2	7.41	25	92.59
	Widowed	2	50.00	2	50.00
Educational level					
	Primary school studies	24	70.59	10	29.41
	Secondary school studies	137	84.57	25	15.43
	High school studies	428	82.79	89	17.21
	College studies				
	Undergraduate studies	250	56.18	195	43.28
Main activity					
	Studying	502	91.11	49	8.89
	Studying and working	237	72.48	90	27.52
	Working	64	21.48	234	78.52
	Unemployed, inactive or retired	38	48.10	41	51.10

All differences are statistically significant ($p < 0.01$), except for Widowed and Unemployed, inactive or retired.

(74). So as to calculate the descriptive correlations among items and with criteria, Cronbach's alpha, the item-total corrected correlations, and the percentiles by groups, IBM SPSS 27 was used.

RESULTS

Table 2 shows the descriptive statistics of the items, the item-total corrected correlations, and the inter-item correlations. Note that the scores are not normally distributed, with values above the midpoint of the response scale predominating. In addition, the correlations among items present moderate values, except for the correlation between Items 3 and 4, which presents a high value.

Excellent fit values were found in the CFA for the one-dimensional model [$\chi^2_{(5)} = 15.774$, $p < 0.001$, CFI = 0.992, TLI = 0.985, RMSEA = 0.042, 90% RMSEA CI (0.020, 0.066), SRMR = 0.016]. The factor loadings were all statistically significant ($p < 0.001$), ranging between 0.605 and 0.828. Cronbach's alpha was 0.842, the CRI was 0.851, and the AVE was 0.537. All of these values can be considered adequate.

Table 3 shows the results for the measurement invariance models by gender and age. The results show that the Colombian SWLS had scalar invariance by gender, and the fit of the one-dimensional model for men and women was good. As can be seen, ΔCFI, ΔRMSEA, and ΔSRMR values are lower than 0.010, 0.015, and 0.030, respectively. Thus, the latent mean values were fixed to zero for men and compared. No differences were found by gender ($b = -0.008$, $z = -0.178$, $p = 0.859$). Regarding the measurement of invariance by age, the results showed scalar invariance for the SWLS, and the fit of the one-dimensional model for emerging adults (until 25 years old) and adults (more than 25 years old) was excellent. Thus, the latent mean values were fixed to zero for emerging adults, showing that adults present more satisfaction ($b = 0.155$, $z = 3.114$, $p = 0.002$).

Regarding validity, the SWLS presented statistically significant correlations ($p < 0.001$) that went in the expected direction regarding the other well-being variables (SPANE-P) ($r = 0.603$), SPANE-N ($r = -0.376$), Flourishing ($r = 0.492$), Optimism ($r = 0.566$) and Pessimism ($r = -0.131$). Finally, **Table 4** shows the descriptive statistics and percentiles for each age group and gender.

DISCUSSION

It is suggested that "One reason for the increasing need for short scales could be a changing way to approach psychological research in general. With research questions becoming more and more complex, involving more and more constructs…" (78). Therefore, the effort to obtain valid instruments to assess well-being that are short and easy to answer is worth it to the point that a version of the SWLS with only three items has already been proposed (79). In the present work, the effort was

TABLE 2 | Descriptive statistics, item-total corrected correlations, and inter-item correlations among the items (Valid $N = 1,222$).

	Item 1	Item 2	Item 3	Item 4	Item 5
Mean	3.67	4.05	4.00	3.94	3.47
Mode	4	4	4	4	4
Standard deviation	1.025	0.899	1.016	0.987	1.266
Skewness	−0.747	−1.031	−0.945	−0.892	−0.412
SE of skewness	0.070	0.070	0.070	0.070	0.070
Kurtosis	0.035	1.140	0.353	0.429	−0.942
SE of kurtosis	0.140	0.140	0.140	0.140	0.140
Minimum	1	1	1	1	1
Maximum	5	5	5	5	5
Item-total corrected correlation	0.663	0.642	0.736	0.682	0.557
Inter-item correlations					
	Item 1	**Item 2**	**Item 3**	**Item 4**	
Item 2	0.539				
Item 3	0.590	0.571			
Item 4	0.587	0.514	0.658		
Item 5	0.435	0.458	0.513	0.437	

SE, standard error; Item 1 = In most ways my life is close to my ideal; Item 2 = The conditions of my life are excellent; Item 3 = I am satisfied with my life; Item 4 = So far, I have got the important things I want in life; Item 5 = If I could live my life over, I would change almost nothing.

TABLE 3 | Measurement invariance models of the SWLS by gender (reference group: men) and by age (reference group: under 25).

Models for gender	χ^2	df	$\Delta\chi^2$	Δgl	CFI	RMSEA	SRMR	ΔCFI	ΔRMSEA	ΔSRMR
Men	11.219*	5			0.985	0.052	0.024			
Women	21.764*	5			0.985	0.064	0.020			
Configural	33.103*	10			0.985	0.060	0.020			
Metric	38.383*	14	3.952	4	0.984	0.052	0.032	−0.001	−0.008	0.012
Scalar	45.330*	18	5.459	4	0.982	0.049	0.038	−0.002	−0.003	0.005
Models for age	χ^2	**df**	$\Delta\chi^2$	**Δgl**	**CFI**	**RMSEA**	**SRMR**	**ΔCFI**	**ΔRMSEA**	
Under 25	14.409*	5			0.991	0.047	0.017			
26–59 (adults)	8.652	5			0.992	0.042	0.020			
Configural	22.993*	10			0.991	0.045	0.018			
Metric	25.145*	14	0.628	4	0.992	0.035	0.021	0.001	−0.010	0.003
Scalar	28.236*	18	1.436	8	0.993	0.030	0.021	0.001	−0.005	0.000

*df, degrees of freedom; $\Delta\chi^2$, Chi Square increase; Δgl, increase in degrees of freedom; CFI, comparative fit index; RMSEA, root-mean-square error of approximation; SRMR, standardized root-mean-square residual; ΔCFI, CFI increase; ΔRMSEA, RMSEA increase; ΔSRMR, SRMR increase. *$p < 0.001$.*

aimed at evaluating the validity of the SWLS with fewer response options in order to facilitate its application in the Colombian context and offer data that facilitate researchers, clinicians, and, in general, professionals interested in the study of well-being to work on interpretation for evaluation purposes and in the design of interventions.

Despite the importance of how the person is asked to scale their response to a question or statement, there is little consensus in the literature regarding the number of points to include on a Likert response scale. Longer response scales have been suggested as preferable because they will increase variability in total scores and therefore would maximize precision and validity (80–83). However, what is important is that the variation in the scores allows for distinctions between individuals on the psychological characteristic that is evaluated. As early as the middle of the last century, Bendig (84) reported the same reliability for three, five, six, or nine answer options, but a decrease in reliability for 11 options. More recently, various studies (29, 85–88) have concluded that many response options can cause difficulties among participants in perceiving differences between alternatives written in a similar way (for example, agree vs. moderately agree), and induce biases in attributing lower numerical values to variables associated with social inequities and gender (89).

TABLE 4 | Descriptive statistics and percentiles for the satisfaction with life scale (SWLS) by age group and by gender.

		Teens/emerging adults (*N* = 817)	Adults (*N* = 405)	Male (*N* = 435)	Female (*N* = 787)
Mean		18.86	19.66	19.24	19.06
Median		19	20	20	20
Mode		20	25	20	20
Standard deviation		4.028	4.168	3.769	4.259
Skewness		−0.545	−0.843	−0.366	−0.722
SE of Skewness		0.086	0.121	0.117	0.087
Kurtosis		0.069	0.773	−0.235	0.339
SE of Kurtosis		0.171	0.242	0.234	0.174
Minimum		5	5	5	5
Maximum		25	25	25	25
Percentiles	5	12	12	13	11
	10	13	14	14	13
	15	15	15	15	15
	20	16	17	16	16
	25	16	17.5	17	16
	30	17	18	17	17
	35	17	18	18	18
	40	18	19	18	18
	45	19	19	19	19
	50	19	20	20	20
	55	20	20	20	20
	60	20	21	20	20
	65	21	22	21	21
	70	21	22	22	22
	75	22	23	22	22
	80	23	24	23	23
	85	23	24	24	24
	90	24	25	24.4	24
	95	25	25	25	25

SE, standard error.

Our study shows that the validity of the SWLS is not affected by a reduction in the number of response options. In fact, this version shows excellent psychometric properties, including evidence of construct and concurrent validity with other measures related to well-being, which coincides with numerous studies carried out in various countries with different variants of the SWLS (18, 20–22, 90) and specifically in the Ibero-American context, with scales with a different number of response options, for example, in Spain (32, 33, 35–37, 91), in Chile (45, 92), in Peru (93), in Mexico (42, 94), in Puerto Rico (41) and in Argentina (95). In this sense, it will be necessary, in future studies, to determine the optimal number of response options, for which it may be useful to work using the Item Response Theory (IRT) so as to delve into the invariant properties of items and optimize the comparison of the results of the scale in different populations, as it is possible that other scales work better in different cultures (31, 83, 96). The use of IRT models in the study of response options can be done by estimating models that do not assume an order in the response categories (such as the nominal Bock model or the rating scale model), which allow estimating a location parameter for each of these response categories. This would allow us to check if the order of the response options is presented as it is assumed when dealing with items with a Likert-type response scale, or if said order is altered (97, 98). Likewise, these location parameters would allow us to check if each answer option has the maximum probability of being chosen for certain values in the trait. In the event that an answer option was less likely to be chosen than other adjacent options, this would indicate that said answer option has no relevance, since people would always prefer to choose one of the adjacent categories. This has been proven in various studies, in which the intermediate category is much less likely to be chosen than the adjacent categories (98, 99). And this can happen depending on the verbal anchor of the intermediate category.

Cross-cultural studies have confirmed the invariance of the unifactorial structure among nations, and initially enabled the detection of large differences that were attributed to sociocultural and socio-economic factors, such as national wealth and democratic governance (100). More recently, cross-cultural research revealed that these differences are also due to how different populations make judgments about satisfaction with life. For example, Emerson et al. (24) conducted a literature review on the cross-cultural invariance of the SWLS that encompassed works published in the last 30 years and included a sample of 27 articles with data from 66,380 respondents across 24 nations. This review corroborated the unifactorial structure of the scale, as well as the invariance by gender, but showed that there was no invariance between age groups and cultures. Similarly, other authors examined the invariance across 26 countries using three different methods, consistently finding configural and metric invariance, but not scalar invariance (101).

In our study, overall gender invariance was found, which was also confirmed for the seven-response option version in the Colombian population, and the total invariance among the youngest population (adolescents and emerging adults) and established adults (47). In this sense, our results are in line with those reported by other investigations with Spaniards (12, 91, 102), as well as with studies in other populations (103–105). However, it differs from other research that has not confirmed age invariance, such as a study carried out in Norwegian population (106). Contradictory results may be explained not only by specific cultural characteristics, but also by the characteristics of the samples, which is why this is a topic that should be studied in depth in the future.

As has been reported in studies with the SWLS around the world, the values of the items above the midpoint of the scale predominate (51). In addition, in our study, we identified differences in satisfaction with life attributable to age, consistent with the results reported by numerous studies (107). However, we did not find differences related to gender. According to Joshanloo and Jovanović (108), research on the relationship between satisfaction with life and gender has shown inconsistent results both in national studies and in

large international studies and meta-analyses, since it can be influenced by various moderators such as sociocultural conditions, income, education, and marital status, among others (109). In Colombia, some studies have found that women have less satisfaction with life than men (110, 111) but it has been evaluated indirectly, based on data derived from econometric analyses. In any case, it is a topic that should continue to be studied, considering the limitations of the present study.

LIMITATIONS

The scope of the results of our study is limited by the type of non-probability sampling used, as the sample was selected because of its accessibility, which restricts generalization, considering the cultural diversity of Colombia. In addition, the sample differs in the proportion in which age is distributed in the Colombian population. Considering the data from the last National Population and Housing Census of Colombia, our sample presents a higher proportion of young adults, as 82% of participants were under 30 years of age, and <1% of participants was above 65 years. However, nationally, those over 65 represent around 9% of the population, while young people between 18 and 30 years constitute ∼16%. Similarly, in our sample, there are more people with a higher educational level than in the population, and it does not include any illiterate people, even though 5.9% of Colombians cannot read or write. Additionally, the online administration restricted participants by allowing only people with Internet access. In this sense, the psychometric properties of the scale should be studied in other populations, such as the rural population, and include representatives from the various ethnic groups that inhabit the national territory. Likewise, for future studies, temporal stability with a test-retest strategy is recommended.

CONCLUSIONS

The five-choice SWLS maintains the excellent psychometric properties of the seven-choice version, with the advantage of being easier to answer. In addition, it presents invariance by gender and age groups, and provisional normative data are offered.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comité de Bioética de Investigación de la Universidad Cooperativa de Colombia. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MM-C, IC, and BE: conceptualization, methodology, and writing—review and editing. IC and BE: data curation. MM-C and BE: writing—original draft preparation. YP and MF-D: collect data in Santa Marta. AA and AC: collect data in Bucaramanga. JH: collect data in Antioquia. MM-C: project administration. All authors have read and agreed to the published version of the manuscript.

FUNDING

The research data were part of a broader study, Project: An Analysis of the Psychometric Proper-ties of Diener's Well-Being Scales among the Colombian Population, funded by the Cooperative University of Colombia, in cooperation with the University of Valencia (Spain) and the Pontifical Bolivarian University of Bucaramanga, Grant Number INV2371.

ACKNOWLEDGMENTS

We would like to thank to our students for their cooperation in data collection. We also thank Crimson Interactive Pvt. Ltd. (Enago) – https://www.enago.com/es/ for their assistance in manuscript translation and editing.

REFERENCES

1. Landes X. Building happiness indicators some philosophical and political issues. *Les ateliers l'éthique.* (2016) 10:4–37. doi: 10.7202/1035325ar
2. Kushlev K, Drummond DM, Diener E. Subjective well-being and health behaviors in 2.5 million Americans. *Appl Psychol Heal Well-Being.* (2020) 12:166–87. doi: 10.1111/aphw.12178
3. Cobo-Rendón R, Pérez-Villalobos M V., Páez-Rovira D, Gracia-Leiva M. A longitudinal study: affective wellbeing, psychological wellbeing, self-efficacy and academic performance among first-year undergraduate students. *Scand J Psychol.* (2020) 61:518–26. doi: 10.1111/sjop.12618
4. Nangoy R, Mursitama TN, Setiadi NJ, Pradipto YD. Creating sustainable performance in the fourth industrial revolution era: the effect of employee's work well-being on job performance. *Manag Sci Lett.* (2020) 10:1037–42. doi: 10.5267/j.msl.2019.11.006
5. Kusier AO, Folker AP. The satisfaction with life scale: philosophical foundation and practical limitations. *Heal Care Anal.* (2021) 29:21–38. doi: 10.1007/s10728-020-00420-y
6. Forgeard MJC, Jayawickreme E, Kern ML, Seligman MEP. Doing the right thing: measuring well-being for public policy. *Int J Wellbeing.* (2011) 1:79–106. doi: 10.5502/ijw.v1i1.15
7. Diener E. Subjective well-being. *Psychol Bull.* (1984) 95:542–75. doi: 10.1037/0033-2909.95.3.542
8. Kobau R, Sniezek J, Zack MM, Lucas RE, Burns A. Well-being assessment: an evaluation of well-being scales for public health and population estimates of well-being among US adults. *Appl Psychol Heal Well Being.* (2010) 2:272–97. doi: 10.1111/j.1758-0854.2010.01035.x
9. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. *J Pers Assess.* (1985) 49:71–5. doi: 10.1207/s15327752jpa4901_13

10. Pavot W, Diener E. Review of the satisfaction with life scale. *Psychol Assess.* (1993) 5:164–72. doi: 10.1037/1040-3590.5.2.164
11. Sumner LW. *Welfare, Happiness, and Ethics*. Oxford: Oxford University Press (1999).
12. Atienza FL, Martinez N, Silva C. Life satisfaction and self-rated health in adolescents: the relationships between them and the role of gender and age. *Span J Psychol.* (2020) 23:1–9. doi: 10.1017/SJP.2020.10
13. Pavot W, Diener E. Review of the satisfaction with life scale. In: Diener E, editor. *Assessing Well-Being: The Collected Works of Ed Diener*. New York, NY: Springer Science + Business Media. 117, Chapter xi, 274 Pages.
14. Triwidyati H, Tentama F. Validity and reliability construct of subjective well-being scale. *Int J Sci Basic Appl Res.* (2020) 51:191–200. Available online at: http://eprints.uad.ac.id/20150/
15. Jebb AT, Morrison M, Tay L, Diener E. Subjective well-being around the world: trends and predictors across the life span. *Psychol Sci.* (2020) 31:293–305. doi: 10.1177/0956797619898826
16. Diener E, Inglehart R, Tay L. Theory and validity of life satisfaction scales. *Soc Indic Res.* (2013) 112:497–527. doi: 10.1007/s11205-012-0076-y
17. Kim ES, Delaney SW, Tay L, Chen Y, Diener E, Vanderweele TJ. Life satisfaction and subsequent physical, behavioral, and psychosocial health in older adults. *Milbank Q.* (2021) 99:209–39. doi: 10.1111/1468-0009.12497
18. Blais MR, Vallerand RJ, Pelletier LG, Brière NMThe satisfaction scale: Canadian-French validation of the satisfaction with life scale. *Can J Behav Sci.* (1989) 21:210–23. doi: 10.1037/h0079854
19. Glaesmer H, Grande G, Braehler E, Roth M. The German version of the satisfactionwith life scale (SWLS) psychometric properties, validity, and population-based norms. *Eur J Psychol Assess.* (2011) 27:127–32. doi: 10.1027/1015-5759/a000058
20. Sancho P, Galiana L, Gutierrez M, Francisco EH, Tomás JM. Validating the portuguese version of the satisfaction with life scale in an elderly sample. *Soc Indic Res.* (2014) 115:457–66. doi: 10.1007/s11205-012-9994-y
21. Durak M, Senol-Durak E, Gencoz T. Psychometric properties of the satisfaction with life scale among Turkish University students, correctional officers, and elderly adults. *Soc Indic Res.* (2010) 99:413–29. doi: 10.1007/s11205-010-9589-4
22. Bai X, Wu C, Zheng R, Ren X. The psychometric evaluation of the satisfaction with life scale using a nationally representative sample of China. *J Happiness Stud.* (2011) 12:183–97. doi: 10.1007/s10902-010-9186-x
23. Wilson JF, Embree J, Guthmann D, Sligar SR, Titus JC, Taylor KC. Satisfaction with life scale in American sign language: validation and normative data. *J Am Deaf Rehabil Assoc.* (2020) 53:1–10. doi: 10.1037/t80262-000
24. Emerson SD, Guhn M, Gadermann AM. Measurement invariance of the satisfaction with life scale: reviewing three decades of research. *Qual Life Res.* (2017) 26:2251–64. doi: 10.1007/s11136-017-1552-2
25. Schutte L, Negri L, Delle Fave A, Wissing MP. Rasch analysis of the satisfaction with life scale across countries: findings from South Africa and Italy. *Curr Psychol.* (2019) 40:4908–17. doi: 10.1007/s12144-019-00424-5
26. Bernal H, Wooley S, Schensul J. The challenge of using Likert-type scales with low-literate ethnic populations. *Nurs Res.* (1997) 46 3:179–81. doi: 10.1097/00006199-199705000-00009
27. Flaskerud JH. Cultural bias and likert-type scales revisited. *Issues Ment Health Nurs.* (2012) 33:130–2. doi: 10.3109/01612840.2011.600510
28. Simms LJ, Zelazny K, Williams TF, Bernstein L. Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychol Assess.* (2019) 31:557–66. doi: 10.1037/pas0000648
29. Claveria O. A new metric of consensus for Likert-type scale questionnaires: an application to consumer expectations. *J Bank Financ Technol.* (2021) 5:35–43. doi: 10.1007/s42786-021-00026-5
30. Matas A. Diseño del formato de escalas tipo Likert: un estado de la cuestión. *Rev Electron Investig Educ.* (2018) 20:38–47. doi: 10.24320/redie.2018.20.1.1347
31. Hernández-Baeza A, Muñiz J, García-Cueto E. Comportamiento del modelo de respuesta graduada en función del número de categorías de la escala. *Psicothema.* (2000) 12:288–91. Available online at: https://www.redalyc.org/pdf/727/72797068.pdf
32. Vázquez C, Duque A, Hervás G. Satisfaction with life scale in a representative sample of Spanish adults: validation and normative data. *Span J Psychol.* (2013) 16:E82. doi: 10.1017/sjp.2013.82
33. Cabañero-Martínez MJ, Richart-Martínez M, Cabrera-García J, Orts-Cortés MI, Reig-Ferrer A, Tosal-Herrero B. Fiabilidad y validez de la Escala de Satisfacción con la Vida de Diener en una muestra de mujeres embarazadas y puérperas. *Psicothema.* (2004) 16:449–55. Available online at: https://www.redalyc.org/pdf/727/72716318.pdf
34. Núñez JL, Martín-Albo J, Domínguez E. Propiedades psicométricas de la escala de satisfacción con la vida en sujetos practicantes de actividad física. *Rev Psicol del Deport.* (2010) 19:291–304. Available online at: https://www.redalyc.org/pdf/2351/235116352009.pdf
35. Atienza FL, Pons D, Balaguer I, García-Merita M. Propiedades Psicométricas de la Escala de Satisfacción con la Vida en Adolescentes. *Psicothema.* (2000) 12:314–9. Available online at: https://reunido.uniovi.es/index.php/PST/article/view/7597
36. Atienza FL, Balaguer I, Garcia-Merita ML. Satisfaction with life scale: analysis of factorial invariance across sexes. *Pers Individ Dif.* (2003) 35:1255–60. doi: 10.1016/S0191-8869(02)00332-X
37. Pons D, Atienza FL, Balaguer I, García-Merita ML. Satisfaction with life scale: analysis of factorial invariance for adolescents and elderly persons. *Percept Mot Skills.* (2000) 91:62–8. doi: 10.2466/pms.2000.91.1.62
38. Álvarez CA, Briceño AM, Álvarez K, Abufhele M, Delgado I. Transcultural adaptation and validation of a satisfaction with life scale for chilean adolescents. *Rev Chil Pediatr.* (2018) 89:51–8. doi: 10.4067/S0370-41062018000100051
39. Calderón-De la Cruz G, Lozano-Chávez F, Cantuarias-Carthy A, Ibarra-Carlos L. Validación de la Escala Satisfacción con la Vida en trabajadores peruanos. *Lib Rev Peru Psicol.* (2018) 24:249–64. doi: 10.24265/liberabit.2018.v24n2.06
40. Blanco M, Salazar M. Escala de Satisfacción con la Vida para adultos mayores de Emmons, Larsen, y Griffin. In: *Compendio de Instrumentos de Medición IIP-2014*. p. 227–231.
41. González-Rivera JA, Rosario-Rodríguez A. Psychometric characteristics of the satisfaction with life scale in a sample of puerto ricans adults. *Rev Puertorriquena Psicol.* (2020) 1:128–39. Available online at: https://dialnet.unirioja.es/servlet/articulo?codigo=7872123
42. López-Ortega M, Torres-Castro S, Rosas-Carrasco O. Psychometric properties of the satisfaction with life scale (SWLS): secondary analysis of the Mexican health and aging study. *Health Qual Life Outcomes.* (2016) 14:1–7. doi: 10.1186/s12955-016-0573-9
43. Hernández-Mateus LV, Salamanca-Chacón SA, Umbarila-Contréras LI, Venegas Muñoz CF, Riveros-Munévar F. Validez de constructo y confiabilidad de la Escala de Satisfacción con la Vida, en escolares de zonas rurales y urbanas de Cundinamarca. *Actual en Psicol.* (2018) 32:71. doi: 10.15517/ap.v32i125.29848
44. Bernal-Romero T, Daza-Pinzón CD, Jaramillo-Acosta PK. Satisfacción con la vida y resiliencia en jóvenes en edad escolar. *Rev Iberoam Psicol Cienc y Tecnol.* (2015) 8:43–53. doi: 10.33881/2027-1786.rip.8204
45. Sanín-Posada A, Salanora M, Vera-Villarroel P. Satisfacción con la Vida: El rol de la Felicidad en el Trabajo y la Vida. *Rev Interam Psicol Ocup.* (2019) 38:54–66. doi: 10.21772/ripo.v38n1a05
46. Vinaccia-Alpi S, Parada N, Quiceno JM, Riveros-Munévar F, Vera-Maldonado LA. Satisfaction with life scale (SWLS): validity, reliability and assessment analysis in college students from Bogotá (Col) as sample. *Psicogente.* (2019) 22:1–13. doi: 10.17081/psico.22.42.3468
47. Ruiz FJ, Suárez-Falcón JC, Flórez CL, Odriozola-González P, Tovar D, López-González S, et al. Validity of the satisfaction with life scale in Colombia and factorial equivalence with spanish data. *Rev Latinoam Psicol.* (2019) 51:58–65. doi: 10.14349/rlp.2019.v51.n2.1
48. Muñiz J, Elosúa P, Padilla JL, Hambleton RK. Test adaptation standards for cross-lingual assessment. In: Wells CS, Faulkner-Bond M. *Educational Measurement. From Foundations to Future.* New York, NY: Guilford Press. p. 291–304.
49. Bailey TC, Eng W, Frisch MB, Snyder CR. Hope and optimism as related to life satisfaction. *J Posit Psychol.* (2007) 2:168–75. doi: 10.1080/17439760701409546

50. Lucas RE, Diener E, Suh E. Discriminant validity of well-being measures. *J Pers Soc Psychol.* (1996) 71:616–28. doi: 10.1037/0022-3514.71.3.616
51. Pavot W, Diener E, Colvin CR, Sandvik E. Further validation of the satisfaction with life scale: evidence for the cross-method convergence of well-being measures. *J Pers Assess.* (1991) 57:149–61. doi: 10.1207/s15327752jpa5701_17
52. Diener E, Heintzelman SJ, Kushlev K, Tay L, Wirtz D, Lutes LD, et al. Findings all psychologists should know from the new science on subjective well-being. *Can Psychol.* (2017) 58:87–104. doi: 10.1037/cap0000063
53. Keyes CLM, Annas J. Feeling good and functioning well: distinctive concepts in ancient philosophy and contemporary science. *J Posit Psychol.* (2009) 4:197–201. doi: 10.1080/17439760902844228
54. Diener E, Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, et al. New well-being measures: short scales to assess flourishing and positive and negative feelings. *Soc Indic Res.* (2010) 97:143–56. doi: 10.1007/s11205-009-9493-y
55. Checa I, Perales J, Espejo B. Spanish validation of the flourishing scale in the general population. *Curr Psychol.* (2018) 37:949–56. doi: 10.1007/s12144-017-9581-0
56. Martín-Carbonell M, Espejo B, Checa I, Fernández-Daza M. Adaptation and measurement invariance by gender of the flourishing scale in a Colombian sample. *Int J Environ Res Public Health.* (2021) 18:2664. doi: 10.3390/ijerph18052664
57. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the life orientation test. *J Pers Soc Psychol.* (1994) 67:1063–78. doi: 10.1037/0022-3514.67.6.1063
58. Valdelamar-Jiménez J, Sánchez-Pedraza R. Traducción y adaptación transcultural de la escala life orientation test-revised para medir optimismo disposicional en cuidadores de pacientes con cáncer en Colombia. *Rev Fac Nac Salud Pública.* (2017) 35:2016–24. doi: 10.17533/udea.rfnsp.v35n2a06
59. Zenger M, Finck C, Zanon C, Jimenez W, Singer S, Hinz A. Evaluation of the Latin American version of the life orientation test-revised. *Int J Clin Heal Psychol.* (2013) 13:243–52. doi: 10.1016/S1697-2600(13)70029-2
60. Ventura-León J, Peña-Calero BN. The world should not revolve around CronbachĄfs alpha ≥ .70. *Adicciones.* (2021) 33:369–72. doi: 10.20882/adicciones.1576
61. Martín-Carbonell M, Checa I, Fernández-Daza M, Paternina Y, Espejo B. Adaptation and psychometric properties of the scale of positive and negative experience (SPANE) in the general Colombian population. *Int J Environ Res Public Heal.* (2021) 18:6449. doi: 10.3390/ijerph18126449
62. Johnson DR, Creech JC. Ordinal measures in multiple indicator models: a simulation study of categorization error. *Am Sociol Rev.* (1983) 48:398. doi: 10.2307/2095231
63. Rhemtulla M, Brosseau-Liard PÉ, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol Methods.* (2012) 17:354–73. doi: 10.1037/a0029315
64. Rigdon EE. Structural equation modeling. In: *Modern Methods for Business Research Quantitative Methodology Series*. Mahwah, NJ: Lawrence Erlbaum Associates. p. 251–94.
65. Raykov T. Scale construction development using structural equation modeling. In: Hoyle RH, editor. *Handbook of Structural Equation Modeling*. New York, NY: The Guilford Press. p. 472–92.
66. Li CH. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res Methods.* (2016) 48:936–49. doi: 10.3758/s13428-015-0619-7
67. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model.* (1999) 6:1–55. doi: 10.1080/10705519909540118
68. Rigdon EE. CFI versus RMSEA: a comparison of two fit indexes for structural equation modeling. *Struct Equ Model.* (1996) 3:369–79. doi: 10.1080/10705519609540052
69. Rigdon EE, Schumacker RE, Wothke W. A comparative review of interaction nonlinear modeling. In: Schumacker RE, Marcoulides GA, editors. *Interaction Nonlinear Effects in Structural Equation Modeling Modern Methods for Business Research*. Mahwah, NJ: Lawrence Erlbaum. p. 251–94.
70. Werts CE, Linn RL, Jöreskog KG. Intraclass reliability estimates: testing structural assumptions. *Educ Psychol Meas.* (1974) 34:25–33. doi: 10.1177/001316447403400104
71. McDonald RP. *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers (1999).
72. Raykov T. Estimation of composite reliability for congeneric measures. *Appl Psychol Meas.* (1997) 21:173–84. doi: 10.1177/01466216970212006
73. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res.* (1981) 18:39–50. doi: 10.1177/002224378101800104
74. Muthén LK, Muthén B. *Mplus User's Guide*. 8th ed. Los Angeles, CA: Muthén & Muthén (2017).
75. Wu H, Estabrook R. Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika.* (2016) 81:1014–45. doi: 10.1007/s11336-016-9506-0
76. Li CH. *The Performance of MLR, USLMV, and WLSMV Estimation in Structural Regression Models With Ordinal Variables*, New York, NY: George A. Marcoulides (2014).
77. Abubakar A, van de Vijver F, Alonso-Arbiol I, He J, Adams B, Aldhafri S, et al. Measurement invariance of the brief multidimensional student's life satisfaction scale among adolescents and emerging adults across 23 cultural contexts. *J Psychoeduc Assess.* (2016) 34:28–38. doi: 10.1177/0734282915611284
78. Ziegler M, Kemper CJ, Kruyen P. Short scales - five misunderstandings and ways to overcome them. *J Individ Differ.* (2014) 35:185–9. doi: 10.1027/1614-0001/a000148
79. Kjell ONE, Diener E. Abbreviated three-item versions of the satisfaction with life scale and the harmony in life scale yield as strong psychometric properties as the original scales. *J Pers Assess.* (2021) 103:183–94. doi: 10.1080/00223891.2020.1737093
80. Cox EP. The optimal number of response alternatives for a scale: a review. *J Mark Res.* (1980) 17:407–22. doi: 10.1177/002224378001700401
81. Bisquerra R. ¿Pueden las escalas Likert aumentar en sensibilidad? *REIRE Rev d'Innovació i Recer en Educ.* (2015) 8:129–47. doi: 10.1344/reire2015.8.2828
82. Wu H, Leung SO. Can likert scales be treated as interval scales?—a simulation study. *J Soc Serv Res.* (2017) 43:527–32. doi: 10.1080/01488376.2017.1329775
83. Calleja N, Reskala-Sánchez FJ, Rivera-Fong L, Buenrostro-Mercado D. Efecto del número de opciones de respuesta en las propiedades psicométricas de cuatro escalas psicosociales. *Rev Psicol y Ciencias del Comport la Unidad Académica Ciencias Jurídicas y Soc.* (2019) 10:100–13. doi: 10.29059/rpcc.20191126
84. Bendig AW. The reliability of self-ratings as a function of the amount of verbal anchoring and of the number of categories on the scale. *J Appl Psychol.* (1953) 37:38. doi: 10.1037/h0057911
85. Dawes J. Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *Int J Mark Res.* (2008) 50:61–77. doi: 10.1177/147078530805000106
86. Hernández A, Espejo B, González-Romá V, Gómez-Benito J. Escalas de respuesta tipo Likert: ¿Es relevante la alternativa "indiferente"? *Metodol Encuestas.* (2001) 3:135–50. Available online at: https://roderic.uv.es/handle/10550/39081
87. Weijters B, Millet K, Cabooter E. Extremity in horizontal and vertical Likert scale format responses. Some evidence on how visual distance between response categories influences extreme responding. *Int J Res Mark.* (2021) 38:85–103. doi: 10.1016/j.ijresmar.2020.04.002
88. Chyung SYY, Roberts K, Swanson I, Hankinson A. Evidence-based survey design: the use of a midpoint on the likert scale. *Perform Improv.* (2017) 56:15–23. doi: 10.1002/pfi.21727
89. Rivera LA, Tilcsik A. Scaling down inequality: rating scales, gender bias, and the architecture of evaluation. *Am Sociol Rev.* (2019) 84:248–74. doi: 10.1177/0003122419833601
90. Arrindell WA, Meeuwesen L, Huyse FJ. The satisfaction with life scale (SWLS): psychometric properties in a non-psychiatric medical outpatients sample. *Pers Individ Dif.* (1991) 12:117–23. doi: 10.1016/0191-8869(91)90094-R

91. Checa I, Perales J, Espejo B. Measurement invariance of the satisfaction with life scale by gender, age, marital status and educational level. *Qual Life Res.* (2019) 28:963–8. doi: 10.1007/s11136-018-2066-2
92. Barrientos-Delgado JE, Cárdenas M. A Confirmatory factor analysis of the spanish language version of the attitudes toward lesbians and gay men (ATLG) measure. *Univ Psychol.* (2012) 11:579–86. doi: 10.11144/Javeriana.upsy11-2.cfas
93. Inga-Aranda J, Vara-Horna A. Factors associated to life satisfaction in adults over 60 years old In Lima-Peru. *Univ Psychol.* (2006) 5:475–86. Available online at: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-92672006000300004
94. Padrós-Blázquez F, Soriano-Mas C, Navarro-Contreras G. Afecto positivo y negativo: ¿Una dimensión bipolar o dos dimensiones unipolares independientes? *Interdiscip Rev Psicol y Ciencias Afines.* (2012) 29:151–64. doi: 10.16888/interd.2012.29.1.9
95. Moyano NC, Martínez-Tais M, Muñoz MP. Propiedades psicométricas de la Escala de Satisfacción con la Vida de Diener. *Rev Argentina Clínica Psicológica.* (2013) 22:161–8. Available online at: https://www.redalyc.org/pdf/2819/281931436008.pdf
96. Asún-Inostroza R, Zúñiga-Rivas C. Ventajas de los modelos politómicos de teoría de respuesta al ítem en la medición de actitudes sociales. *El análisis de un caso*. *Psykhe.* (2008) 17:103–15. doi: 10.4067/S0718-22282008000200009
97. González-Romá V, Espejo B. Testing the middle response categories "Not sure", "in between" and "?" in polytomous items. *Psicothema.* (2003) 15:278–84. Available online at: https://roderic.uv.es/handle/10550/39080
98. Murray AL, Booth T, Molenaar D. When middle really means "top" or "bottom": an analysis of the 16PF5 using Bock's nominal response model. *J Pers Assess.* (2016) 98:319–31. doi: 10.1080/00223891.2015.1095197
99. Hernández A, Espejo B, González-Romá V. The functioning of central categories middle level and sometimes in graded response scales: does the label matter? *Psicothema.* (2006) 18:300–6. Available online at: https://roderic.uv.es/handle/10550/39123
100. Inglehart R. *Culture Shift in Advanced Industrial Society*. Princeton, New Jersey: Princeton University Press (1990).
101. Jang S, Kim ES, Cao C, Allen TD, Cooper CL, Lapierre LM, et al. Measurement invariance of the satisfaction with life scale across 26 countries. *J Cross Cult Psychol.* (2017) 48:560–76. doi: 10.1177/0022022117697844
102. Esnaola I, Benito M, Antonio-Agirre I, Axpe I, Lorenzo M. Longitudinal measurement invariance of the satisfaction with life scale in adolescence. *Qual Life Res.* (2019) 28:2831–7. doi: 10.1007/s11136-019-02224-7
103. Bronk KC, Hill PL, Lapsley DK, Talib TL, Finch H. Purpose, hope, and life satisfaction in three age groups. *J Posit Psychol.* (2009) 4:500–10. doi: 10.1080/17439760903271439
104. Gestsdottir S, Magnusson K, Arngrimsson SA, Johannsson E, Arnarsson A, Sveinsson T. Gender differences in development of mental well-being from adolescence to young adulthood: an eight-year follow-up study. *Scand J Public Health.* (2015) 43:269–75. doi: 10.1177/1403494815569864
105. Tomás JM, Gutiérrez M, Sancho P, Romero I. Measurement invariance of the satisfaction with life scale (SWLS) by gender and age in Angola. *Pers Individ Dif.* (2015) 85:182–6. doi: 10.1016/j.paid.2015.05.008
106. Clench-Aas J, Nes RB, Dalgard OS, Aarø LE. Dimensionality and measurement invariance in the satisfaction with life scale in Norway. *Qual Life Res.* (2011) 20:1307–1317. doi: 10.1007/s11136-011-9859-x
107. Steptoe A, Deaton A, Stone AA. Subjective wellbeing, health, and ageing. *Lancet.* (2015) 385:640–8. doi: 10.1016/S0140-6736(13)61489-0
108. Joshanloo M, Jovanović V. The relationship between gender and life satisfaction: analysis across demographic groups and global regions. *Arch Womens Ment Health.* (2020) 23:331–8. doi: 10.1007/s00737-019-00998-w
109. Graham C, Chattopadhyay S. Gender and well-being around the world. *Int J Happiness Dev.* (2013) 1:212–32. doi: 10.1504/IJHD.2013.055648
110. Cardona L, Gómez-Toro C, Henao-Duue JF. ¿Costumbre o Reacción? El efecto de ser víctima de delitos y corrupción en la satisfacción con la vida de los colombianos. In: *Center for Research in Economics and Finance (CIEF). Working Papers.* Available online at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2849148 (accessed July 25).
111. Peláez-Madrigal N. *Caracterización del reporte de la satisfacción con la vida de las mujeres en Colombia.* (2019). Available online at: https://repository.eafit.edu.co/handle/10784/14947 (accessed July 25).

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

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

Copyright © 2022 Espejo, Martín-Carbonell, Checa, Paternina, Fernández-Daza, Higuita, Albarracín and Cerquera. 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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 07 March 2022
doi: 10.3389/fpsyg.2022.795686

The Life Crafting Scale: Development and Validation of a Multi-Dimensional Meaning-Making Measure

Shi Chen[1], Leander van der Meij[1], Llewellyn E. van Zyl[1,2,3,4] and Evangelia Demerouti[1,5]*

[1] Human Performance Management Group, Department of Industrial Engineering and Innovation Sciences, University of Eindhoven, Eindhoven, Netherlands, [2] Optentia Research Focus Area, North-West University (VTC), Vanderbijlpark, South Africa, [3] Department of Human Resource Management, University of Twente, Enschede, Netherlands, [4] Department of Social Psychology, Institut für Psychologie, Goethe University, Frankfurt am Main, Germany, [5] Department of Industrial Psychology and People Management, University of Johannesburg, Johannesburg, South Africa

Finding meaning in our lives is a central tenet to the human experience and a core contributor to mental health. Individuals tend to actively seek the sources of meaning in their lives or consciously enact efforts to create or "craft" meaning in different life domains. These overall "Life Crafting" behaviors refer to the conscious efforts individuals exert to create meaning in their lives through (a) cognitively (re-)framing how they view life, (b) seeking social support systems to manage life challenges, and (c) actively seeking challenges to facilitate personal growth. Specifically, these behaviors are actioned to better align life goals, personal needs, values, and capabilities. However, no psychological assessment instrument currently exists to measure overall life crafting. As such, the purpose of this paper was twofold: to conceptualize life crafting and to develop, validate and evaluate a robust measure of overall life crafting. A mixed-method, multi-study research design was employed. *First*, nine participants were interviewed to determine the methods or techniques used to craft meaningful life experiences. These methods/techniques were used as indicators to create an initial item pool which was then reviewed by a panel of experts to ensure face validity. *Second*, in Study 1, the factorial structure of the instrument was explored by gathering data from a convenience sample (N = 331), with the results showing support for a three-factor structure of life crafting, consisting of (a) cognitive crafting, (b) seeking social support, and (c) seeking challenges. *Finally*, in Study 2 (N = 362), the aim was to confirm the factorial structure of the Life Crafting scale and to determine its level of internal consistency, partial measurement invariance across genders, and criterion validity [meaning in life (β = 0.91), mental health (β = 0.91), work engagement (β = 0.54), and job burnout (β = −0.42)]. The results supported a second-order factorial model of Life Crafting, which comprised of three first-order factors (cognitive crafting, seeking social support, and seeking challenges). Therefore, the Life Crafting Scale can be used as a valid and reliable instrument to measure- and track the effectiveness of life crafting interventions.

Keywords: life crafting, seeking challenges, seeking resources, cognitive crafting, meaning in life, wellbeing, scale development, meaning making

OPEN ACCESS

Edited by:
Paul T. P. Wsong, Trent University, Canada

Reviewed by:
Rita Berger, University of Barcelona, Spain
Emanuela Ingusci, University of Salento, Italy

****Correspondence:***
Shi Chen
s.chen2@tue.nl

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

Received: *15 October 2021*
Accepted: *11 January 2022*
Published: *07 March 2022*

Citation:
Chen S, van der Meij L, van Zyl LE and Demerouti E (2022) The Life Crafting Scale: Development and Validation of a Multi-Dimensional Meaning-Making Measure. Front. Psychol. 13:795686. doi: 10.3389/fpsyg.2022.795686

INTRODUCTION

The COVID-19 pandemic has fundamentally changed the way individuals view life and approach work (Frenzel et al., 2022). In the absence of validated treatment strategies or vaccines, governments across the globe opted to introduce a series of non-pharmaceutical interventions (NPIs) to manage the spread of the disease (Chowdhury et al., 2020). These NPIs aimed to control the transmissibility of the disease through social distancing, case-based isolation strategies, quarantine, and community containment procedures (Frenzel et al., 2022). These NPIs led to ever-increasing restrictions on personal freedoms ranging from (inter) national travel bans, large-scale business/school/university closures, limitations on daily social engagements, and instructions to work-from-home (Van Zyl, 2021). This, in turn, resulted in large-scale layoffs, a major decline in the global economy, and radical changes in individuals' daily activity patterns (Frenzel et al., 2022). In a relatively short period, most (if not all) individuals experienced drastic changes in how they worked, studied, shopped, and connected to others which significantly impacted their mental health and wellbeing (de Jong et al., 2020; Van Zyl, 2021). de Jong et al. (2020) argued that as loneliness and boredom set in during the initial stages of the pandemic, it led to an increase in depression, general mental health issues, and irrational decision making, which in turn increased social monitoring and eroded social bonds. In order to cope with the adverse effects of these radical changes, Lin (2021) argued that individuals started to alter the meaning they attach to and derive from these life events. Understanding the meaning or purpose of these radical changes and how individual actions, such as wearing a face mask, may contribute to the greater good, may help buffer against the negative effects these NPIs have on individuals' mental health (Lin, 2021). Searching for or creating/crafting meaning is, therefore, an essential personal resource that individuals can employ to help cope with or make sense of the misery caused during the COVID-19 pandemic (de Jong et al., 2020; Lin, 2021).

Therefore, having a sense of purpose and meaning in our lives is a central tenet to the human experience and a core contributor to enhancing or maintaining mental health or wellbeing during times of extreme uncertainty (van Zyl et al., 2020a). Meaning, defined as "the sense that people make of their existence and having an overarching life purpose they pursue" (Steger et al., 2014, p. 27), has shown to be a critical element for better functioning in almost every life domain ranging from home to work (Steger, 2009, 2013). When individuals are actively engaging in activities which they deem to be meaningful, they are more likely to be happier (Steger, 2019), and physically healthier (Czekierda et al., 2017) as well as less likely to be depressed, stressed, or anxious (Steger et al., 2014; van Zyl et al., 2020a). When individuals are facilitated to discover what truly matters to them and are provided with the flexibility to pursue these life goals/aspirations, they show less psychopathology and show more organizational citizenship behavior, work engagement, job satisfaction, and even perform better at work (Maharaj and Schlechter, 2007; Van Zyl et al., 2010; David and Iliescu, 2020). Research has also shown that having a sense of meaning or purpose during the COVID-19 pandemic is associated with increased levels of life satisfaction, more pro-social behaviors, less psychological distress, and lower levels of negative affect over time (Lin, 2021). Further, when controlling for the presence of meaning, individuals' perceptions of the outbreak and the adverse effects of self-quarantine did not affect individuals' wellbeing (Lin, 2021). Given its importance, it is not surprising that practitioners, researchers, and organizations have become interested in finding practical ways to aid individuals to cultivate meaning in their lives (van Zyl et al., 2020a,b).

Jacob and Steger (2021) argued that individuals could cultivate meaning through either (a) identifying the sources of meaning in one's life and aiding individuals to actively pursue activities aligned to such or (b) aiding individuals to craft meaning in various life domains. A considerable amount of attention has been placed on aiding individuals in identifying the sources of meaning in their lives ranging from meaning-centered therapy and positive psychology coaching to self-help activities such as photo-ethnography (c.f., Steger et al., 2014; van Zyl et al., 2020a; Richter et al., 2021). These approaches are designed to help individuals find activities which they deem to be meaningful and are facilitated to pursue these more actively to help buffer against the impact radical life challenges such as the COVID-19 pandemic has on their mental health/wellbeing (Steger et al., 2014). In contrast, aiding individuals to craft meaning in specific life domains (e.g., work-, home-, leisure-, or relationships) has only recently started to gain popularity in the literature (Tims and Bakker, 2010; Demerouti et al., 2020). From this perspective, individuals are empowered not necessarily to pursue new sources of meaning, but rather to take active steps to change the characteristics of specific life domains to be better aligned to the personal needs, values, goals, and capabilities of the individual (Tims and Bakker, 2010; Demerouti et al., 2020).

Although this crafting approach to meaning-making is becoming increasingly popular within the literature, it is stringently domain-specific and negates the dynamic interaction between life domains (de Jong et al., 2020). For example, when crafting behaviors are applied to the work context (i.e., *Job Crafting*), the aim is to align the employee's personal needs, goals, and skills to the characteristics of the job (Tims and Bakker, 2010). This individually driven work design process provides employees with a means through which to change the nature of the tasks they engage in, change the nature of interactions at work, or modify perceptions one has about the job itself (Tims and Bakker, 2010). This approach, however, ignores the impact of other life domains such as home-life (Petrou and Bakker, 2016). When individuals are unable to experience meaning in one domain of their lives, they are likely to pursue activities in other domains such as the home environment-, leisure, or relationships to compensate (Petrou and Bakker, 2016; Demerouti et al., 2020). As an alternative, Demerouti et al. (2020) suggested that individuals may pursue

"Home Crafting" to compensate for lack of meaning at work. Still viewing crafting as primarily a work-centered activity, Demerouti et al. (2020, p. 1013) defined home crafting as efforts "employees make to balance their home demands and home resources with their personal abilities and needs in order to experience meaning and create or restore their person-environment fit." This (home) domain-specific approach also assumes that "home crafting" is fundamentally different from the behaviors associated with Job Crafting as the application domain differs. However, Demerouti et al. (2020) argue that home crafting shares similarities with job crafting in that individual actively also "seek (home) resources," "seek (home) challenges," and wants to "reduce (home) demands." Therefore, it can be seen that the overarching behavioral approaches to crafting are similar. However, how and where they are applied differs. Again, this approach negates another important life domain: "Leisure." Petrou and Bakker (2016) argued that if meaning cannot be pursued at work, it would be pursued during leisure time or activities. From this perspective, Leisure Crafting refers to "the proactive pursuit of leisure activities targeted at goal setting, human connection, learning, and personal development" (Petrou and Bakker, 2016, p. 508) pursued during one's off-work or leisure time. From this definition, similar elements or behaviors associated with Job and Home Crafting are apparent: reshaping the task, seeking challenges, and seeking relationships. Therefore, it is apparent that people engage in similar behaviors to craft meaning in various domains of their lives.

Given the overlap in these behaviors, we, therefore, argue that crafting behaviors should be seen as a meta-level concept that transcends the confinement to particular or specific life domains. In other words, crafting should be regarded as a process of conscious efforts individuals exert to create meaning in their lives through (a) cognitively (re-) framing how they view life, (b) seeking social support systems to manage life challenges, and (c) actively seeking challenges to facilitate personal growth. Specifically, these behaviors are actioned to better align an individual's life goals, personal needs, values, and overall capabilities. In addition, since these behaviors are directly located to an individual's general needs, values and self-development, it may benefit all social roles individual plays *via* spillover or crossover effects from one domain to another. We call this meta-approach to meaning-making: "Life Crafting."

As such, the present study has two goals: (1) to establish a theoretical framework for life crafting by contrasting and comparing different domain-specific crafting approaches apparent within the literature and (2) to develop and validate a positive psychological assessment measure aimed at measuring overall life crafting. Our study aims to make several contributions. First, contributing to positive psychology literature, we plan to explore how people create meaning in life by establishing the construct of life crafting. Second, we expand job crafting to the whole life domain by testing whether crafting behaviors at work are linked to crafting behaviors in life, contributing to proactivity/job crafting literature. Third, we will provide empirical evidence on the relationship between life crafting, meaning in life, and mental health. All these contributions are possible by introducing a reliable, valid, and flexible tool for empirical research on life crafting.

THE CONCEPTUALIZATION OF LIFE CRAFTING

Life crafting has emerged as a relatively new concept in the literature, with little to no theoretical grounding underpinning its use (c.f., Schippers and Ziegler, 2019; de Jong et al., 2020; Dekker et al., 2020). Only three academic papers explicitly refer to life crafting as a specific strategy aimed to pursue meaning. First, Schippers and Ziegler (2019) viewed life crafting as a process to reflect on life and take action to increase fit among their life, values, and wishes. Second, de Jong et al. (2020) developed a four-stage theoretical life crafting intervention: discovering the values and passions, reflecting on one's ideal life, setting specific goals and plans, and making the public commitment to the goals set. Finally, Dekker et al. (2020) argued that life crafting might enhance the individual's goal pursuit, performance, and mental health. From these approaches, the core premise of life crafting seems to focus on proactive actions individuals take to discover their values/passions, look for challenges, and accumulate resources needed to further their personal growth and development. Although these three papers showed promise, the conceptual construction of life crafting and what it entails is severely lacking. A clear conceptual model for life crafting is needed, highlighting the concept's theoretical foundation (and measurement).

Given that no theoretical model for life crafting exists, we turn to "crafting" in other domains (i.e., job crafting, home crafting, and leisure crafting) and use these as a reference for constructing a conceptual definition and model for life crafting. By reviewing job crafting (Wrzesniewski and Dutton, 2001; Tims and Bakker, 2010), home crafting (Demerouti et al., 2020), and leisure crafting literature (Petrou and Bakker, 2016), we attempt to find the conceptual overlap between the different crafting strategies employed in each, the definitions and structures of crafting in other domains. These are briefly summarized in **Table 1**. This, in turn, would act as a foundation for life crafting.

By contrasting and comparing the different approaches toward crafting in different domains, we found that there are clear conceptual overlap in strategies between job-, home- and leisure crafting: (a) cognitive crafting, (b) relational crafting, (c) resources crafting, (d) challenges crafting, and (e) demands crafting. There are small differences in how similar crafting behaviors manifest from these five overlapping strategies. For example, although relational crafting was an essential component of job crafting, the perspective that the researcher used to explain it is slightly different. Wrzesniewski and Dutton (2001) highlighted how people expand or restrict their social network, whereas Tims and Bakker (2010) were curious about the individual's social resources-seeking behaviors. We, therefore, proceeded to look at the overlap and differences and derived 8 possible crafting strategies from the literature.

Cognitive crafting is the effort of individuals to redefine or reframe their life in such a manner that it provides more meaning.

TABLE 1 | Definitions and structures of crafting in other domains.

	Definitions	Structure
Job crafting	Job crafting is the physical and cognitive changes individuals make in the task or relational boundaries of their work (Wrzesniewski and Dutton, 2001).	Cognitive crafting Relational crafting Task crafting
	The actions employees take to change their levels of job demands and job resources in order to align them with their own abilities and preferences (Tims and Bakker, 2010; Tims et al., 2012).	Increasing structural job resources Increasing social job resources Increasing challenging job demands Decreasing hindering job demands
Home crafting	Changes that employees make to balance their home demands and home resources with their personal abilities and needs, in order to experience meaning and create or restore their person-environment fit (Demerouti et al., 2020)	Seeking resources at home Seeking challenge at home Reducing demands at home
Leisure crafting	The proactive pursuit of leisure activities is targeted at goal setting, human connection, learning, and personal development (Petrou and Bakker, 2016).	Single dimension: leisure crafting

The perception of meaning in life is primarily influenced by how people think about or define their life (Beck, 1995). For example, a psychologist believing his/her work or life serves a broader purpose through mitigating mental pain or stimulating others' flourishing. In this case, they may evaluate that their life is more meaningful because their work contributes to something conceptually larger than themselves (Wrzesniewski et al., 1997). Bruning and Campion (2018) found that cognitive job crafting was positively related to meaning. Berg et al. (2015) proposed three ways to craft individuals' cognition: expanding perceptions (e.g., look for the holistic purpose of job), focusing perceptions (e.g., narrow the mental scope of the purpose for dislike work), and liking perceptions (e.g., connect specific tasks to adoring outcomes). Wellman and Spreitzer (2011) believed that cognitive crafting, such as thinking about best-self, could enhance scholars' joy and meaning. Furthermore, we reviewed the items from the job crafting scales (Slemp and Vella-Brodrick, 2013; Bindl et al., 2019), and conclude that two overall cognitive crafting strategies can be distilled: positive thinking (e.g., proactively looking for the positive aspect of adverse events) and transcending personal goals (e.g., think about how your life contributes to society).

Relational crafting refers to people seeking social support to further pursue personal or life goals. Human beings are social animals, and interaction with others plays a vital role in their daily lives (Rofcanin et al., 2019). Relationship with others is one of the most important sources of meaning (Steger, 2009; Sacco et al., 2014). Individuals may actively choose to spend more time with preferred people and seek assistance when encountering difficulties (Laurence, 2010). Prior studies found that relational crafting can increase meaningfulness, extra-role performance, work engagement and decrease job boredom (Harju et al., 2016; Tims et al., 2016). Berg et al. (2015) proposed three ways to craft individuals' cognition: building relationships (e.g., increasing the amount of interaction), reframing relationships (e.g., thinking about their social environment in different ways), and adapting relationships (e.g., assisting others). Further, we reviewed the items from job crafting (Tims et al., 2012; Slemp and Vella-Brodrick, 2013; Bruning and Campion, 2018; Bindl et al., 2019), and the literature about networking (Porter and Woo, 2015; Wolff and Spurk, 2020). From this review we concluded that three relational crafting strategies are relevant: crafting strategies are relevant: creating new relationships (e.g., try to meet new people), optimizing current relationships (e.g., improving the quality of my interactions with people), and utilizing social resources (e.g., seeking support from family when feeling down).

Resources crafting aims to achieve life goals and fulfill life's potential by increasing or optimizing available resources. For example, individuals could look for more autonomy and seek more technological competence. Resources refer to the life circumstances people value in the pursuit of meaningful goals (Hobfoll, 1989), such as the opportunity for development and autonomy. Demerouti et al. (2001) proposed that resources in the work domain could facilitate the achievement of work goals, reduce job demands and the associated physiological and psychological costs, and stimulate personal growth and development. It can also promote essential outcomes, such as job performance (Demerouti et al., 2010) and wellbeing (Nielsen et al., 2017). Moreover, previous studies found that resources are an essential source of positive meaning (Clausen and Borg, 2011; Sacco et al., 2014). Thus, we argue that seeking life resources is a valuable crafting strategy for individuals to increase meaningful life experiences.

Challenge crafting is the proactive behavior that aims to help individuals experience personal growth, achievement, and accomplishment. Examples are working hard on challenging activities and seeking a new challenge in life. Lepine et al. (2005) claimed that not all job demands are related to adverse outcomes. Some types of job demands may lead to positive results, such as personal growth and positive emotions. They coined those job demands as challenging demands. Challenging demands can be viewed as a barrier in life that can be overcome with effort. If people do so, they may experience a sense of personal accomplishment. Previous studies have shown that seeking challenges increases work engagement (Petrou et al., 2012), academic performance (Ingusci et al., 2020), and job performance (Petrou et al., 2015). We argue that seeking challenging life demands will also promote a positive self-image and create meaning in life.

Demands crafting aims to reduce hindering life demands and avoid excessive resource loss more effectively. Examples

are avoiding intense mental work and using one's strengths to achieve life goals (Richter et al., 2021). Demands typically cost resources, but unlike the challenge demands we mentioned above, some demands will decrease individuals' motivation and engagement (Lepine et al., 2005; Tims et al., 2012). Once this type of demand exceeds one's capability, it will deplete mental health and wellbeing (Tims and Bakker, 2010). Although life demands cannot be avoided, we can interpret and deal with them positively or effectively, such as simplifying the work processes to make them more efficient (Demerouti and Peeters, 2018). This strategy may free people from intensive life demands and avoid stress and burnout. Notably, in previous studies, demands crafting sometimes had a negative or no relationship with other dimensions (Tims et al., 2012; Lichtenthaler and Fischbach, 2019). It may be because demands crafting was driven by avoiding motivation (Petrou et al., 2012).

Overall, the eight dimensions we included from the literature were: positive thinking, transcending personal goals, creating new relationships, optimizing current relationships, utilizing social resources, resources crafting, challenges crafting, and demands crafting.

Nomological Network of Life Crafting

Life crafting can be embedded in the broader meaning-making literature based on several key attributes. First, life crafting is a general type of crafting, so it should be similar to crafting in other life domains. Second, life crafting highlights how people self-initiate and deliberately create meaning in life, which might overlap with proactive behaviors. Third, life crafting happens not only after adverse life events but also throughout everyday life when people proactively try to create meaning in life. Therefore, to further clarify the concept of life crafting, we will compare life crafting to job crafting, proactive personality and coping, and meaning-making theory (i.e., Meaning Maintenance Model and Global-situational Meaning-making Model).

Life Crafting and Job Crafting

Job crafting can be interpreted as life crafting applied to one specific domain in life. In the literature, job crafting refers to the actions employees proactively take to make their own job more meaningful, engaging, and satisfying (Wrzesniewski and Dutton, 2001; Tims and Bakker, 2010; Wrzesniewski et al., 2013). Wrzesniewski et al. (2013) proposed that job crafting is an effective strategy that helps individuals gain positive work meaning from specific sources, such as themself, their social circumstance, work context, and spirituality. Tims et al. (2016) also found that job crafting could increase meaningfulness by promoting person-job fit. Compared to job crafting, we argue that life crafting is a more general and holistic concept because life meaning could be drawn from multiple sources (Steger and Dik, 2009), instead of from a single source (e.g., the work domain). For example, Bronfenbrenner (1992) posited that meaning in life is constructed around various domains (microsystem) and the interactions of these domains (mesosystem). Thus, we conceptualized life crafting as a domain-unspecific concept.

Life Crafting and Proactive Personality

The proactive personality is a behavioral disposition toward taking action or changing one's environment (Bateman and Crant, 1993). The core premise of the proactive personality is that people, behaviors, and environment have lastingly influenced each other. Therefore, people can proactively change the environment they live in Fuller and Marler (2009). Prior studies found that people with proactive personalities were more likely to attach great value to their job and precept high-level work meaning (Akgunduz et al., 2018, 2020). In line with proactive personality, life crafting emphasizes people's initiative to change or shape their external environment. For instance, how people proactively change their environment to make it more challenging and appealing. As such, we argue that life crafting also includes a cognitive component, whereby individuals can actively change their views about life. In addition, compared to proactive personality, life crafting, as a behavior, is easier to emerge, change or enhance. Therefore, individuals could always look for the chance to craft their daily life.

Life Crafting and Proactive Coping

Proactive coping refers to individuals building up available resources in order to achieve challenging goals and personal growth (Schwarzer and Knoll, 2003). The basic proposition of proactive coping is that people view demands as a challenge to promotion instead of threatening resource loss (Hambrick and McCord, 2010). Greenglass and Fiksenbaum (2009) found that proactive coping could improve positive affect and psychological functioning. Similar to life crafting, proactive coping also stresses individuals' initiative to control the situation and to seek challenges. However, the ultimate purpose of proactive coping is to handle a situation successfully or transform the potential threatens into opportunities, whereas life crafting aims to increase an individual's positive meaning in life.

Life Crafting and Reactive Meaning-Making Theory

The mechanisms of meaning-making have been studied for decades. Two prevalent theories in the field were the Meaning Maintenance Model (Heine et al., 2006) and the Global-situational Meaning-making Model (Park, 2010). The Meaning Maintenance Model's core assumption is that people tend to reaffirm alternative frameworks while experiencing meaningless or meaningful disruption. In comparison, the Global-situational Meaning-making model is used to explain how individuals' global meaning (e.g., beliefs) interact with situational meaning (e.g., a meaningless context). They were devoted to exploring how people respond and recover from meaningless situations or mental trauma. Conversely, life crafting emphasizes people's initiative or proactive efforts to search for meaning, it assumes that motivation for living worth will lastingly force people to pursue a better and meaningful life, instead of just when bad things happen. Moreover, the former two meaning-making theories mainly underlined reflective- or cognitive exercises. However, life crafting provided a practice-friendly framework that values action than exposed facto reflection in traumatic events. This distinguishes life crafting from other recover-oriental meaning-making theories.

Consequences of Life Crafting

We propose that life crafting leads to many positive outcomes for individuals. Firstly, life crafting is a strategy people use to increase meaning in life. Thus, more life crafting should be related to more meaning in life. Secondly, life crafting may lead to higher levels of mental health. Crafting is driven by individuals' own needs (De Bloom et al., 2020). Therefore, people who craft their lives more are also more likely to experience life satisfaction and positive affect. This reasoning is in line with Dekker et al. (2020) view that life crafting is an essential strategy to improve and maintain overall mental health. Thirdly, we propose that life crafting is positively related to work-related variables (i.e., work engagement and job burnout). On the one hand, life crafting has similar effects or relationships to job crafting in the work domain. Ample studies have shown that job crafting will enhance work engagement and reduce job burnout (Bakker and Costa, 2014; Tims et al., 2015; Harju et al., 2016). On the other hand, life crafting in the non-work domain may also enhance work engagement and reduce job burnout through crossover or compensation effects (De Bloom et al., 2020). For instance, Abdel Hadi et al. (2021) found that leisure crafting behaviors were negatively related to employees' emotional exhaustion and mitigated the undermining effect of job and home demands on the emotional exhaustion.

THE CURRENT STUDY

Given the potential benefits of life crafting, the current study aimed to conceptualize life crafting and to develop, validate and evaluate a robust measure of overall life crafting.

MATERIALS AND METHODS

Research Approach

A mixed-method, multi-study design was employed to develop and validate the Life Crafting scale. In the preliminary study, we reviewed prior research and interviewed people to create the item pool. In study 1, we performed a cross-sectional study to explore the construct of life crafting and its underlying factors. In study 2, we ran a cross-sectional study to confirm and validate the factor structure of life crafting found in study 1.

Preliminary Study: Item Generation

The current study developed potential items through both deductive (e.g., literature review) and inductive (e.g., interview) techniques. Firstly, we collected insights from the literature on job crafting, home crafting, leisure crafting, and meaning in life. After this, we retrieved items from the Job Crafting Scale (Bindl et al., 2019), JD-R based Job Crafting Scale (Tims et al., 2012), Role–Resource Job Crafting Measure (Bruning and Campion, 2018), and the Job Crafting Questionnaire (Slemp and Vella-Brodrick, 2013). To make such items fit the life domain, we also adopted and reframed some of them. Secondly, since all of the published theoretical work on crafting behaviors are based on Western cultures, we interviewed nine participants from China (four men and five women) to explore whether there were differences in the crafting strategies these people employ. These findings were used to supplement our initial item pool. Convenient sampling was used to select interviewees. The interviewees included a painter, a novelist, two college counselors, an HRM practitioner, and four secondary school teachers. Their ages ranged from 26 to 41, and three of them were parents. We created an interview protocol based on the literature review's findings, and in-depth semi-structured interviews were employed. We first presented and explained our definition of life crafting. Following this, we asked participants a set of open-ended questions about their approaches to creating meaning and how they experienced this. All interviewees were interviewed in Mandarin. The first author translated the interview manuscripts into English and all authors coded the materials together in English. We used content analyses (Hsieh and Shannon, 2005) as a means to process the data with the qualitative data analysis program Nvivo11. See **Table 2** for the interview questions, coded answers, and

TABLE 2 | Examples of life crafting techniques.

Questions	Typical thoughts and behaviors	Typical illustrative quotations
Can you recall specific examples of when you sought meaning by reinterpreting or reflecting on work, family, or life events?	1. Seek support from family; 2. Seek advice from others; 3. Share my life with friends or family; 4. Expand my social network.	*"I like to expand my social network, especially meet senior leaders. Because if you have a good relationship with them, it will be easier for you to deal with the problem at work."*
Can you recall specific examples of when you sought meaning by expanding/limiting your social network or seeking support from your social network?	1. Recognize me; 2. Think about the influence on others; 3. Find a balance between life and dream; 4. Think about the influence on others.	*"I (a teacher) recognized that my actions or words might influence others. Most of my students can solve the problems by themself. So I begin to trust my students, give them positive feedback, and mentor them."*
Can you recall specific examples of when you sought meaning by challenging yourself or fitting you and your life?	1. Learn new skills; 2. Take control of your life; 3. Look for the chance to challenge yourself; 4. Expand hobbies; 5. Take extra works.	*"I used to be scared of public speaking. I sometimes behaved in such a way that I seemed to deserve less respect. However, as a teacher, you cannot avoid public speaking. I finally find that if I am in charge of the topic, I can control my audience. So I train myself to be more dominant in large meetings. Finally, I am not nervous anymore when I have to speak to groups."*

the typical illustrative quotations. Following up on a reviewer's suggestion, we also interviewed five European people to check if these strategies were consistent with those that individuals from western cultural backgrounds exhibit. The results showed that there was considerable overlap between the original findings from the Chinese sample interviews and those from Europe.

In step three, we created a life crafting item pool with 64 items based on our literature review (step 1) and interviews (step 2). By speculating the content of items, the first author independently classified these 64 items into eight theoretical dimensions in Round 1. There are six items for creating new relationships, seven items for optimizing current relationships, utilizing social resources, and challenges crafting, respectively, eight items for transcending personal goals, nine items for positive thinking and resources crafting respectively, and 11 items for the demands crafting. After this the other three co-authors checked the definition and category of each dimension. In Round 3, we invited a panel of experts to review our items pool to assess content validity. The panel consisted of 5 psychologists who had specifically researched crafting behaviors or meaning in life. These five experts were asked to assess the consistency between the definitions and items on a 5-point Likert scale from 1 (not representative of the concept definition) to 5 (very representative of the concept definition). We first checked the interrater reliability by calculating Cohen's (1960) kappa values. Cohen's kappa values ranged from −1 to 1, and values ≤ 0 as indicating no agreement, 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement (McHugh, 2012). Based on these criteria, we removed 18 items for which Cohen's (1960) kappa values were ≤ 0. We then computed the mean score of the experts' grades on the remaining 46 items and kept 33 items who got a three or higher mean score. Considering the experts' comments, we also removed four items that were regarded as redundant. Moreover, two items received a low score on representativeness (2.8). We, therefore, decided to rephrase these items as we did believe the items were relevant to life crafting. Eventually, an item pool with 31 items (eight dimensions) was established and used in the follow-up studies.

Study 1 Exploratory Factor Analysis and Reliability

Methods

The purpose of study 1 is to develop and examine a generic scale that can be used to measure life crafting. An exploratory factor analysis was conducted to screen the items and explore the structure of life crafting. Additionally, we tested the reliability of the life crafting scale with Cronbach's alpha and Composite Reliability.

Participants

A convenience sampling strategy was used to collect data for the study. Inclusion criteria were that participants had to be 18 years or older, English-speaking, and currently employed. Three hundred eighty-five people responded to our questionnaire, and 86% of them completed all questions. In total, 331 people participated in Study 1. See **Table 3** for the participants' gender, age, marital status, employment status, and whether they had children. Almost half of the sample were women (42.9%). The participants' ages ranged from 18 to 71, and the average age was 27.55 (*SD* = 9.85). 50.8% of participants were single, whereas 32.6% of them were in a relationship. The majority of the participants did not have children (81.0%). 44.4% of the participants worked for an organization, and 9.4% were self-employed. The most of participants were recruited from the UK and other European countries, such as Portugal, Poland, and the Netherlands.

Procedure

The participants were recruited through Prolific and electronic surveys were administered through Qualtrics. The electronic questionnaire consisted of questions relating to participants' demographic information and the 31 life crafting items. The Ethical Review Board at the Eindhoven University of Technology approved this study. This study was registered under this code: ERB2020IEISSHI20. Inclusion criteria were that participants had to be 18 years or older, English-speaking, and employed.

Measures

Life Crafting

Participants answered each of the 31 life crafting items, stemming from the original eight dimentions, in respect to how frequently they engaged in each of the mentioned behaviours. Each item was rated on a five-point Likert scale (1 = never, 2 = sometimes, 3 = regularly, 4 = often, 5 = always). Example items were 'I change the way I think about challenges to make myself feel more positive about them' (positive thinking), "I think about how my life contributes to society" (transcending personal goals), "I try to meet new people" (creating new relationships), "I spend more time with people who give me energy" (optimizing current relationships), "I use my social network to more effectively achieve my life goals" (utilizing social resources), "I try to learn new things" (resources crafting), "I undertake or seek extra tasks to expand my vision" (challenges crafting), and "I structure my tasks to achieve my goals" (demands crafting).

Statistical Analyses

To explore the factorial structure of the life crafting scale, we performed an exploratory factor analysis (EFA) with SPSS 25.0. First, the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's sphericity test were used to determine factorability. A KMO value larger than 0.60 and a statistically significant chi-square value on Barlett's test of sphericity would indicate that the data are factorable (Dziuban and Shirkey, 1974). Thereafter, we determined the multivariate normality of the data by reviewing the absolute ranges for skewness and kurtosis. According to George and Mallery (2019), a skewness/kurtosis range between $\pm$ 2.0 indicates that the data is relatively normally distributed.

Second, an EFA was conducted with varimax rotation to extract factors. Since there are eight factors in our hypotheses, we first extracted eight factors to check the quality of the items. We only retained factors with an eigenvalue of at least 1, and the total combined explained variance of all the retained factors was set to at least explain 50% of the overall variance (Carpenter, 2018; Youssef-Morgan et al., 2022). Furthermore, we removed

TABLE 3 | Demographic and biographic characteristics.

		Study 1		Study 2		Study1 vs. Study 2
Item	**Category**	**Frequency**	**Percentage (%)**	**Frequency**	**Percentage (%)**	***p*-value**
Gender	Male	189	57.1	188	51.9	0.01
	Female	142	42.9	173	47.8	
	Other	0	0	1	0.3	
Age (years)	18∼30	245	74.0	166	45.9	0.01
	31∼45	65	19.7	46	12.7	
	46∼	21	6.3	150	41.4	
Marital status	Single	168	50.8			
	Married or in a relationship	157	47.4			
	Divorced	5	14.8			
	Widowed	1	0.3			
Have children	Yes	63	19.0			
	No	268	81.0			
Employment status	Work for an organization/company	147	44.4	317	87.6	0.01
	Self-employed	31	9.4	37	10.2	
	Other	153	46.2	8	2.2	

those items with a factor loading and commonality smaller than 0.4 or when they loaded more than 0.4 on more than one factor (Carpenter, 2018).

Finally, Corrected item-total correlation (CITC), Cronbach's alpha, and Composite Reliability (CR) were used to examine reliability. CITC is the correlation of the designated item with the sum of other items, and the value of CITC for each item should be above 0.3 (Field, 2013; George and Mallery, 2019). Alpha is the lower bound, and CR is the upper bound of the internal consistency. Their values should be all above 0.7 (Hair et al., 1998). We ran a confirmatory analysis to get the standardized factor loading with the first sample to calculate CR with Mplus 8.0.

Results Study 1

Exploratory Factor Analysis and Reliability

Since we theorized that life crafting is a model with eight first-order factors, we first fixed the number of factors to eight and then used principal factor analysis with a varimax rotation. The eight-factor model was factorable as the KMO was 0.93, and Barlett's test indicated sphericity. The eight factors explained 49.10% of the overall variance. After that, we removed 17 items because either their loading on one of the factors was smaller than 0.4 (11 items), their commonality was smaller than 0.4 (2 items), or when an item had a high loading on more than one factor (4 items). Following Carpenter's (2018) suggestion, we deleted the factors that included less than three items. One two-item factor, two one-item factors, and one zero-item factor (totalling four items) were deleted based on this, resulting in a four-factor model with 16 items.

Consequently, we ran a principal axis factor analysis with varimax rotation and screened the items again. The results indicated that only three factors had an eigenvalue larger than 1. Therefore, we removed the factor that got an eigenvalue smaller than 1 (3 items). Finally, we ended up with a three-factor model with nine items. The three factors explained 53.87% of the overall variance. Results showed that meaningful factors could be extracted from the data because the KMO value was larger than 0.60 (KMO = 0.82) and a significant chi-square [$\chi^2_{(331)} = 1139.81$, $df = 36$, $p < 0.001$] was produced. The mean, SD, CITC, and factor loading for each item were reported in **Table 4**. We reported the Cronbach's alpha, CR, and correlations among three factors in **Table 5**. We found that these three factors were identical to the factors cognitive crafting, relational crafting, and challenges crafting of the initial eight factors. Moreover, by inspecting the contents of the nine remaining items, we labeled the three factors: cognitive crafting, seeking social support, and seeking challenges.

Study 2. Confirmatory Factor Analysis, Validation, and Measurement Invariance

Methods

In Study 2, we investigated if the three-factor structure found in study 1 could also be confirmed in a different sample. To this end, we performed a second study with confirmatory factor analysis. We compared the model fit between the three-factor model and other alternative models: the one-factor model, two-factor model, second-order factor model, and Bi factor model. Second, we examined the second-order factor model's convergent validity and discriminant validity. There is no other life crafting scale yet, so we tested convergent validity and discriminant validity by comparing life crafting to similar concepts: job crafting (Tims et al., 2012) and proactive personality (Bateman and Crant, 1999). Furthermore, we tested measurement invariance across gender. Finally, we examined criteria validation by computing the standardized regression values among life crafting, job burnout, work engagement, meaning in life, and mental health.

TABLE 4 | Item level descriptive statistics and factor loading.

Items	Mean	SD	Skewness	Kurtosis	CITC	Factor loading		
						CC	SSS	SC
Cognitive crafting								
LF 10. I think about how my life helps others	3.04	1.24	0.07	−1.09	0.63	**0.61**	0.34	0.19
LF 42. I think about how my actions positively impact my community	2.78	1.19	0.22	−0.96	0.71	**0.79**	0.18	0.27
LF 44. I think about how my life contributes to society	2.86	1.20	0.19	−1.04	0.69	**0.75**	0.10	0.25
Seeking social support								
LF 12. I actively ask people for advice when I encounter difficulties	3.11	1.23	0.07	−1.11	0.56	0.12	**0.68**	0.10
LF 25. I seek support from my family when I am down	2.92	1.37	0.13	−1.27	0.55	0.20	**0.61**	0.17
LF 40. I am willing to ask others for help when things become too difficult to bear	3.09	1.17	0.07	−1.12	0.63	0.11	**0.78**	0.10
Seeking challenges								
LF 27. I try to work hard on challenging activities	3.33	1.03	−0.04	0.13	0.58	0.10	0.13	**0.65**
LF 41. I change my activities so that they are more challenging	2.40	1.00	0.67	0.13	0.64	0.34	0.12	**0.67**
LF 43. I seek out opportunities that challenge my skills and abilities	2.95	1.11	0.10	−0.08	0.69	0.34	0.17	**0.76**

CITC, Corrected item total correlation; λ, Standardized factor loadings; CC, cognitive crafting; SSS, seeking social support; SC, seeking challenges. Bold: Significant item loadings ($p < 0.01$).

TABLE 5 | Factor correlations and internal consistencies of life crafting.

No	Factor	CR	Cronbach's alpha	1	2
1	Cognitive crafting	0.83	0.82	–	
2	seeking social support	0.76	0.75	0.25*	–
3	Seeking challenges	0.79	0.79	0.27*	0.19*

$^*p < 0.01$.

Participants

Four hundred thirty-one employees participated in Study 2, and 78% of participants filled all questionnaires. The final sample consisted of 362 participants after we deleted unfinished responses. Almost all the participants were Dutch, and their gender, age, and employment status are summarized in **Table 3**. Almost half of the participants were women (47.8%), and the average age was 38.60 (SD = 14.14). Most of the sample was working for an organization (87.6%). The average workload of the participants was 35.81 h (SD = 9.13) per week. We compared the biographic characteristics between participants of Study 1 and Study 2 *via* t-test and chi-square test, (c.f. **Table 3**).

Procedure

For Study 2, participants were recruited by students who participated in a Master's course in Performance Management at the Eindhoven University of Technology in the Netherlands. Each student recruited approximately 5 participants from their social network (e.g., parents and friends), and the students then used the collected data for their assignment. Participants had to be over the age of 18 and had to work a minimum of 3 days per week. Questionnaires were administered through Qualtrics. Because we conducted the study in the Netherlands, we also asked them to report their English level with a 7-point Likert scaling ranging from 1 (Not Sufficient) to 7 (Sufficient). Eight participants who reported an English level below three were removed from the analyses. The Ethical Review Board at the Eindhoven University of Technology approved this study, and this study was registered under this code: ERB2020IEIS20.

Measures

The Following Scales Were Administered in Study 2:

Life crafting. The life crafting scale developed in study 1 was used to measure life crafting. The nine-item scale was rated on a five-point Likert-type agreement scale ranging from 1 (Never) to 5 (Always). The scale consisted of nine items and comprised four subscales: cognitive crafting (3 items), seeking social support, and seeking challenges (3 items). Example items were "I think about how my life helps others" (cognitive crafting), "I actively ask people for advice when I encounter difficulties" (seeking social support), and "I try to work hard on challenging activities" (seeking challenges). The Cronbach's α of the three subscales were 0.82, 0.75, and 0.79. The final set of items is presented in **Appendix Table 1**.

Job crafting. We adopted items from the daily job crafting scale (Petrou et al., 2012) to measure job crafting. The scale consisted of thirteen items and comprised four subscales: seeking job resources (5 items), seeking challenges (4 items), and reducing demands (4 items). Responses were given on a 5-point scale with 1 (Never) – 5 (Always). Example items were "I try to learn new things at work' (seeking job resources), "I ask for more tasks if I finish my work" (seeking challenges), and "I try to ensure that my work is emotionally less intense" (reducing demands), The Cronbach's α of the three subscales in this study were 0.85, 0.82, and 0.76, respectively.

Meaning in Life. The Meaning in Life Questionnaire developed by Steger et al. (2006) was used to measure meaning in life. The ten-item questionnaire is rated on a seven-point Likert-type scale ranging from 1 (absolutely untrue) to 7 (absolutely true). It measures the two components of meaning in life with five items each. Example items are "I understand my life's meaning" (presence of meaning), and "I am always looking to find my life's purpose" (search for meaning). The Cronbach's α of the two subscales were both 0.93.

Proactive personality. The six-item short version of the Proactive Personality Scale (Bateman and Crant, 1993) was used to measure proactive personality. This 6-item short version was validated by Claes et al. (2005). Ratings were made on a 5-point scale that ranged from 1 (Totally disagree) to 5 (Totally agree). Cronbach's α in this study was 0.79. Example items were "If I see something I don't like, I fix it." and "I excel at identifying opportunities."

Job burnout. The Oldenburg Burnout Inventory (Demerouti et al., 2003) was used to measure job burnout. The sixteen-item scale was rated on a four-point Likert-type scale ranging from 1 (Strongly disagree) to 4 (Strongly agree). The scale consists of two components of job burnout with eight items each. Example items were "It happens more and more often that I talk about my work in a negative way" (disengagement) and "There are days when I feel tired before I arrive at work" (exhaustion). Half of the items were reversed coded. The Cronbach's α of the two subscales in this study were 0.73 and 0.78, respectively.

Work engagement. The 9-item version of the Utrecht Work Engagement Scale (Schaufeli et al., 2006) was used to measure work engagement. The nine-item scale was rated on a seven-point Likert-type scale ranging from 0 (Never) to 6 (Always). It measured the three components of work engagement with three items each. Example items were "At my work, I feel bursting with energy (vigor)," "I am enthusiastic about my job" (dedication), and "I get carried away when I am working" (absorption). The Cronbach's α of the three subscales in this study were 0.82, 0.88, and 0.67, respectively.

Mental health. The Mental Health Continuum-Short Form validated by Lamers et al. (2011) was used to measure mental health. The form consists of fourteen items that were derived from Midlife Development in the United States (Keyes, 2002). Respondents rated the frequency of every feeling in the past month on a 6-point Likert scale from 1 (never) to 7 (every day). It measures the three components of mental health with three items (emotional wellbeing), five items (psychological wellbeing), and six items (social wellbeing), respectively. Example items are "In the past month, how often did you feel happy," "In the past month, how often did you feel that you liked most parts of your personality," and "In the past month, how often did you feel that our society is becoming a better place for people." The instrument showed to be a reliable measure in other contexts with McDonald omegas ranging from 0.76 to 0.92 on the various subscales (Van Zyl and Ten Klooster, 2022). The Cronbach's a of the three subscales in this study was 0.89 for emotional wellbeing, 0.87 for psychological wellbeing, and 0.78 for social wellbeing.

Statistical Analyses

First, we estimated factorial validity by conducting a confirmatory factor analysis in Mplus v 8.0 (Muthén and Muthén, 2020). The parameters were calculated through the maximum likelihood estimation method. Several fit indices, which we illustrated in **Table 6**, were used to evaluate model fit. We also calculated factor loadings, item-level statistics, and internal consistency to investigate the three-factor life crafting model and higher-order life crafting model in SPSS and Mplus.

Second, we investigated measurement invariance or factor equivalence across gender by computing and comparing configural- (similar factor structures), metric- (similar factor loadings), and scalar (similar intercepts) models in Mplus. Invariance was determined through a non-significant difference in chi-square, CFI ($\Delta < 0.01$), TLI (< 0.01), RMSEA ($\Delta < 0.015$), and SRMR ($\Delta < 0.015$) (Cheung and Rensvold, 2002). If full invariance could not be established, partial invariance would be pursued by releasing some constraints on the various models (van de Schoot et al., 2012; Van Zyl and Ten Klooster, 2022).

Third, to investigate convergent validity, we created a structural equation model in which we regressed subfactors of job crafting on life crafting and life crafting on proactive personality. Additionally, we assessed discriminant validity following (Rönkkö and Cho, 2020) approach. We calculated the confidence intervals of the correlation between life crafting and similar concepts [i.e., seek resources, seek (job) challenges, reduce demands, and proactive personality], then investigated whether the upper limit values of confidence intervals were smaller than 0.90.

Finally, the same method to establish convergent validity was applied to establish concurrent validation. Using a structural equation model, we determined the relationship between life crafting and related theoretical variables (i.e., job burnout, work engagement, mental health, and meaning in life).

Result Study 2

The confirmatory factor analyses, convergent validity, discriminant validity, measurement invariance, and criteria validity are reported in this section. The results are presented in the tabulated format with brief subsequent interpretations.

Confirmatory Factor Analysis

We employed a competing measurement modeling strategy to establish the factorial validity of the life crafting scale. The following models were estimated:

a) Model 1: All items load on a single factor.
b) Model 2a: Items load on two factors: factor 1 (cognitive crafting + seeking social resources) and factor 2 (seeking challenges).
c) Model 2b: Items load on two factors: factor 1 (cognitive crafting + seeking challenges) and factor 2 (seeking social resources).

TABLE 6 | Model fit indices.

Fit indices	Cut-off criterion
Absolute fit indices	
Chi-Square (χ^2)	• Lowest comparative value between measurement models • Non-significant chi-square ($p > 0.01$) • Significant difference in chi-square between models • For model comparison: retain model with lowest chi-square
Approximate fit indices	
Root-means-square error of approximation (RMSEA)	• 0.06–0.08 (marginally acceptable); 0.01–0.05 (excellent) • Not-significant ($p > 0.01$) • 90% Confidence interval range should not include zero • For model comparison: retain model where ΔRMSEA ≤ 0.015
Standardized root mean square residual (SRMR)	• 0.06–0.08 (marginally acceptable); 0.01–0.05 (excellent) • For model comparison: retain model where ΔSRMR ≤ 0.015
Incremental fit indices	
Comparative fit index (CFI)	• 0.90–0.95 (marginally acceptable fit); 0.96–0.99 (excellent) • For model comparison: retain model with highest CFI value (ΔCFI > 0.01)
Tucker-Lewis index (TLI)	• 0.90–0.95 (marginally acceptable fit); 0.96–0.99 (excellent) • For model comparison: retain model with highest TLI value (ΔTLI > 0.01)
Akaike information criterion (AIC)	• Lowest value in comparative measurement models
Bayes information criterion (BIC)	• Lowest value in comparative measurement models
Sample-size adjusted BIC (aBIC)	• Lowest value in comparative measurement models

These indices and criteria were adapted from Wong and Wong (2020).

TABLE 7 | Confirmatory factor analysis.

Model	χ^2	*df*	χ^2/df	CFI	TLI	RMSEA		SRMR	AIC	BIC	aBIC
Model 1	300.87	27	11.14	0.70	0.59	0.17	[0.151–0.185]	0.09	8777.59	8882.67	8797.01
Model 2a	169.85	26	6.53	0.84	0.78	0.12	[0.106–0.142]	0.07	8648.57	8757.54	8668.71
Model 2b	171.05	26	6.58	0.84	0.78	0.12	[0.107–0.142]	0.07	8649.78	8758.74	8669.91
Model 2c	188.46	26	7.25	0.82	0.75	0.13	[0.114–0.149]	0.08	8667.18	8776.15	8687.32
Model 3	29.67	24	1.24	0.99	0.99	0.03	[0.000–0.052]	0.02	8512.40	8629.14	8533.97
Model 4	29.67	24	1.24	0.99	0.99	0.03	[0.000–0.052]	0.02	8512.40	8629.14	8533.97
Model 5	23.76	18	1.32	0.99	0.99	0.03	[0.000–0.059]	0.02	8521.90	8662.00	8547.79

χ^2, Chi-square; df, degrees of freedom; TLI, Tucker-Lewis Index; CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation [90%CI]; SRMR, Standardized Root Mean Square Residual; AIC, Akaike Information Criterion; BIC, Bayes Information Criterion; aBIC, Adjusted Bayes Information Criterion.

d) Model 2c: Items load on two factors: factor 1 (seeking social resources + seeking challenges) and factor 2 (cognitive crafting).
e) Model 3: Items load on three factors: cognitive crafting, seeking social resources, and seeking challenges.
f) Model 4: Second-order factor model: Items load on three factors: cognitive crafting, seeking social resources, and seeking challenges. Moreover, these three factors load on a second-order factor: life crafting.
g) Model 5: Bi factor model, items load on a general factor life crafting and three specific factors: cognitive crafting, seeking social resources, and seeking challenges.

Table 7 presents the model fit indices for each of the six estimated models. The results showed that Model 3, Model 4, and Model 5 had a good fit and were significantly better than Model 1 and Model 2a–2c. Furthermore, Model 3 and Model 4 have identical model fits because we only have three first-level factors (Wong and Wong, 2020), and there were no statistically significant difference when compared to Model 5 ($\Delta\chi^2 = 5.91$, $\Delta df = 6$, $p = 0.43$). In addition, we examined the measurement quality (**Table 8**) of Model 3 and Model 4. Both model 3 and 4 showed acceptable standardized factor loadings ($\lambda > 0.40$, $p < 0.001$), standard errors, and item uniqueness ($0.10 > \delta < 0.90$, $p < 0.001$) (Asparouhov and Muthén, 2009). A defect of the measurement quality is that the AVE for Seeking social support was slightly less than 0.5. However, according to Fornell and Larcker (1981), AVE is a more conservative measure indicator than CR. If the CR value is adequate, the results should be accepted. Although model 3 and Model 5 had a good model fit and measurement quality, we argue that the second-order factor could account for the variation among the first-order factors. Therefore, we used model 4 (Second-order factor model) for the subsequent tests of measurement quality (c.f. **Figure 1**).

TABLE 8 | Item level descriptive statistics, standardized factor loadings, average value explained, and internal consistency for the second-order model.

Items	x̄	σ	Skewness	Kurtosis	CITC	λ	SE	δ	AVE	CR	α
Cognitive crafting									0.53	0.77	0.77
I think about how my life helps others	3.09	1.04	0.08	−0.79	0.55	0.63	0.04	0.61			
I think about how my actions positively impact my community	2.86	0.99	0.27	−0.59	0.63	0.77	0.03	0.41			
I think about how my life contributes to society	2.75	1.01	0.19	−0.85	0.60	0.78	0.03	0.39			
Seeking social support									0.45	0.71	0.69
I actively ask people for advice when I encounter difficulties	3.13	1.02	0.20	−0.95	0.56	0.77	0.04	0.40			
I seek support from my family when I am down	3.09	1.23	0.15	−1.07	0.42	0.50	0.05	0.75			
I am willing to ask others for help when things become too difficult to bear	3.17	1.04	0.17	−1.05	0.53	0.72	0.04	0.49			
Seeking challenges									0.52	0.76	0.75
I try to work hard on challenging activities	3.46	0.98	−0.22	−0.66	0.53	0.60	0.04	0.64			
I change my activities so that they are more challenging	2.54	0.88	0.42	−0.34	0.57	0.68	0.04	0.54			
I seek out opportunities that challenge my skills and abilities	3.02	0.96	0.45	−0.70	0.64	0.85	0.03	0.28			
Life crafting											
Cognitive crafting						0.73	0.07	0.47			
Seeking social support						0.57	0.06	0.68			
Seeking challenges						0.78	0.07	0.36			

x̄, Mean; σ, Standard deviation; CITC, Corrected item total correlation; λ, Standardized factor loadings; SE., Standard Error; δ, Item Uniqueness; AVE, average value explained; CR, Composite Reliability; α, Cronbach's Alpha.

Measurement Invariance

Measurement invariance across genders (males: 188 vs. females: 173) was computed for the second-order CFA model, and the results were reported in **Table 9**. The results showed that all invariance models fitted the data based on the criteria mentioned in **Table 6**. After that, the χ^2 difference test suggested no significant difference between the configural invariance model, the first-order metric invariance model, and the second-order metric invariance model could be found. Moreover, the difference in CFI and TLI did not exceed 0.01, RSMEA and SRMR did not exceed 0.015 for these model comparisons. Thus, we provided evidence for the configural and metric invariance.

However, the imposed scalar invariance model showed a substantial deterioration in model fit in terms of the χ^2, CFI, TLI, RSMEA, and SRMR for the second-order factor model. For this reason, the full second-order scalar invariance had to be rejected. By checking the model comparison results, we found that the misfit between the first-order scalar model and the second-order scalar model may attribute to unequal intercepts of seeking social support. The estimated intercept of seeking social support was 3.16 in the female group, whereas the intercept in the male group was 3.01. After we freed the interception in model 6 (Partial Invariance Model), the model fit improved, and there was no difference in the χ^2, CFI, TLI, RSMEA, and SRMR compared to model 4. Thus, partial invariance was established, and all subsequent models were based on this partial model. Finally, we set all errors in the first and second-factor levels to be equal across genders. The results showed no difference in the χ^2, CFI, TLI, RSMEA, and SRMR. Therefore, strict invariance was established.

Convergent Validity and Discriminant Validity

The relationships between the life crafting and similar concepts (i.e., seeking job resources, seeking job challenges, reducing job demands, and proactive personality) can be found in **Table 10**. The structural model of these variables showed adequate fit [$\chi^2_{(331)}$ = 600.80, df = 340, χ^2/df = 1.77, CFI = 0.93, TLI = 0.92, RMSEA = 0.46 (0.040, 0.052), SRMR = 0.06, AIC = 25116.47, BIC = 25482.29]. First, life crafting was directly associated with seeking job resources (β = 0.69, $p < 0.001$, R^2 = 0.47) and seeking job challenges (β = 0.58, $p < 0.001$, R^2 = 0.33). Second, there was no significant relationship between life crafting and reducing job demands (β = 0.06, p = 0.34, R^2 = 0.00). Finally, proactive personality positively predicted life crafting (β = 0.46, $p < 0.001$, R^2 = 0.21).

Table 11 presents the correlations among the three life crafting subfactors with the three job crafting scales and proactive personality. The upper limits ranged from 0.11 to 0.69. According to Rönkkö and Cho (2020), the upper limits of confidence intervals should be smaller than 0.9. Therefore, discriminant was established.

Criterion Validity

To establish criterion validity, we examined the relationship between life crafting and job burnout, work engagement, mental health, and meaning in life. The results are summarized in

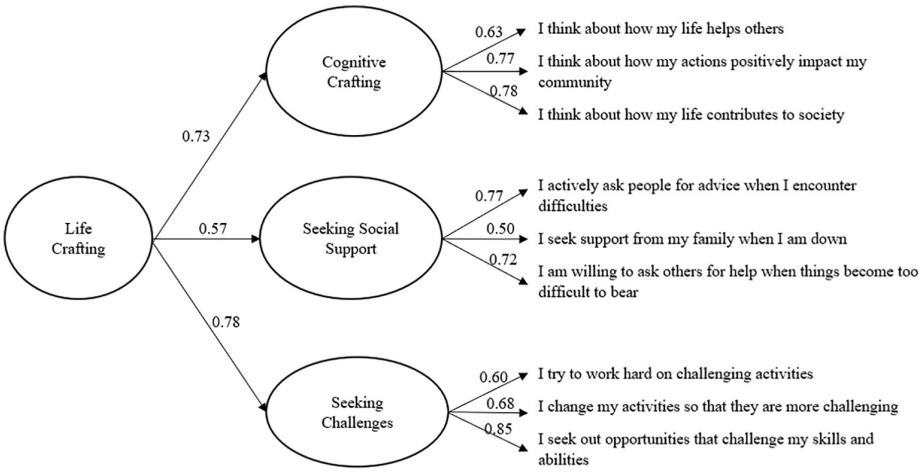

FIGURE 1 | The factorial model of the Life Crafting Scale.

Table 12. We first conducted a structure model regressing life crafting on mental health and meaning in life in sample 1 (n = 331). The model showed adequate fit [$\chi^2_{(331)}$ = 1090.09, df = 484, χ^2/df = 2.25, CFI = 0.92, TLI = 0.91, RMSEA = 0.62 (0.057, 0.066), SRMR = 0.08, AIC = 32245.93, BIC = 32664.16]. Life crafting was directly associated with mental health (β = 0.65, $p < 0.001$, R^2 = 0.42) and meaning in life (β = 0.91, $p < 0.001$, R^2 = 0.83). Following this, we examined the structure model which life crafting regressed on job burnout and work engagement in sample 2 (n = 362). The model showed adequate fit [$\chi^2_{(362)}$ = 914.29, df = 501, χ^2/df = 1.82, CFI = 0.91, TLI = 0.90, RMSEA = 0.48 (0.043, 0.053), SRMR = 0.07, AIC = 26349.81, BIC = 26847.94]. Life crafting was directly associated with job burnout (β = −0.42, $p < 0.001$, R^2 = 0.18) and work engagement (β = 0.54, $p < 0.001$, R^2 = 0.30).

DISCUSSION

The purpose of this paper was to conceptualize life crafting and to develop, validate and evaluate a robust measure of overall life crafting. The results showed that a second-order factorial model of overall life crafting, comprised of three first-order factors called cognitive crafting, seeking social support, and seeking challenges, fitted the data best. Our results, therefore, support the conceptual definition of Life Crafting as the conscious efforts individuals exert to create meaning in their lives through (a) cognitively (re-)framing how they view life, (b) seeking social support systems to manage life challenges and (c) actively seeking challenges to facilitate personal growth. The results further support partial measurement invariance across genders. Moreover, life crafting was substantially different but related to job crafting and proactive personality. Furthermore, life crafting was positively associated with meaning in life, mental health, and work engagement and was negatively related to job burnout. Our results therefore support the notation that life crafting could be a valuable meaning-making strategy which people could employ to create more meaningful life experiences.

The Life Crafting Framework

The first objective of this paper was to conceptualize and validate a conceptual framework for Life Crafting. Our results support the notion that "Life Crafting" refers to the conscious efforts individuals exert to create meaning in their lives through (a) cognitively (re-)framing how they view life, (b) by seeking social support systems to manage life challenges, and (c) to actively seeking challenges to facilitate personal growth. From this definition, and supported by our empirical findings, Life Crafting is seen to consist out of three factors that provide individuals with the means to both search for new sources of meaning in their lives but also affords the opportunity to (re)craft life in such a way to allow for the self to something larger than themselves:

1. *Cognitive crafting*. Our results indicate that cognitive crafting is an essential component of one's life crafting strategy. Cognitive crafting is defined as the individual's ability to proactively reshape or cognitively re-frame the physical, cognitive or social features of work or life in order for it to be perceived as more meaningful.
2. *Seeking social support*. Human beings are fundamentally social animals with a desire to connect to others and its therefore not surprising that seeking social support was found to be a component of life crafting. Seeking social support refers to the extent to which individuals seek out social support systems and networks to achieve personal/professional goals and aid in managing adversity. Meaning is therefore crafted through establishing mutually beneficial relationships with others.
3. *Seeking challenges*. The inherent need to grow and develop ourselves is at the core of most meaning-making strategies (van Zyl et al., 2020). Seeking challenges refers to the active efforts implemented by individuals to stretch their

TABLE 9 | Measurement invariance across genders.

Model	χ^2	*df*	CFI	TLI	RMSEA	SRMR	AIC	BIC	Model comparison	$\Delta\chi^2$	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
M1 configural invariance	58.907	48	0.986	0.978	0.035 [0.000–0.063]	0.040	8492.916	8726.249	–	–	–	–	–	–
M2 metric invariance: first order	67.920	54	0.982	0.976	0.038 [0.000–0.063]	0.052	8489.751	8699.750	M1 vs. M2	9.01	0.004	0.002	0.003	0.012
M3 metric invariance: second order	68.623	56	0.983	0.979	0.035 [0.000–0.061]	0.053	8486.330	8688.551	M2 vs. M3	0.70	0.001	0.003	0.003	0.001
M4 scalar invariance: first order	73.279	62	0.983	0.985	0.032 [0.000–0.057]	0.057	8478.791	8657.679	M3 vs. M4	4.66	0.000	0.006	0.003	0.004
M5 scalar invariance: second order	95.830	64	0.958	0.953	0.052 [0.029–0.073]	0.075	8498.865	8669.976	M4 vs. M5	22.55*	0.025	0.028	0.020	0.018
M6 partial invariance	76.314	63	0.982	0.980	0.034 [0.000–0.059]	0.059	8479.884	8654.884	M4 vs. M6	3.04	0.001	0.005	0.002	0.002
M7 strict invariance first order	80.600	69	0.985	0.984	0.031 [0.000–0.055]	0.058	8471.886	8623.552	M6 vs. M7	4.29	0.003	0.004	0.003	0.001
m8 strict invariance Second order	82.517	71	0.985	0.985	0.030 [0.000–0.055]	0.057	8469.788	8613.676	M7 vs. M8	1.92	0.000	0.001	0.001	0.001

*χ^2, Chi-square; df, degrees of freedom; TLI, Tucker-Lewis Index; CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation [90%CI]; SRMR, Standardized Root Mean Square Residual; AIC, Akaike Information Criterion; BIC, Bayes Information Criterion; aBIC, Adjusted Bayes Information Criterion. *$p < 0.05$.*

current capabilities and to learn new skills/abilities aimed at facilitating personal growth and environmental mastery.

These factors conceptually overlaps with both the three factors of Wrzesniewski and Dutton's (2001) conceptualization of job crafting (i.e. cognitive-, relational-, and task crafting) and two of Tims and Bakker (2010)'s conceptualization (i.e. seeking social resources, increasing challenges). This is probably not a surprise as these forms of (job) crafting may lead to the satisfaction of basic human needs, i.e., cognitive crafting (need for positive self-image), relational crafting (need for relatedness), and task crafting (need for competence). In addition, according to Baumeister and Vohs (2002), the essence of the meaning is the need to establish connections with others. When individuals feel that their lives are connected with something larger than themselves and that they are making a significant contribution to society, they intend to appraise their lives as full of meaning (Wong, 2020). When individuals actively seek challenges in their lives, it could lead to either a physical- or perceptive increase in their available resources. Further, we believe that when individuals are able to cognitively craft their lives and are able to seek out the necessary social support needed to facilitate goal achievement, that it could close the gap between one's current life and desired life. As such, individuals would be able to more actively see how current life tasks relate to their overall goals and therefore result in life feeling more meaningful.

Psychometric Properties of the Life Crafting Scale

The results further showed that second-order factorial model for overall Life Crafting, comprised out of three first-order factors (cognitive crafting, seeking social support, and seeking challenges) fitted the data best. This was in contrast to the initial expectation that life crafting would be a multi-dimensional construct comprised out of eight factors. Study 1 showed that only three factors could be meaningfully extracted from the data, which was confirmed in Study 2. The reason for such may be threefold. Firstly, the original eight dimensions and their items were primarily derived from other domain-specific studies on crafting behaviors and their associated scales. Given the conceptual overlap between these different crafting approaches, creating an item pool with similarly worded items may have created factors that look rather homogenous to participants. This was in contrast to initial expectations as we expected different approaches to cognitive crafting, for example, to produce different factorial structures. For example, Bindl et al. (2019) measured cognitive crafting with items such as "I thought about ways in which my job as a whole contributed to society" and "I thought about how my job contributed to the organization's goals." Whereas Slemp and Vella-Brodrick (2013) measured cognitive crafting with items such as "Think about the ways in which your work positively impacts your life" and "Reflect on the role your job has for your overall wellbeing." From these two sets of items, it would seem as though the strategies employed to cognitively craft work would consist of two subdimensions: transcending personal goals and positive thinking. In a similar vein, relational

TABLE 10 | Relationships with job crafting and proactive personality.

Regressive path	Standardized					Validity established
	β	SE	*t*-value	*p*	R^2	
Life crafting→ seek job challenges	0.578	0.051	11.447	<0.01	0.334	Yes
Life crafting→ reduce job demands	0.064	0.066	0.963	<0.01	0.004	No
Life crafting→ seek job resources	0.686	0.051	13.505	<0.01	0.470	Yes
Proactive personality → life crafting	0.457	0.057	8.035	<0.01	0.209	Yes

TABLE 11 | Confidence intervals of the correlation among life crafting, three job crafting subfactors, and proactive personality.

Variable	Seek job resources	Seek job challenges	Reduce job demands	Proactive personality
Life Crafting	[0.610 0.812]	[0.463 0.668]	[−0.067 0.199]	[0.344 0.577]

TABLE 12 | Relationships with burnout, engagement, meaning in life, and mental health.

Regressive path	Standardized					Validity established
	β	SE	*t*-value	*p*	R^2	
Life crafting→ job burnout	−0.42	0.06	−6.96	<0.01	0.18	Yes
Life crafting→ work engagement	0.54	0.05	10.11	<0.01	0.30	Yes
Life crafting→ mental health	0.65	0.05	13.90	<0.01	0.42	Yes
Life crafting →meaning in life	0.91	0.20	4.47	<0.01	0.83	Yes

crafting could also be conceptually divided into creating new relationships, optimizing current relationships, and utilizing social resources. However, it would seem as though participants do not differentiate between micro-level crafting behaviors but rather just focus on changing the way in which they view life, seeking resources to support their meaning-making processes, and engaging in challenges to stretch their current capabilities in order for them to grow.

Secondly, the results of EFA showed that resources crafting could be removed as a potential factor of life crafting. This dimension was derived from job crafting's dimension, i*ncreasing structural job resources* (Tims et al., 2012). Increasing structural job resources refers to employees' proactive behaviors initiated in order to increase resource variety, develop new opportunities, and enhance autonomy at work. However, in the global life domain, there are relatively few situations where people are required to deal with "structure," which might be why this specific crafting domain did not manifest as initially expected. Further, whilst controlling for environmental factors, one would also assume that individuals are rather autonomous in the way in which they approach life. This is in contrast to work where the roles, functions, processes, and procedures are usually relatively well defined and leave little room for autonomy (Van Zyl et al., 2010). As such, there may be no need to correct for a lack of autonomy through "seeking structural resources" in general life.

Thirdly, demands crafting was also not found to be a component of life crafting. We think that this is because demands crafting is driven by different motivations in contrast to the other factors. In previous studies, demands crafting related to individuals' avoiding motivation, whereas seeking resources and seeking challenges were connected with approaching motivation (Tims et al., 2012; Lichtenthaler and Fischbach, 2019). Therefore, some researchers found that reduced demands always showed insignificance results with other dimensions (Tims et al., 2012; Lichtenthaler and Fischbach, 2019). Similar to the findings of job crafting, reducing demands was not found to be an essential element of life crafting as it probably does not represent a purposeful behavior contributing to meaning but perhaps occurs due to other reasons and needs, e.g., creating sufficient time and energy for meaningful activities.

Finally, the results showed that the second-order life crafting model demonstrated configural, metric, and partial scalar invariance across gender. These findings imply that overall life crafting, showed similar factor structures, factor loadings, intercepts, and residual errors for both male and female samples. However, seeking social support showed a difference when we tested the scalar invariance. This finding showed that women, on average, seek more social help than men. This result is consistent with other studies showing that women are more inclined than men to seek help when they encounter destructive/difficult/challenging issues (Koydemir-Özden, 2010; Liddon et al., 2018). The potential reason for the difference is that the female participants held more positive attitudes toward help-seeking behaviors and more easily recognized their needs for help than male participants (Ang et al., 2004).

The Relationship Between Life Crafting and Individual Outcomes

As we expected, life crafting showed a positive relationship with seeking job resources, seeking job challenges, and proactive personality. This result indicated that life crafting might tap the

same conceptual area as job crafting and proactive personality. There was no significant relationship between life crafting and a reduction in job demands in the current study. This result implies that life crafting and reducing demands might be two independent variables. One reason for this is that we did not include any demands-based items into the final scale. Another reason may be that reducing job demands is driven by the motivation to avoid, whereas seeking job resources and seeking job challenges have a focus on a more proactive motivation (Hu et al., 2020).

Finally, the current study confirmed the criterion validity of life crafting. We found a positive relationship between life crafting and meaning in life, mental health, and work engagement. These results are consistent with previous studies. For example, Tims et al. (2016) found that crafting behaviors in the work domain could increase meaningfulness. Slemp and Vella-Brodrick (2014) pointed out that job crafting may improve mental health by satisfying personal intrinsic needs. In addition, we found a negative relationship between life crafting and job burnout. This result is in line with Tims et al. (2013) finding that job crafting plays a vital role in preventing job burnout. Overall, the validity of the Life Crafting Scale is further supported by the results and implies that life crafting could be an important predictor for people's mental condition or state.

Limitations and Recommendations

Despite the novelty of the current study, it is not without its limitations. First, we only collected cross-sectional data. Therefore, the scale's stability over time is unknown. According to our definition, life crafting is a self-driven strategy to create meaning, which means life crafting might change over time while the individuals' motivation changes. Therefore, the longitudinal stability or invariance of the life crafting scale should be investigated in future studies.

Second, all of our criterion indicators relied upon self-report measures. Self-report measures are more open to positive bias, and research has shown that there could be a discrepancy between what is being felt and what is being reported (Van Zyl and Ten Klooster, 2022). This may lead to higher levels of common method effects and positive reporting bias. Future research should aim to validate the Life Crafting Scale against more objective indicators of positive mental health, meaning, and job performance.

Thirdly, the current study checked the measurement invariance of life crafting across gender. However, life crafting may also vary across age and occupations. Future researchers could examine the factorial equivalence of the Life Crafting scale across these different demographic factors. Finally, the cross-cultural validity of the instrument should be investigated. In the current study, the empirical validation of the instrument was only conducted within a western, predominantly individualistic, cultural context. Therefore, future studies should attempt to validate the scale in other cultural contexts such as in eastern countries in order to provide more evidence as to its cross-cultural applicability.

Finally, our approach to life crafting is founded in the philosophical tenents of the meaning-making theory, the conservation of resources theory, and the extension of the Job-Demands and Resources model. Although we believe that our approach is holistic and encompassing, there may be other interpretative frameworks that may also explain life crafting behaviors (e.g., positive existential psychology or chaos theory). For example, Gleick (1988) proposed that we are living in a chaotic world; therefore, our plans are consistently changing because of disruption by external, unplanned events. A typical example is the Covid-19 crisis which hindered our life in various ways yet made us rethink the value and purpose of suffering, the meaning of our lives, and how these affect our mental health or wellbeing (Wong, 2020; Wong et al., 2021). Positive existentialism indicated that the pandemic also resulted in people adopting new strategies to form-, search for or create meaning through reframing the purpose and function of suffering (Wong et al., 2021). These approaches could also be seen or interpreted as life crafting strategies, which reframes how people view essential life events or create meaning from suffering (Wong et al., 2021). Therefore, we urge future researchers to expand upon life crafting theory by approaching such from different philosophical traditions.

CONCLUSION

In conclusion, our first attempt to conceptualize and measure life crafting as a global meaning-making strategy has shown promising results. Our results support the importance of life crafting as a tool individuals can employ to enhance their mental health. We also found that crafting behaviors transcends context and that individuals approach meaning-making from a holistic, integrated perspective. Life crafting could therefore be important and alternative strategies researchers and practitioners could use to aid their individuals to find more meaning in their lives.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Review Committee of Eindhoven University of Technology (under registration number: ERB2020IEIS20). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SC collected and analyzed the data and drafted the first version of the manuscript. LM, LZ, and ED made substantial contributions to the drafting of the final manuscript and to the revisions. All authors contributed to the conceptualization and design of the study.

REFERENCES

Abdel Hadi, S., Bakker, A. B., and Häusser, J. A. (2021). The role of leisure crafting for emotional exhaustion in telework during the COVID-19 pandemic. *Anxiety Stress Coping* 34, 530–544. doi: 10.1080/10615806.2021.1903447

Akgunduz, Y., Alkan, C., and Gök, Ö. A. (2018). Perceived organizational support, employee creativity and proactive personality: the mediating effect of meaning of work. *J. Hosp. Tour. Manage.* 34, 105–114. doi: 10.1016/j.jhtm.2018.01.004

Akgunduz, Y., Gök, Ö. A., and Alkan, C. (2020). The effects of rewards and proactive personality on turnover intentions and meaning of work in hotel businesses. *Tour. Hosp. Res* 20, 170–183. doi: 10.1177/1467358419841097

Ang, R. P., Lim, K. M., Tan, A. G., and Yau, T. Y. (2004). Effects of gender and sex role orientation on help-seeking attitudes. *Curr. Psychol.* 23, 203–214.

Asparouhov, T., and Muthén, B. (2009). Exploratory structural equation modeling. *Struct. Equ. Model. Multidiscip. J.* 16, 397–438.

Bakker, A. B., and Costa, P. L. (2014). Chronic job burnout and daily functioning: a theoretical analysis. *Burn. Res.* 1, 112–119. doi: 10.11124/jbisrir-2015-2380

Bateman, T. S., and Crant, J. M. (1999). Proactive behavior: meaning, impact, recommendations. *Bus. Horiz.* 42, 63–70.

Bateman, T. S., and Crant, J. M. (1993). The proactive component of organizational behavior: a measure and correlates. *J. Org. Behav.* 14, 103–118.

Baumeister, R. F., and Vohs, K. D. (2002). The pursuit of meaningfulness in life. *Handb. Posit. Psychol.* 1, 608–618.

Beck, J. S. (1995). *Cognitive Therapy: Basics and Beyond*. New York, NY: Guilford press.

Berg, J., Dutton, J., and Wrzesniewski, A. (2015). Purpose and meaning in the workplace. *Rev. Am. Psychol. Assoc.* 230–232. doi: 10.1037/14183-000

Bindl, U. K., Unsworth, K. L., Gibson, C. B., and Stride, C. B. (2019). Job crafting revisited: implications of an extended framework for active changes at work. *J. Appl. Psychol.* 104, 605–628. doi: 10.1037/apl0000362

Bronfenbrenner, U. (1992). "Ecological systems theory," in *Six Theories of Child Development: Revised Formulations and Current Issues*, ed. R. Vasta (London: Jessica Kingsley Publishers), 187–249.

Bruning, P. F., and Campion, M. A. (2018). A role-resource approach-avoidance model of job crafting: a multimethod integration and extension of job crafting theory. *Acad. Manage. J.* 61, 499–522. doi: 10.5465/amj.2015.0604

Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Commun. Methods Meas.* 12, 25–44.

Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Model.* 9, 233–255.

Chowdhury, R., Heng, K., Shawon, M. S. R., Goh, G., Okonofua, D., Ochoa-Rosales, C., et al. (2020). Dynamic interventions to control COVID-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries. *Eur. J. Epidemiol.* 35, 389–399. doi: 10.1007/s10654-020-00649-w

Claes, R., Beheydt, C., and Lemmens, B. (2005). Unidimensionality of abbreviated proactive personality scales across cultures. *Appl. Psychol.* 54, 476–489.

Clausen, T., and Borg, V. (2011). Job demands, job resources and meaning at work. *J. Manag. Psychol.* 26, 665–681.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177/001316446002000104

Czekierda, K., Banik, A., Park, C. L., and Luszczynska, A. (2017). Meaning in life and physical health: systematic review and meta-analysis. *Health Psychol. Rev.* 11, 387–418. doi: 10.1080/17437199.2017.1327325

David, I., and Iliescu, D. (2020). The influence of religiosity and meaning making on work outcomes: a path analysis. *Curr. Psychol.* [Epub ahead of print].

De Bloom, J., Vaziri, H., Tay, L., and Kujanpää, M. (2020). An identity-based integrative needs model of crafting: crafting within and across life domains. *J. Appl. Psychol.* 105, 1423–1446. doi: 10.1037/apl0000495

de Jong, E. M., Ziegler, N., and Schippers, M. C. (2020). From shattered goals to meaning in life: life crafting in times of the COVID-19 Pandemic. *Front. Psychol.* 11:577708. doi: 10.3389/fpsyg.2020.577708

Dekker, I., De Jong, E. M., Schippers, M. C., De Bruijn-Smolders, M., Alexiou, A., and Giesbers, B. (2020). Optimizing students' mental health and academic performance: AI-enhanced life crafting. *Front. Psychol.* 11:1063. doi: 10.3389/fpsyg.2020.01063

Demerouti, E., Bakker, A. B., Nachreiner, F., and Schaufeli, W. B. (2001). The job demands-resources model of burnout. *J. Appl. Psychol.* 86, 499–512.

Demerouti, E., Bakker, A. B., Vardakou, I., and Kantas, A. (2003). The convergent validity of two burnout instruments: a multitrait-multimethod analysis. *Eur. J. Psychol. Assess.* 19, 12–23.

Demerouti, E., Bakker, A. B., and Voydanoff, P. (2010). Does home life interfere with or facilitate job performance? *Eur. J. Work Organ. Psychol.* 19, 128–149. doi: 10.1080/13594320902930939

Demerouti, E., Hewett, R., Haun, V., de Gieter, S., Rodríguez-Sánchez, A., and Skakon, J. (2020). From job crafting to home crafting: a daily diary study among six European countries. *Hum. Relat.* 73, 1010–1035. doi:10.1177/0018726719848809

Demerouti, E., and Peeters, M. C. (2018). Transmission of reduction-oriented crafting among colleagues: a diary study on the moderating role of working conditions. *J. Occup. Organ. Psychol.* 91, 209–234. doi: 10.1111/joop.12196

Dziuban, C. D., and Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychol. Bull.* 81, 358–361. doi: 10.1037/h0036316

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. London: Sage.

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

Frenzel, S., Junker, N. M., Avanzi, L., Bolatov, A., Haslam, S. A., Häusser, J. A., et al. (2022). A trouble shared is a trouble halved: the role of family identification and identification with humankind in wellbeing during the COVID-19 pandemic. *Br. J. Soc. Psychol.* 61, 55–85. doi: 10.1111/bjso.12470

Fuller, B. Jr., and Marler, L. E. (2009). Change driven by nature: a meta-analytic review of the proactive personality literature. *J. Vocat. Behav.* 75, 329–345.

George, D., and Mallery, P. (2019). *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference*. London: Taylor & Francis Ltd.

Gleick, J. (1988). *Chaos: Making a New Science*. London: William Heinemann.

Greenglass, E. R., and Fiksenbaum, L. (2009). Proactive coping, positive affect, and wellbeing: testing for mediation using path analysis. *Eur. Psychol.* 14, 29–39. doi: 10.1027/1016-9040.14.1.29

Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C. (1998). *Factor Analysis Multivariate Data Analysis*, 4 Edn. Upper Saddle River, NJ: Upper Saddle River.

Hambrick, E. P., and McCord, D. M. (2010). Proactive coping and its relation to the five-factor model of personality. *Individ. Differ. Res.* 8, 67–77.

Harju, L. K., Hakanen, J. J., and Schaufeli, W. B. (2016). Can job crafting reduce job boredom and increase work engagement? A three-year cross-lagged panel study. . *J. Vocat. Behav.* 95, 11–20. doi: 10.1016/j.jvb.2016.07.001

Heine, S. J., Proulx, T., and Vohs, K. D. (2006). The meaning maintenance model: on the coherence of social motivations. *Pers. Soc. Psychol. Rev.* 10, 88–110. doi: 10.1207/s15327957pspr1002_1

Hobfoll, S. E. (1989). Conservation of resources: a new attempt at conceptualizing stress. *Am. Psychol.* 44, 513–524. doi: 10.1037//0003-066x.44.3.513

Hsieh, H. F., and Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qual. Health Res.* 15, 1277–1288. doi: 10.1177/1049732305276687

Hu, Q., Taris, T. W., Dollard, M. F., and Schaufeli, W. B. (2020). An exploration of the component validity of job crafting. *Eur. J. Work Organ. Psychol.* 29, 776–793. doi: 10.1080/1359432x.2020.1756262

Ingusci, E., Signore, F., Pasca, P., Catalano, A., De Masi, S., Gioia, G. M., et al. (2020). Boosting human and social capital in the university: the role of seeking challenges on academic performance. *BPA Appl. Psychol. Bull.* 68, 30–42.

Jacob, Y., and Steger, M. F. (2021). "Meaning-centred coaching in the workplace," in *Positive Psychology Coaching in the Workplace*, eds W. A. Smith et al. (Cham: Springer), 555–574. doi: 10.1007/978-3-030-79952-6_29

Keyes, C. L. (2002). The mental health continuum: from languishing to flourishing in life. *J. Health Soc. Behav.* 43, 207–222. doi: 10.2307/3090197

Koydemir-Özden, S. (2010). Self-aspects, perceived social support, gender, and willingness to seek psychological help. *Int. J. Ment. Health* 39, 44–60. doi: 10.2753/imh0020-7411390303

Lamers, S. M., Westerhof, G. J., Bohlmeijer, E. T., ten Klooster, P. M., and Keyes, C. L. (2011). Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF). *J. Clin. Psychol.* 67, 99–110. doi: 10.1002/jclp.20741

Laurence, G. A. (2010). *Workaholism and Expansion and Contraction Oriented Job Crafting: The Moderating Effects of Individual and Contextual Factors*. Ph.D. dissertation. Syracuse, NY: The Graduate School Syracuse University.

Lepine, J. A., Podsakoff, N. P., and Lepine, M. A. (2005). A meta-analytic test of the challenge stressor-hindrance stressor framework: an explanation for inconsistent relationships among stressors and performance. *Acad. Manage. J.* 48, 764–775. doi: 10.5465/amj.2005.18803921

Lichtenthaler, P. W., and Fischbach, A. (2019). A meta-analysis on promotion-and prevention-focused job crafting. *Eur. J. Work Organ. Psychol.* 28, 30–50. doi: 10.1080/1359432x.2018.1527767

Liddon, L., Kingerlee, R., and Barry, J. A. (2018). Gender differences in preferences for psychological treatment, coping strategies, and triggers to help-seeking. *Br. J. Clin. Psychol.* 57, 42–58.

Lin, L. (2021). Longitudinal associations of meaning in life and psychosocial adjustment to the COVID-19 outbreak in China. *Br. J. Health Psychol.* 26, 525–534. doi: 10.1111/bjhp.12492

Maharaj, I., and Schlechter, A. F. (2007). Meaning in life and meaning of work: relationships with organizational citizenship behaviour, commitment and job satisfaction. *Manage. Dyn.* 16, 24–41. doi: 10.1186/s12960-018-0277-9

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochem. Med.* 22, 276–282. doi: 10.11613/bm.2012.031

Muthén, L. K., and Muthén, B. O. (2020). *Mplus (Version 8.3). [Statistical Software]*. Los Angeles, CA: Muthén and Muthén.

Nielsen, K., Nielsen, M. B., Ogbonnaya, C., Känsälä, M., Saari, E., and Isaksson, K. (2017). Workplace resources to improve both employee wellbeing and performance: a systematic review and meta-analysis. *Work Stress* 31, 101–120. doi: 10.1080/02678373.2017.1304463

Park, C. L. (2010). Making sense of the meaning literature: an integrative review of meaning making and its effects on adjustment to stressful life events. *Psychol. Bull.* 136, 257–301. 20192563doi: 10.1037/a0018301

Petrou, P., and Bakker, A. B. (2016). Crafting one's leisure time in response to high job strain. *Hum. Relat.* 69, 507–529. doi: 10.1177/0018726715590453

Petrou, P., Demerouti, E., Peeters, M. C., Schaufeli, W. B., and Hetland, J. (2012). Crafting a job on a daily basis: contextual correlates and the link to work engagement. *J. Organ. Behav.* 33, 1120–1141. doi: 10.1002/job.l783

Petrou, P., Demerouti, E., and Schaufeli, W. B. (2015). Job crafting in changing organizations: antecedents and implications for exhaustion and performance. *J. Occup. Health Psychol.* 20, 470–480.

Porter, C. M., and Woo, S. E. (2015). Untangling the networking phenomenon: a dynamic psychological perspective on how and why people network. *J. Manage.* 41, 1477–1500.

Richter, S., Van Zyl, L. E., Roll, L. C., and Stander, M. W. (2021). Positive psychological Coaching Tools: a systematic literature review. *Front. Psychiatry* 12:667200. doi: 10.3389/fpsyt.2021.667200

Rofcanin, Y., Bakker, A. B., Berber, A., Gölgeci, I., and Las Heras, M. (2019). Relational job crafting: exploring the role of employee motives with a weekly diary study. *Hum. Relat.* 72, 859–886. doi: 10.1177/0018726718779121

Rönkkö, M., and Cho, E. (2020). An updated guideline for assessing discriminant validity. *Organ. Res. Methods* 25, 6–14. doi: 10.1177/1094428120968614

Sacco, S. J., Park, C. L., Suresh, D. P., and Bliss, D. (2014). Living with heart failure: Psychosocial resources, meaning, gratitude and wellbeing. *Heart Lung* 43, 213–218. doi: 10.1016/j.hrtlng.2014.01.012

Schaufeli, W. B., Bakker, A. B., and Salanova, M. (2006). The measurement of work engagement with a short questionnaire: a cross-national study. *Educ. Psychol. Meas.* 66, 701–716.

Schippers, M. C., and Ziegler, N. (2019). Life crafting as a way to find purpose and meaning in life. *Front. Psychol.* 10:2778. doi: 10.3389/fpsyg.2019.02778

Schwarzer, R., and Knoll, N. (2003). "Positive coping: mastering demands and searching for meaning," in *Positive Psychological Assessment: A Handbook of Models and Measures*, eds S. J. Lopez and C. R. Snyder (Washington, DC: American Psychological Association), 393–409.

Slemp, G. R., and Vella-Brodrick, D. A. (2013). The job crafting questionnaire: a new scale to measure the extent to which employees engage in job crafting. *Int. J. Wellbeing* 3, 126–146. doi: 10.5502/ijw.v3i2.1

Slemp, G. R., and Vella-Brodrick, D. A. (2014). Optimizing employee mental health: the relationship between intrinsic need satisfaction, job crafting, and employee wellbeing. *J. Happiness Stud.* 15, 957–977.

Steger, M. F. (2009). "Meaning," in *The Encyclopedia of Positive Psychology*, ed. S. J. Lopez (New York, NY: John Wiley & Sons).

Steger, M. F. (2013). "Experiencing meaning in life: optimal functioning at the nexus of wellbeing, psychopathology, and spirituality," in *The Human Quest for Meaning*, 2nd Edn, ed. P. T. P. Wong (New York, NY: Routledge), 211–230.

Steger, M. F. (2019). "Meaning in life and in work," in *The Oxford Handbook of Meaningful Work*, eds A. Madden, R. Yeoman, K. Bailey, and M. Thompson (Oxford: Oxford University Press), 208.

Steger, M. F., and Dik, B. J. (2009). If one is looking for meaning in life, does it help to find meaning in work? *Appl. Psychol. Health Well Being* 1, 303–320. doi: 10.1111/j.1758-0854.2009.01018.x

Steger, M. F., Frazier, P., Kaler, M., and Oishi, S. (2006). The meaning in life questionnaire: assessing the presence of and search for meaning in life. *J. Couns. Psychol.* 53, 80–93. doi: 10.1037/0022-0167.53.1.80

Steger, M. F., Shim, Y., Barenz, J., and Shin, J. Y. (2014). Through the windows of the soul: a pilot study using photography to enhance meaning in life. *J. Contextual Behav. Sci.* 3, 27–30. doi: 10.1016/j.jcbs.2013.11.002

Tims, M., and Bakker, A. B. (2010). Job crafting: towards a new model of individual job redesign. *SA J. Ind. Psychol.* 36, 1–9. doi: 10.4102/sajip.v36i2.841

Tims, M., Bakker, A. B., and Derks, D. (2012). Development and validation of the job crafting scale. *J. Vocat. Behav.* 80, 173–186. doi: 10.1016/j.jvb.2011.05.009

Tims, M., Bakker, A. B., and Derks, D. (2013). The impact of job crafting on job demands, job resources, and wellbeing. *J. Occup. Health Psychol.* 18, 230–240.

Tims, M., Bakker, A. B., and Derks, D. (2015). Job crafting and job performance: a longitudinal study. *Eur. J. Work Organ. Psychol.* 24, 914–928.

Tims, M., Derks, D., and Bakker, A. B. (2016). Job crafting and its relationships with person–job fit and meaningfulness: a three-wave study. *J. Vocat. Behav.* 92, 44–53. doi: 10.1016/j.jvb.2015.11.007

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

Van Zyl, L. E. (2021). Social study resources and social wellbeing before and during the intelligent COVID-19 lockdown in the Netherlands. *Soc. Indic. Res.* 157, 1–23. doi: 10.1007/s11205-021-02654-2

Van Zyl, L. E., Deacon, E., and Rothmann, S. (2010). Towards happiness: experiences of work-role fit, meaningfulness and work engagement of industrial/organizational psychologists in South Africa. *SA J. Ind. Psychol* 36:890.

Van Zyl, L. E., and Ten Klooster, P. M. (2022). Exploratory structural equation modelling: practical guidelines and tutorial with a convenient online tool for Mplus. *Front. Psychiatry* 12:795672. doi: 10.3389/fpsyt.2021.795672

Van Zyl, L. E., Zondervan-Zwijnenburg, M. A. J., Dickens, L. R., and Hulshof, I. (2020). Exploring meaning in life through a brief photo-ethnographic intervention using Instagram: a Bayesian growth modelling approach. *Int. Rev. Psychiatry* 32, 723–745. doi: 10.1080/09540261.2020.1809357

van Zyl, L. E., Zondervan-Zwijnenburg, M. A., Dickens, L. R., and Hulshof, I. L. (2020a). Exploring meaning in life through a brief photo-ethnographic intervention using Instagram: a Bayesian growth modelling approach. *Int. Rev. Psychiatry* 32, 723–745. doi: 10.1080/09540261.2020.1809357

van Zyl, L. E., Roll, L. C., Stander, M. W., and Richter, S. (2020b). Positive psychological coaching definitions and models: a systematic literature review. *Front. Psychol.* 11:793. doi: 10.3389/fpsyg.2020.00793

Wang, J., and Wang, X. (2020). *Structural Equation Modeling: Applications Using Mplus*. Chichester: John Wiley & Sons.

Wellman, N., and Spreitzer, G. (2011). Crafting scholarly life: strategies for creating meaning in academic careers. *J. Organ. Behav.* 32, 927–931.

Wolff, H. G., and Spurk, D. (2020). Developing and validating a short networking behavior scale (SNBS) from Wolff and Moser's (2006) measure. *J. Career Assess.* 28, 277–302.

Wong, P. T. (2020). Existential positive psychology and integrative meaning therapy. *Int. Rev. Psychiatry* 32, 565–578. doi: 10.1080/09540261.2020.1814703

Wong, P. T. P., Mayer, C.-H., and Arslan, G. (2021). Editorial: COVID-19 and existential positive psychology (PP2.0): the new science of self-transcendence. *Front. Psychol.* 12:800308. doi: 10.3389/fpsyg.2021.800308

Wrzesniewski, A., and Dutton, J. E. (2001). Crafting a job: revisioning employees as active crafters of their work. *Acad. Manage. Rev.* 26, 179–201. doi: 10.5465/AMR.2001.4378011

Wrzesniewski, A., LoBuglio, N., Dutton, J. E., and Berg, J. M. (2013). "Job crafting and cultivating positive meaning and identity in work," in *Advances in Positive Organizational Psychology*, ed. A. B. Bakker (Bingley: Emerald Group Publishing Limited).

Wrzesniewski, A., McCauley, C., Rozin, P., and Schwartz, B. (1997). Jobs, careers, and callings: people's relations to their work. *J. Res. Personal.* 31, 21–33.

Youssef-Morgan, C., Van Zyl, L. E., and Ahrens, B. (2022). The work gratitude scale: development and evaluation of a multidimensional measure. *Front. Psychol.* 13:795328. doi: 10.3389/fpsyg.2021.795328

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

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

Copyright © 2022 Chen, van der Meij, van Zyl and Demerouti. 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.

APPENDIX

Appendix 1 | The Life Crafting Scale.

Instructions	**For the following set of questions, think about how you approach different components of your life. Consider the strategies you employ to create meaningful life experiences. Then rate each item as it applies to your life.**					
No	**Item**	**Never**	**Sometimes**	**Regularly**	**Often**	**Always**
Cognitive crafting						
1	I think about how my life helps others	1	2	3	4	5
2	I think about how my actions positively impact my community	1	2	3	4	5
3	I think about how my life contributes to society	1	2	3	4	5
Seeking social support						
4	I actively ask people for advice when I encounter difficulties	1	2	3	4	5
5	I seek support from my family when I am down	1	2	3	4	5
6	I am willing to ask others for help when things become too difficult to bear	1	2	3	4	5
Seeking challenges						
7	I try to work hard on challenging activities	1	2	3	4	5
8	I change my activities so that they are more challenging	1	2	3	4	5
9	I seek out opportunities that challenge my skills and abilities	1	2	3	4	5

Scoring: Create an average or "mean" score of the following items to create a score for each of the components of the Life Crafting Scale

1. **Cognitive Crafting** = (Item 1 + Item 2 + Item 3)/3
2. **Seeking Social Support** = (Item 4 + Item 5 + Item 6)/3
3. **Seeking Challenges** = (Item 7 + Item 8 + Item 9)/3

To create an overall score of Life Crafting, create an average score of the means for each of the aforementioned components

4. **Overall Life Crafting** = (Cognitive Crafting + Seeking Social Support + Seeking Challenges)/3

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 31 March 2022
doi: 10.3389/fpsyg.2022.795452

Flourishing Scale: Adaptation and Evidence of Validity in a Chilean High School Context

Marcos Carmona-Halty, Mauricio Marín-Gutierrez, Patricio Mena-Chamorro, Geraldy Sepulveda-Páez and Rodrigo Ferrer-Urbina*

Escuela de Psicología y Filosofía, Universidad de Tarapacá, Arica, Chile

This study aimed to adapt the Flourishing Scale to a Chilean high school context and provide evidence of its validity. Data were collected from 1,348 students (52% girls) from three different Chilean schools. The results of confirmatory factor analysis (CFA) supported a one–factor solution, multiple–group CFA supported gender invariance, and structural equation model indicated that the FS is related to positive and negative academic feelings. Overall, the evidence indicates that the Flourishing Scale adapted to the high school context is an instrument that produces valid and reliable scores in our high school Chilean sample.

Keywords: flourishing, psychometric analyses, high school students, Chilean students, gender invariance

OPEN ACCESS

Edited by:
Llewellyn Ellardus Van Zyl, North West University, South Africa

Reviewed by:
Johannes Alfons Karl, Victoria University of Wellington, New Zealand
Jenny Marcionetti, University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Switzerland

****Correspondence:***
Marcos Carmona-Halty
mcarmonah@academicos.uta.cl

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

Received: *15 October 2021*
Accepted: *21 February 2022*
Published: *31 March 2022*

Citation:
Carmona-Halty M, Marín-Gutierrez M, Mena-Chamorro P, Sepulveda-Páez G and Ferrer-Urbina R (2022) Flourishing Scale: Adaptation and Evidence of Validity in a Chilean High School Context. Front. Psychol. 13:795452. doi: 10.3389/fpsyg.2022.795452

INTRODUCTION

Positive education is an emerging area of study aimed at encouraging—without ignoring the negative aspects inherent in all human activity—members of the educational community to flourish and develop their full potential (Jacobs and Renandya, 2019). More specifically, positive education is a discipline that emerges from positive psychology and aims to complement the traditional emphasis on developing academic skills with initiatives to promote well–being and optimal functioning (Seligman et al., 2009). In this line, recent research has shown that the development of personal strengths and resources are potential variables for increasing performance and other desirable outcomes in the high school setting (e.g., Steinmayr et al., 2018; Widlund et al., 2018; Su et al., 2019). In addition, the efficacy of programs aimed at increasing levels of well–being and reducing depressive symptomatology, which favor academic performance, has been confirmed (e.g., Shoshani and Steinmetz, 2013; Shoshani and Slone, 2017; Schoeps et al., 2018).

One of the concepts that has received increasing attention from educational contexts is the so–called *flourishing* (e.g., De la Fuente et al., 2017; Shoshani and Slone, 2017; Datu, 2018; Garzón–Umerenkova et al., 2018; Datu et al., 2019; Chamizo–Nieto et al., 2021; Holliman et al., 2021). Flourishing is synonymous with a high mental well–being level and reflects positive mental health and positive development (Huppert and So, 2013; Hone et al., 2014). More specifically, flourishing is the combination (in a single construct) of feeling good and functioning effectively in one's life. The first refers to feel interest in and a commitment to the activities of daily living, self–confidence, and affect, while the second refers to feeling in control of the course of one's life, having a purpose, and establishing and maintaining positive relationships with others (Ryff and Singer, 1998; Keyes, 2002; Huppert, 2009; Huppert and So, 2013; De la Fuente et al., 2017).

Recent positive education research has shown that flourishing is positively related to desired academic outcomes, such as performance (Datu, 2018), personal resources

(Ouweneel et al., 2011), engagement (Datu, 2018), achievement goal orientation (Datu et al., 2019), adaptability and social support (Holliman et al., 2021), positive teacher–student relationships (Chamizo–Nieto et al., 2021), basic psychological needs (Herrera et al., 2021), and passion for learning (Chen et al., 2021). Conversely, it is negatively related to undesired academic outcomes, such as depression and distress (De la Fuente et al., 2017), procrastination (Garzón–Umerenkova et al., 2018), and psychotic experiences (Oh et al., 2021). Together, these studies show that flourishing is a key construct applicable to the high school context and to the aims of positive education. Therefore, flourishing could help understand the processes underlying the optimal functioning of children and adolescents in school contexts.

One of the most widely used instruments to evaluate flourishing is the Flourishing Scale (FS) developed by Diener et al. (2010). This scale is a brief self–reported measurement that assesses the key components of psychosocial well–being: meaning and purpose in life, supportive and rewarding relationships, engaged and interested, contribute to the well–being of others, competency, self–acceptance, optimism, and being respected (Diener et al., 2009). Initial validation studies support FS as a one–factor solution with adequate psychometric properties (see Diener et al., 2010). More recently, additional validation studies have supported its psychometric properties (e.g., Romano et al., 2020; Martín–Carbonell et al., 2021; Tan et al., 2021) and shown its cross–cultural validity (e.g., Brazil—da Fonseca et al., 2015; China—Lin, 2015; France—Villieux et al., 2016; Egypt—Salama–Younes, 2017; India—Singh et al., 2016; New Zealand—Hone et al., 2014; Russia—Didino et al., 2019).

Despite the contribution that the studies have made to flourishing research, more research efforts are needed, specially, in Spanish–speaking South American countries where minimal research was done to assess the psychometric properties of FS (e.g., Colombia—Martín–Carbonell et al., 2021; Peru—Cassaretto and Martínez Uribe, 2017). The present study attempts to fill the gap on the scarcity of flourishing measures by adapting the FS to the Chilean high school context and examining its psychometric properties. We hope to contribute to increasing the scarce research on positive education in South American countries. More specifically, we aim to adapt the FS to the usual conditions of Chilean high school students and provide evidence of its validity following both a within–network and between–network construct validity. The first refers to assessing reliability, factor structure, and gender invariance, while the second refers to assessing the extent to which flourishing is associated with theoretically related constructs. In this line, given that the FS measures (only) the psychosocial components of well–being, the Scale of Positive and Negative Experiences (SPANE) —developed by Diener et al. (2010)— complements this indicator by measuring a range of positive and negative emotions and feelings in a specific time range (for example, during the past 4 weeks). Accordingly, the FS score has shown positive and negative significant relationships with the positive and negative feelings dimensions of the SPANE, respectively. For example, 0.69 and −0.48 (Giuntoli et al., 2017); 0.67 and −0.47 (Howell and Buro, 2015); and 0.58 and −0.42 (Silva and Caetano, 2013).

Based on the arguments presented, we hypothesize the following: The FS adapted to the high school context will demonstrate adequate psychometric properties in a sample of Chilean high school students. Also, we expect positive and negative relationships between FS scores and study–related positive and negative feelings (measured with the SPANE), respectively.

METHOD

Sample

The sample comprised 1,348 (52% girls) Chilean high school students between grades 7–12 (i.e., 13–18 years old, M = 15.04, SD = 1.43). The students were from three different secondary schools (each of them hosted approximately 600 students) from two urban centers in the country's northern regions: Arica and Iquique. Of 1,348 students, 17% were 13 years old, 19% were 14 years old, 18% were 15 years old, 21% were 16 years old, 22% were 17 years old, and 3% were 18 years old. In addition, 13% correspond to low, 79% to medium, and 8% to high socioeconomic levels.

Instruments

The *Flourishing Scale* (Diener et al., 2010) is composed of eight–item. Each item is rated by respondents using a 7–point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*). In this study, 3 expert judges were asked to compare both the Spanish and English language version of the FS (available on Ed Diener's website[1]) to establish whether both versions did not differ from each other. Furthermore, they checked the instrument's legibility. Subsequently, the FS was adapted to the educational setting of the students following the recommendations described in the literature associated with the adaptation of instruments (see Muñiz et al., 2013; Vallejo–Medina et al., 2017). More specifically, a rewording of the items from the general context to the school context was conducted. For example, "*I am engaged and interested in my daily activities*" was changed to "*I am engaged and interested in my daily school activities*." Finally, a pilot test was conducted with the FS adapted version (see **Table 1**) where 30 Chilean high school students were encouraged to answer the scale and indicate possible comprehension issues. At this stage, none of the participants expressed problems with understanding the items or the answering format of the FS.

The *Scale of Positive and Negative Experiences* (Diener et al., 2010) is composed of 12 items. Each item is rated by respondents using a 5–point Likert scale (1 = *never*, 5 = *always*). The scale is integrated by two subscales (six items each): positive (e.g., "*I have had pleasant feelings*") and negative (e.g., "*I have had unpleasant feelings*") feelings. This study used an adaptation to the Chilean high school context of the original SPANE, which demonstrated adequate psychometric properties (see Carmona–Halty and Villegas–Robertson, 2018). In our sample, internal consistency —for alpha and omega index— was 0.931 and 0.931,

[1]https://eddiener.com

TABLE 1 | Descriptive and reliability statistics information of the flourishing scale.

	Descriptive statistics				Reliability statistics			CFA factor loadings	S.E.
	Mean (SD)	S	K	W	CHI	α if item is dropped	ω if item is dropped		
1. I lead a purposeful and meaningful school life.	5.83 (1.359)	−1.187	0.977	0.811*	0.631	0.850	0.852	0.744*	0.015
2. My social relationships at school are supportive and rewarding.	5.73 (1.409)	−1.182	1.027	0.824*	0.499	0.865	0.865	0.598*	0.020
3. I am engaged and interested in my daily school activities.	5.88 (1.250)	−1.140	0.986	0.819*	0.679	0.845	0.846	0.784*	0.014
4. At school I actively contribute to the happiness and well-being of others.	5.81 (1.237)	−1.043	0.861	0.840*	0.592	0.854	0.856	0.692*	0.017
5. I am competent and capable in school activities that are important to me.	6.30 (0.972)	−1.667	3.151	0.722*	0.607	0.853	0.854	0.748*	0.017
6. At school I am a good person and live a good life.	6.10 (1.124)	−1.512	2.592	0.770*	0.663	0.847	0.848	0.769*	0.014
7. I am optimistic about my school future.	6.07 (1.292)	−1.771	3.211	0.729*	0.665	0.846	0.847	0.771*	0.015
8. People at school respect me.	5.92 (1.238)	−1.351	1.848	0.805*	0.613	0.853	0.854	0.712*	0.016
Flourishing	47.68 (7.110)	−19.04	16.47	0.901*					

$p < 0.001$; SD, standard deviation; S, skewness standardized; K, kurtosis standardized; W, Shapiro–Wilk test; CHI, corrected homogeneity index; CFA, confirmatory factor analysis; and S.E., standard error.

TABLE 2 | Fit Indexes for single–group and multiple–group CFA of the flourishing scale.

	χ^2	*df*	χ^2/df	RMSEA	90% CI	CFI	TLI	SRMR	CMs	Δ CFI	Δ RMSEA
Single–group CFA											
M1 One factor solution	215.24	20	107.62	0.085	[0.075, 0.096]	0.976	0.967	0.024	–	–	–
Multiple–group CFA											
M2 Configural invariance	261.60	40	5.415	0.091	[0.080, 0.101]	0.946	0.924	0.037	–	–	–
M3 Metric invariance	278.20	47	5.919	0.085	[0.076, 0.095]	0.944	0.933	0.055	M2–M3	0.002	0.006
M4 Scalar invariance	313.47	54	5.805	0.084	[0.076, 0.094]	0.937	0.934	0.065	M3–M4	0.007	0.001

χ2, Chi-square; df, degree of freedom; RMSEA, root mean square error of approximation; CI, 90% confidence interval; CFI, comparative fit index; TLI, Tucker–Lewis index; SRMR, standardized root mean square residual; and CMs, comparisons between models.

for study–related positive feelings, and for study–related negative feelings was 0.849 and 0.855, respectively.

Procedure

The procedure included contacting the principals of schools to explain to them the research's aim, scope, and needs. Once the proposal was accepted, a written authorization was requested from the principals, students, and parents. Data collection was carried out in group sessions of 25 students through an electronic procedure. For this purpose, each student had a computer at their disposal where the questionnaires had been previously uploaded. The students took about 10 min to answer the questionnaire and data collection lasted approximately 3 weeks.

Analysis

Sequential analyses were conducted using Jamovi 1.2 (The Jamovi Project, 2020) and Mplus 8.2 (Muthén and Muthén, 1998/2017). First, mean scores, standard deviation, standardized skewness and standardized kurtosis, gender differences, and Shapiro–Wilk test were calculated. Second, the internal consistency was estimated using Cronbach's alpha (α) and McDonald's omega (ω) coefficients, the corrected homogeneity index, and the alpha and omega indexes if any of the items were eliminated. Third, to determine whether the model proposed by the FS adequately represents the data collected, a confirmatory factor analysis (CFA) was performed using the weighted least square with mean estimation method (WLSMV) —which is robust to significant deviations from the normal distribution— and the polychoric correlation matrix. The model fit was interpreted according to the cut–off points proposed by Schreiber (2017) (e.g., CFI > 0.95; TLI > 0.95; RMSEA < 0.06). Fourth, to explore gender invariance, a multiple–group CFA was performed, where three levels of equivalence (i.e., configural invariance, metric invariance, and scalar invariance) were evaluated (Chen and West, 2008), using changes in CFI and RMSEA ($\Delta < 0.010$) as criteria to determine whether measurement invariance was established (Cheung and Rensvold, 2002; Chen, 2007;

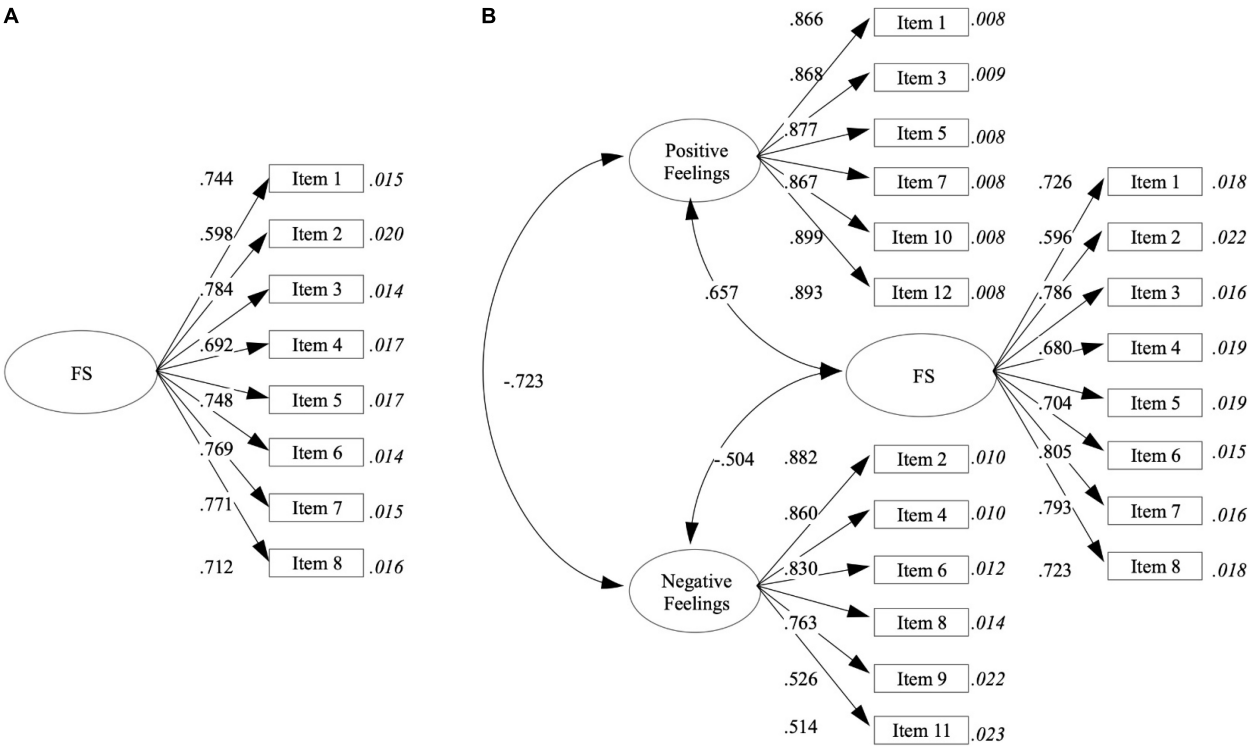

FIGURE 1 | Graphical representation of flourishing scale model **(A)** and the SEM covariate between FS and SPANE **(B)**. Values in italics correspond to the standard error.

Dimitrov, 2010). Finally, to examine the criterion validity of the FS, a structural equation model (SEM) was conducted between the covariations of positive and negative feelings and the FS score.

RESULTS

Descriptive Analysis

Table 1 shows the descriptive statistics for the FS at item level, including reliability and factor loading as they emerged in the CFA analysis described below. The Shapiro–Wilk test showed that none of the items had a normal distribution. Following previous research (e.g., Diener et al., 2010), gender differences were considered. However, independent sample *t*-test reveal that there are not statistical significance differences between boys' (M = 6.015, SD = 0.728) and girls' (M = 5.929, SD = 0.988) FS scores: $t_{(1346)} = 1.947$, $p > 0.05$.

Within–Network Construct Validation

The FS adapted to the Chilean high school context showed adequate internal consistency for Cronbach's alpha ($\alpha = 0.868$) and McDonald's omega ($\omega = 0.869$) index. In addition, as shown in **Table 1**, the results of the corrected homogeneity index suggests that it is not necessary to delete any items. **Table 2** (M1) and **Figure 1A** shows the CFA results for a model assuming one latent factor underlying all FS items. According to the standards recommended by Schreiber (2017), this model showed adequate fit index, reflecting a sufficient explanation for the observed covariate matrix. Indeed, the factorial loadings show adequate representations ($\lambda > 0.50$). In addition, the multiple–group CFA shows that the differences in the CFI and RMSEA —across the three invariance models (i.e., configural, metric, and scalar)— were lower than 0.010, which indicates gender invariance.

Between–Network Construct Validation

The SEM model showed satisfactory comparative and absolute fit indexes: χ^2 (167, 1,348) = 959.437, $p < 0.05$; CFI = 0.980; TLI = 0.978; RMSEA = 0.059, 90% CI (0.056–0.063). As shown in **Figure 1B**, there are positive ($\gamma = 0.657$, $p < 0.001$) and negative ($\gamma = -0.504$, $p < 0.001$) relationships between FS scores and positive and negative feelings, respectively.

DISCUSSION

The current study aimed to adapt the FS to the Chilean school context and obtain evidence of its validity to address the lack of measures and facilitate flourishing research in educational settings.

Our results are consistent with previous research in terms of the reliability indices and factor structure of the FS (e.g., Diener et al., 2010; Silva and Caetano, 2013; Howell and Buro, 2015; Villieux et al., 2016; Giuntoli et al., 2017; Checa et al., 2018). Also, gender invariance was demonstrated, leading to the conclusion that flourishing can be measured with the same

precision in boys and girls, which is consistent with recent studies (e.g., Romano et al., 2020; Martín–Carbonell et al., 2021; Tan et al., 2021). Furthermore, students who report higher levels of flourishing are more likely to experience study–related positive feelings (e.g., happiness, pleasure, and satisfaction) and less likely to experience study–related negative feelings (e.g., sadness, displeasure, and anger).

The main strength of the present study is the large sample used. However, there are also some limitations that highlighting possible paths for future research. First, we use a convenience sample, and our results should be generalized with caution. Therefore, future research could use a representative and diverse sample to generalize its results to the Chilean high school population. Second, the cross–sectional nature of the design does not allow to prove the temporal stability of the FS. Therefore, future research may include longitudinal designs to analyze their stability and temporal invariance. Third, the use of self–report data may increase the probability of incurring common method variance. Hence, it would be interesting to move toward an external measure of flourishing.

The results suggest that the FS adapted to the Chilean high school context can thus be considered a valid and reliable tool for researchers and practitioners. For researchers, this measure contains only eight items and is, therefore, a short and practical instrument, which offers a broad view of positive and healthy functioning that has been shown to be important for students' optimal functioning. For practitioners, high schools can take advantage of this measure by including it within their diagnosis and monitoring activities. That is, knowing the state of their student's flourishing will allow them to design and deploy properly grounded actions to foster their well–being and contribute to the building of a healthy and thriving school community.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Comité Ético Científico of the Universidad de Tarapacá (CEC-UTA). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

All authors contributed equally to the research design and wrote the manuscript.

REFERENCES

Carmona–Halty, M., and Villegas–Robertson, J. M. (2018). Scale of positive and negativeexperience (SPANE): adaptation and validation in a chilean school context. *Interciencia* 43, 317–321.

Cassaretto, M., and Martínez Uribe, P. (2017). Validation of the scales of well–being of flourishing and feelings. *Pensam. Psicol.* 15, 19–31.

Chamizo–Nieto, M. T., Arrivillaga, C., Rey, L., and Extremera, N. (2021). The role of emotional intelligence, the teacher student–relationship, and flourishing on academic performance in adolescents: a moderated mediation study. *Front. Psychol.* 12:695067. doi: 10.3389/fpsyg.2021.695067

Checa, I., Perales, J., and Espejo, B. (2018). Spanish validation of the flourishing scale in the general population. *Curr. Psychol.* 37, 949–956. doi: 10.1007/s12144-017-9581-0

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. modeling* 14, 464–504. doi: 10.1080/10705510701301834

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

Chen, X., Vallerand, R. J., and Padilla, A. M. (2021). On the role of passion in second language learning and flourishing. *J. Happiness Stud.* 22, 2761–2779. doi: 10.1007/s10902-020-00339-0

Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness–of–fit indexes for testing measurement invariance. *Struct. Equ. Modeling* 9, 233–255. doi: 10.1207/s15328007sem0902_5

da Fonseca, P. N., Nascimento, B. S., Macedo Barbosa, L. H. G., Vione, K. C., and Veloso Gouveia, V. (2015). Flourishing scale: evidence of its suitability to the Brazilian context. *Soc. Inq. Well–being* 1, 33–40.

Datu, J. A. D. (2018). Flourishing is associated with higher academic achievement and engagement in Filipino undergraduate and high school students. *J. Happiness Stud.* 19, 27–39. doi: 10.1007/s10902-016-9805-2

Datu, J. A. D., Labarda, C. E., and Salanga, M. G. C. (2019). Flourishing is associated with achievement goal orientations and academic delay of gratification in a collectivist context. *J. Happiness Stud.* 21, 1171–1182. doi: 10.1007/s10902-019-00122-w

De la Fuente, R., Parra, A., and Sánchez–Queija, I. (2017). Flourishing scale and measurement invariance between two samples of Spanish university students. *Eval. Health Prof.* 40, 409–424. doi: 10.1177/0163278717703446

Didino, D., Taran, E. A., Barysheva, G. A., and Casati, F. (2019). Psychometric evaluation of the Russian version of the flourishing scale in a sample of older adults living in Siberia. *Health Qual. Life Outcomes* 17:34. doi: 10.1186/s12955-019-1100-6

Diener, E., Wirtz, D., Biswas–Diener, R., Tov, W., Kim–Prieto, C., Choi, D., et al. (2009). "New measures of well–being," in *Assessing Well–Being: The Collected Works of Ed Diener*, ed. E. Diener (Berlin: Springer), 247–266. doi: 10.1007/978-90-481-2354-4

Diener, E., Wirtz, D., Kim–Prieto, C., Choi, D., Oishi, S., and Biswas–Diener, R. (2010). New well–being measures: short scales to assess flourishing and positive and negative feelings. *Soc. Indic. Res.* 97, 143–156. doi: 10.1007/s11205-009-9493-y

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Meas. Eval. Couns. Dev.* 43, 121–149. doi: 10.1177/0748175610373459

Garzón–Umerenkova, A., de la Fuente, J., Amate, J., Paoloni, P. V., Fadda, S., and Pérez, J. F. (2018). A linear empirical model of self–regulation on flourishing, health, procrastination, and achievement, among university students. *Front. Psychol.* 9:536. doi: 10.3389/fpsyg.2018.00536

Giuntoli, L., Ceccarini, F., Sica, C., and Caudek, C. (2017). Validation of the Italian versions of the flourishing scale and of the scale of positive and negative experience. *SAGE Open* 7, 1–12. doi: 10.1177/2158244016682293

Herrera, D., Matos, L., Gargurevich, R., Lira, B., and Valenzuela, R. (2021). Context matters: teaching styles and basic psychological needs predicting flourishing and perfectionism in university music students. *Front. Psychol.* 12:623312. doi: 10.3389/fpsyg.2021.623312

Holliman, A. J., Waldeck, D., Jay, B., Murphy, S., Atkinson, E., Collie, R. J., et al. (2021). Adaptability and social support: examining links with psychological wellbeing among UK students and non–students. *Front. Psychol.* 12:636520. doi: 10.3389/fpsyg.2021.636520

Hone, L., Jarden, A., and Schofield, G. (2014). Psychometric properties of the flourishing scale in a New Zealand sample. *Soc. Indic. Res.* 119, 1031–1045. doi: 10.1007/s11205-013-0501-x

Howell, A. J., and Buro, K. (2015). Measuring and predicting student well–being: further evidence in support of the flourishing scale and the scale of positive and negative experiences. *Soc. Indic. Res.* 121, 903–915. doi: 10.1007/s11205-014-0663-1

Huppert, F. A. (2009). Psychological well–being: evidence regarding its causes and consequences. *Appl. Psychol. Health Well–being* 1, 137–164. doi: 10.1111/j.1758-0854.2009.01008.x

Huppert, F. A., and So, T. T. C. (2013). Flourishing across Europe: application of a new conceptual framework for defining well–being. *Soc. Indic. Res.* 110, 837–861. doi: 10.1007/s11205-012-0030-z

Jacobs, G. M., and Renandya, W. A. (2019). "Positive education: a new way to look at learning," in *Student Centered Cooperative Learning: Linking Concepts in Education to Promote Student Learning*, (Berlin: Springer). 99–110. doi: 10.1007/978-981-13-7213-1_8

Keyes, C. L. M. (2002). The mental health continuum: from languishing to flourishing in life. *J. Health Soc. Behav.* 43, 207–222. doi: 10.2307/3090197

Lin, C. (2015). Validation of the psychological well–being scale for use in Taiwan. *Soc. Behav. Pers.* 43, 867–874. doi: 10.2224/sbp.2015.43.5.867

Martín–Carbonell, M., Espejo, B., Checa, I., and Fernández–Daza, M. (2021). Adaptation and measurement invariance by gender of the flourishing scale in a colombian sample. *Int. J. Environ. Res. Public Health* 18:2664. doi: 10.3390/ijerph18052664

Muñiz, J., Elosua, P., and Hambleton, R. K. (2013). International test commission guidelines for test translation and adaptation: second edition. *Psicothema* 25, 151–157. doi: 10.7334/psicothema2013.24

Muthén, L.K, and Muthén, B.O (1998/2017). *Mplus User's Guide*, 8th Edn. (Los Angeles, CA: Muthén and Muthén).

Oh, H., Banawa, R., Zhou, S., Smith, L., and Koyanagi, A. (2021). Flourishing and psychotic experiences among college students in the United States: findings from the healthy minds study 2020. *J. Posit. Psychol.* doi: 10.1080/17439760.2021.1975162 [Epub ahead of print].

Ouweneel, E., Le Blanc, P. M., and Schaufeli, W. B. (2011). Flourishing students: a longitudinal study on positive emotions, personal resources, and study engagement. *J. Posit. Psychol.* 6, 142–153. doi: 10.1080/17439760.2011.558847

Romano, I., Ferro, M. A., Patte, K. A., Diener, E., and Leatherdale, S. T. (2020). Measurement invariance of the flourishing scale among a large sample of Canadian adolescents. *Int. J. Environ. Res. Public Health* 17:7800. doi: 10.3390/ijerph17217800

Ryff, C. D., and Singer, B. (1998). The contours of positive human health. *Psychol. Inq.* 9, 1–28. doi: 10.1207/s15327965pli0901_1

Salama–Younes, M. (2017). Psychometric properties of the psychological flourishing scale in an Egyptian setting. *J. Psychol. Africa* 27, 310–315. doi: 10.1080/14330237.2017.1347749

Schoeps, K., Villanueva, L., Prado–Gascó, V. J., and Montoya–Castilla, I. (2018). Development of emotional skills in adolescents to prevent cyberbullying and improve subjective well–being. *Front. Psychol.* 9:2050. doi: 10.3389/fpsyg.2018.02050

Schreiber, J. B. (2017). Update to core reporting practices in structural equation modeling. *Res. Social Adm. Pharm.* 13, 634–643. doi: 10.1016/j.sapharm.2016.06.006

Seligman, M. E. P., Ernst, R. M., Gillham, J., Reivich, K., and Linkins, M. (2009). Positive education: positive psychology and classroom interventions. *Oxf. Rev. Educ.* 35, 293–311. doi: 10.1080/03054980902934563

Shoshani, A., and Slone, M. (2017). Positive education for young children: effects of a positive psychology intervention for preschool children on subjective well–being and learning behaviors. *Front. Psychol.* 8:1866. doi: 10.3389/fpsyg.2017.01866

Shoshani, A., and Steinmetz, S. (2013). Positive psychology at school: a school–based intervention to promote adolescents' mental health and well–being. *J. Happiness Stud.* 15, 1289–1311. doi: 10.1007/s10902-013-9476-1

Silva, A. J., and Caetano, A. (2013). Validation of the flourishing scale and scale of positive and negative experiences in Portugal. *Soc. Indic. Res.* 110, 469–478. doi: 10.1007/s11205-011-9938-y

Singh, K., Junnarkar, M., and Jaswal, S. (2016). Validating the flourishing scale and the scale of positive and negative experience in India. *Ment. Health Relig. Cult.* 19, 943–954. doi: 10.1080/13674676.2016.1229289

Steinmayr, R., Heyder, A., Naumburg, C., Michels, J., and Wirthwein, L. (2018). School-related and individual predictors of wellŰbeing and academic achievement. *Front. Psychol.* 9:2631. doi: 10.3389/fpsyg.2018.02631

Su, T., Tian, L., and Huebner, E. S. (2019). The reciprocal relations among prosocial behaviour, satisfaction of relatedness needs at school, and subjective well-being in school: a threewave cross-lagged study among Chinese elementary school students. *Curr. Psychol.* doi: 10.1007/s12144-019-00323-9

Tan, Q., Zhang, L., Li, W., and Kong, F. (2021). Longitudinal measurement invariance of the flourishing scale in adolescents. *Curr. Psychol.* 40, 5672–5677. doi: 10.1007/s12144-021-01754-z

The Jamovi Project. (2020). *Jamovi (version 1.8.1) [Computer Software]*. https://www.jamovi.org. [accessed April 5, 2021]

Vallejo–Medina, P., Gómez–Lugo, M., Marchal–Bertrand, L., Saavedra–Roa, A., Soler, F., and Morales, A. (2017). Developing guidelines for adapting questionnaires into the same language in another culture. *Ter. Psicol.* 35, 159–172.

Villieux, A., Sovet, L., Jung, S., and Guilbert, L. (2016). Psychological flourishing:validation of the French version of the flourishing scale and exploration of itsrelationship with personality traits. *Pers. Indiv. Diff.* 88, 1–5. doi: 10.1016/j.paid.2015.08.027

Widlund, A., Tuominen, H., and Korhonen, J. (2018). Academic well-being, mathematics performance, and educational aspirations in lower secondary education: changes within a school year. *Front. Psychol.* 9:2027. doi: 10.3389/fpsyg.2018.00297x

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

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

Copyright © 2022 Carmona-Halty, Marín-Gutierrez, Mena-Chamorro, Sepulveda-Páez and Ferrer-Urbina. 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.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fpsyg.2022.795478

An Evaluation of the Psychometric Properties of the Temporal Satisfaction With Life Scale

Joline Guitard[1], *Aaron Jarden*[2]*, *Rebecca Jarden*[3] *and Denis Lajoie*[1]

[1] School of Psychology, Université de Moncton, Moncton, NB, Canada, [2] Center for Wellbeing Science, Melbourne Graduate School of Education, University of Melbourne, Melbourne, VIC, Australia, [3] Department of Nursing, Melbourne School of Health Science, University of Melbourne, Melbourne, VIC, Australia

The Temporal Satisfaction with Life Scale measures judgements of life satisfaction using 15 items, according to three temporal dimensions: past, present, and future. However, only seven studies have looked at the psychometric properties of the Temporal Satisfaction with Life Scale, and this has been individually across vastly different countries and cultures (Canada, China, Germany, Spain, Switzerland, Turkey, and United-States), and with different populations, such as undergraduate students, adults, and older adults. In addition, these studies have highlighted issues regarding the replicability of the validity of the scale structure and optimal number of items. In this study we use a large international and multicultural sample (n = 6,912) from the International Wellbeing Study and investigate the scale structure of the Temporal Satisfaction with Life Scale, resulting in the recommendation that a shortened 12-item version provides a better model fit compared to the original 15-item version. More in-depth correlates with aspects of wellbeing and illbeing, in relation to past, present, and future life satisfaction, are also presented than have been previously, which found positive correlations between the temporal dimensions of the Temporal Satisfaction with Life Scale and wellbeing, as well as negative correlations with illbeing measures.

Keywords: life satisfaction, assessment, psychometric, wellbeing, positive assessment

OPEN ACCESS

Edited by:
Arianna Costantini,
University of Trento, Italy

Reviewed by:
Marcos Carmona-Halty,
Universidad de Tarapacá, Chile
Ana Maria Hernandez Baeza,
University of Valencia, Spain
Laurent Sovet,
Université de Paris, France

***Correspondence:**
Aaron Jarden
aaron.jarden@unimelb.edu.au

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

Received: *15 October 2021*
Accepted: *21 March 2022*
Published: *14 April 2022*

Citation:
Guitard J, Jarden A, Jarden R and Lajoie D (2022) An Evaluation of the Psychometric Properties of the Temporal Satisfaction With Life Scale. Front. Psychol. 13:795478. doi: 10.3389/fpsyg.2022.795478

INTRODUCTION

Life satisfaction is amongst the most used concepts to assess subjective wellbeing (SWB). In fact, some authors use life satisfaction and subjective wellbeing as interchangeable synonyms (i.e., Bertoni and Corazzini, 2018). Life satisfaction can be defined as a cognitive evaluation of one's overall satisfaction with their current life, relative to one's own criteria regarding what a satisfactory life means (Diener et al., 1985). The most widely used measure of life satisfaction is the Satisfaction with Life Scale (SWLS; Diener et al., 1985), which assesses individuals' life satisfaction with five items answered on 7-point Likert scales. From this unidimensional scale, Pavot et al. (1998) developed a multidimensional measure to assess life satisfaction over time: The Temporal Satisfaction with Life Scale (TSWLS; Pavot et al., 1998). The TSWLS measures judgements of life satisfaction using 15 items, according to three temporal dimensions: past, present, and future. To create this new measure, the authors kept the original five items of the SWLS to assess present life satisfaction. Then, to create the past and future dimensions, the authors simply added words like "in the past" or "for the future" to the original five questions. For example, the item "I am satisfied

with my life" which is used to measure present life satisfaction was transformed to "I am satisfied with my life in the past" to measure past life satisfaction and to "I will be satisfied with my life in the future" to measure future life satisfaction (Pavot et al., 1998). Therefore, the temporal aspects of the TSWLS aims at assessing one's current subjective perception of their life in the past, present, and expectations for the future. Pavot (2014) argues that this conceptualization is important as our present wellbeing is likely to be impacted by what we subjectively recall from the past and expect for our future (regardless of if these interpretations and expectations are accurate or not).

Since its creation, the TSWLS has rarely been used in research compared to the SWLS. For example, a search in October of 2021 for the SWLS on PsycInfo yielded 10,886 results, while a search for the TSWLS yielded only 61 results (i.e., for every one time the TSWLS is used in research, the SWLS is used 178 times). However, when assessing life satisfaction, researchers may benefit from the temporal dimensions of the TSWLS as it provides more specific information regarding individual differences in levels and experiences of wellbeing (Pavot et al., 1998). For example, let's imagine two individuals who answer the SWLS, and both get an average score of 20. During data analysis, these two individuals would be considered to have the same level of life satisfaction, even if one of them expects their life satisfaction to be better in the future and the other believes their life satisfaction to be worst in the future. Such differences in one's vision of future life satisfaction could have substantial impacts on research (i.e., misinterpretation of results), especially if life satisfaction is used as an independent variable to predict concepts that imply a future-oriented perspective (e.g., optimism, hope; Pavot et al., 1998). The explanation illustrated by the example above also applies to the past dimension of the TSWLS and concepts that imply a past-oriented perspective (i.e., depression, rumination). Furthermore, differences in how one recalls their past or envisions their future can have impact on motivations and coping strategies (Pavot et al., 1998).

In addition to having been used less often than the SWLS, the TSWLS's psychometric properties have rarely been investigated. Specifically, only eight studies, including the original study by Pavot et al. (1998), have investigated the psychometric properties of this scale, and this has been across vastly different cultures and languages in the different countries of Turkey (Akyurek et al., 2019), Canada (McIntosh's, 2001), China (Ye, 2007), Germany (Trautwein, 2004), Spain (Tomás et al., 2016; Carrillo et al., 2021), Switzerland (Proyer et al., 2011), United-States (Pavot et al., 1998), and in different populations, such as undergraduate students (e.g., McIntosh's, 2001), adults (e.g., Akyurek et al., 2019), and older adults (e.g., Carrillo et al., 2021). Moreover, of those studies, only Pavot et al. (1998) and McIntosh's (2001) were conducted using the original English version; all other studies used various non-English adaptations of the scale (i.e., Chinese, German, Spanish, Swiss, Turkish). It is not explicitly known why there has not been more studies using or studying the TSWLS. However, one possible reason is that the temporal aspect of the scale often requires multiple time-points of assessment and therefore, requires more time and resources than a single time-point assessment done by the SWLS. Nevertheless, as depicted by the example above, neglecting to consider the cognitive component that underlies how individuals see their past and future life satisfaction, even in a single time-point assessment, could lead to wrongful conclusions if other study variables have an implied temporal component.

In terms of the TSWLS structure, all studies (except for Tomás et al., 2016-Spanish version) found support for a three-factor structure; factors being past, present, and future life satisfaction. Tomás et al. (2016) instead reported finding a bifactor model comprising one general dimension of satisfaction with life and three specific factors for past, present, and future life satisfaction. A more recent study of Spanish speaking individuals was conducted by Carrillo et al. (2021) who performed their own translation of the original scale into Spanish. Their results supported a three-factor structure which includes all 15 original items. The Turkish adaptation of the scale was found to have a better fit with the first item of each five-item subscale removed, resulting in a 12-item measure (Akyurek et al., 2019), whereas the Chinese adaptation was found to have a better fit with the first and last items of each 5-item subscale removed, with then each subscale resulting in 3-items and in a 9-item measure (Ye, 2007). Thus, it appears there are unresolved issues regarding the replicability of the validity of the scale structure and optimal number of items. Regarding reliability, internal consistency of the scale has been reported as good (Trautwein, 2004) and alpha coefficients have ranged from 0.87 to 0.93 (Pavot et al., 1998; Tomás et al., 2016; Akyurek et al., 2019; Carrillo et al., 2021). For the subscales, good internal consistency has also been reported with alpha coefficients ranging from 0.76 to 0.93 (Pavot et al., 1998; McIntosh's, 2001; Proyer et al., 2011; Tomás et al., 2016; Akyurek et al., 2019; Carrillo et al., 2021). Reports of test-retest reliability have also been respectable (Trautwein, 2004) with correlations between times of measurements of 0.81 (Akyurek et al., 2019) and 0.83 (Pavot et al., 1998).

AIMS AND HYPOTHESES

The current study aims to assess the psychometric properties of the Temporal Satisfaction with Life Scale (TSWLS; Pavot et al., 1998), specifically investigating the scale structure and number of optimal items by testing for measurement invariance between subsets of our sample. More precisely, we will test for measurement invariance between English speakers of different countries as well as between six different translations of the scale. This is in relation to (a) the TSWLS's little use since its creation and little investigation (eight studies) confirming its psychometric properties, (b) the issue of replicability of the validity of the scale structure and optimal item count, and (c) extending the current research base by reporting for the first time on a large and diverse multicultural (rather than single culture) sample.

Regarding the general scale structure and optimal number of items, we are going to test which configuration (the original 15-item three factor structure or a 12-item version) receives support across different cultures given that the 12-item version has been deemed more suitable in several studies. Regarding measurement

TABLE 1 | English-speaking participants' world region.

World region	Relative frequency (%)	Count
Oceania	48.4	1,928
North America	24.0	956
Europe	19.6	779
Asia	5.4	216
Africa	1.3	50
Middle East	0.8	31
Latin America	0.6	22
Total	100%	3,982

Latin America includes countries from Central and South America as well as the Caribbeans.

invariance, we expect to find strict invariance of the English version and at least configural invariance between the different translations. As the TSWLS is a measure of life satisfaction, we expect all three subscales to positively correlate with aspects of wellbeing (strengths use and knowledge, subjective happiness, gratitude, hope, and the presence of meaning in life) and to negatively correlate with aspects of illbeing (search for meaning in life, rumination, depression).

METHOD

Participants

The current study used data from the International Wellbeing Study (IWS; www.wellbeingstudy.com). Our sample in the current study consists of 6,912 individuals who completed the assessment battery at the first assessment timepoint of this five timepoint longitudinal study, with participants who did not complete the whole battery at this timepoint excluded and further timepoint data not included in this analysis. Individuals came from various countries and cultures such that 3,982 were English-speaking participants who completed the English version of the survey, and 2,930 were non-English speaking participants who completed translated versions of the survey in their respective languages. English speaking participants came from 89 different countries and were grouped by their world region for the purpose of analyses. Most of the English-speaking participants were from the regions of Oceania, North America, and Europe as depicted in **Table 1**. The distribution and descriptive statistics of non-English-speaking participants according to their language of assessment is available in **Table 2**. Data was collected between March 2009 and March 2013. All participants were over the age of 16 (81.5% female: mean age 37.4 years old, *SD* 14.3).

Materials

The IWS survey battery consisted of 19 questionnaires (217 items in total) and was completed in 29 min on average. The current study uses nine of the 19 questionnaires from the IWS; each of these nine are described below, and the other 10 we did not view as direct illbeing or wellbeing correlates (e.g., Negative Life Event Scale), and also perceived they were not needed. A full list of the IWS survey battery and copy of the survey questions in each language is available on the IWS website. Regarding translations, the English version of the TSWLS was that provided by Pavot et al. (1998). For the International Wellbeing Study this scale was back-translated into 15 languages, including the six used in this study (Chinese, Czech, Finnish, Hungarian, Slovene, and Spanish). In each case the scale was first translated from English into the relevant language by a native speaker of the language, who also had psychology and scale development knowledge. The translation was then independently translated back into English by a second translator, and then the two translators discussed and resolved any inconsistencies in translation. The detailed reliability statistics are available in supplemental material for each translation of the scale, as well as for each English-speaking subsamples used in the current study.

The Temporal Satisfaction with Life Scale (TSWLS; Pavot et al., 1998) measures past, present and future life satisfaction according to 15 items (five per temporal dimension). Items are answered on 7-point Likert scale ranging from 1—*strongly disagree*, to 7—*strongly agree*. Internal consistency was deemed as mostly good with coefficients alpha and omega ranging from 0.80 to 0.87 for past life satisfaction, 0.88 to 0.91 for present life satisfaction, 0.72 to 0.87 for future life satisfaction and 0.87 to 0.94 for the total scale.

The Strengths Use and Current Knowledge scale (Govindji and Linley, 2007) consists of 10 items (five per subscale) measuring the use (e.g., "I always try to use my strengths") and knowledge (e.g., "I know my strengths well") of one's psychological strengths. The scale can also be used to obtain a global score of strengths use and knowledge. Items are answered on a 7-point Likert scale ranging from 1—*strongly disagree* to 7—*strongly agree*. In the current study, internal consistency coefficients ranged from 0.84 to 0.90 for the use subscale, 0.70–0.85 for the knowledge subscale and 0.86–0.92 for the total scale, showing support for acceptable to good internal consistency.

The Subjective Happiness Scale (SHS; Lyubomirsky and Lepper, 1999) is a 4-item measure of global subjective happiness designed to assess how happy individuals consider themselves to be. Items are answered on a 7-point Likert scale, with scale anchors differing across the four items. An example item is: "Compared to most of my peers, I consider myself:" with answers ranging from 1—*less happy* to 7—*more happy*. Internal consistency of this scale in the present study was adequate with coefficients alpha and omega ranging from 0.74 to 0.86, with the exception of the Finnish translation which showed poor internal consistency ($\alpha = 0.49$, $\omega = 0.66$, 95% CI [0.60, 0.72], SE = 0.03).

The Happiness Measure (HM: Fordyce, 1988), also known as the Fordyce Emotion Questionnaire, is a measure of emotional wellbeing that provides an indication of a person's perceived happiness and measures the affective component of SWB. The HM consists of two questions on happiness; the first one assessing how happy the individual usually feels (intensity), while the second is an estimate of the percentages of time respondents feel happy, unhappy, and neutral (frequency). For the purpose of the current study, only the first item was used, which asks respondents: "In general, how happy or unhappy do you usually feel?." Respondents choose one of 11 descriptive answers ranging from (0) "extremely unhappy (utterly depressed, completely

TABLE 2 | Distribution and descriptive statistics of non-English-speaking participants according to their language of assessment.

Language	Count	Mean age (*SD*)	% Females	Most represented countries per language
Hungarian	1,136	31.5 (11.38)	84.2	1,068 (94.01%) living in Hungary
Spanish	693	35.5 (13.43)	76.8	309 (44.59%) living in Mexico; 200 (28.86%) living in Columbia
Finnish	335	20.5 (14.18)	51.6	314 (93.73%) living in Finland
Slovene	288	23.3 (9.76)	83.0	281 (97.57%) living in Slovenia
Czech	250	27.6 (10.97)	81.6	241 (96.4%) living in Czech Republic
Chinese	228	21.8 (6.73)	58.8	205 (89.91%) living in China

n = 2,930.

down)," to (5) "neutral (not particularly happy or unhappy)," to (10) "extremely happy (feeling ecstatic, joyous, fantastic)." With the item used in the current study, the HM measures individuals' perceptions of their intensity of happiness in general. This is in contrast to the SHS which measures more than intensity in capturing a more global and cognitive aspect of happiness (e.g., one's happiness compared to others).

The Gratitude Questionnaire-6 (GQ-6; McCullough et al., 2002) is a unidimensional 6-item measure of the disposition toward gratitude. Items are answered on a Likert scale ranging from 1—*strongly disagree* to 7—*strongly agree*. An example item is: "I am grateful to a wide variety of people." In the current study, the GQ-6 had adequate internal consistency with coefficients alpha and omega ranging between 0.71 and 0.84.

The Adult Hope Scale (AHS; Snyder et al., 1991) measures hope according to two dimensions: agency and pathways. The agency dimension assesses successful goal-directed determination (e.g., "I energetically pursue my goals"), while the pathways dimension assesses the ability to find ways of surmounting obstacles (e.g., "There are lots of ways around any problem"). The scale consists of 12 items: four agency items, four pathway items, and four fillers not related to hope (e.g., "I feel tired most of the time"). Items are answered on an 8-point Likert scale ranging from 1—*definitely false* to 8—*definitely true.* The AHS provides scores for both dimensions (agency and pathways), as well as a global hope score based on the eight hope related items. Internal consistency for the scale in the current study was mostly acceptable with alpha and omega coefficients ranging from 0.72 to 0.84 for the agency subscale, 0.68 to 0.83 for the pathway subscale and 0.78 to 0.88 for the total scale.

The Meaning in Life Questionnaire (MLQ; Steger et al., 2006) measures the presence and the search for meaning in life with 10 items (five per subscale). The presence subscale assesses "'individuals' feelings of living a" meaningful life (e.g., "My life has a clear sense of purpose"), whereas the search subscale assesses "'individuals' motivations to" find or better understand the meaning in their lives (e.g., "I am looking for something that makes my life feel meaningful). Items are answered on a 7-point Likert scale ranging from 1—*absolutely untrue* to 7— *absolutely true*. In the current study, both the presence and search subscales had good internal consistency with alpha and omega coefficients ranging from 0.79 to 0.93 and 0.83 to 0.91, respectively.

The Rumination Scale used in the current study was a 6-item adaptation from the 22-item Ruminative Response Style subscale of the Response Styles Questionnaire (Treynor et al., 2003), specially created for the IWS by Professor Paul Jose. The scale assesses responses to depressive symptoms focusing on their meanings, causes and consequences (Nolen-Hoeksema, 1991). The scale consists of six items prompted with: "In the past 3 months would you say you…". Two items came from the Brooding-related factor (or moody and self-critical pondering, e.g., "thought: 'Why can't I handle things better?"' and four items came from the Depression-related factor (directly tapping into depression symptoms, e.g., "thought: 'Why can't I get going?"'; Treynor et al., 2003). When put together with the prompt, an example of an item would be: "In the past 3 months, would you say you thought: 'Why can't I handle things better?."' Items were answered on a 7-point Likert scale ranging from 1—*strongly disagree* to 7—*strongly agree*. Internal consistency of this scale in the current study was good with coefficients ranging from 0.82 to 0.89 across subsamples.

The Centre for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977) was used to assess the presence of depressive symptoms in the last week, while focusing on the affective component. The unidimensional scale consists of 20 items (e.g., "I felt tearful") that are answered on a 4-point scale ranging from 0—*rarely or none of the time (<1 day)* to 3—*most or all of the time (5–7 days)*. In the current study, the CES-D had excellent internal consistency with alpha and omega coefficients ranging from 0.90 to 0.93.

Procedure

Ethical approval was granted by the *Open Polytechnic of New Zealand* Ethics Committee in 2009. Participants started the study at different times between March 2009 and March 2012. Self-reported questionnaires were completed at ∼3-month intervals (during an open week period) for a total of five assessments over a year. Incentives for participation included a summary report of their scores on the survey, the chance to win one of 15 Amazon.com vouchers (valued at $100 NZD), and the opportunity to take part in one of three online wellbeing classes for free after the first three assessments.

Statistical Analyses

Analyses were conducted with Lavaan 0.6–9 package in R version 4.1.1 (R Core Team, 2022). We adopted the MLR estimator in our analyses. Indeed, all the items in the TSWLS are statistically non-normal in our sample (either in terms of skew, kurtosis, or

TABLE 3 | Model fit using the Oceania sample across 4 models tested.

Model	Robust X2	Df	Robust CFI	Robust TLI	Robust RMSEA	SRMR	BIC	AIC
Model 1	1274.363	87	0.916	0.899	0.084	0.06	94246.358	93979.274
Model 2	985.242	74	0.931	0.916	0.08	0.039	87256.214	87005.823
Model 3	672.553	51	0.942	0.925	0.08	0.035	74038.989	73821.984
Model 4	289.078	39	0.977	0.961	0.058	0.031	73600.289	73316.513

Excellent fit can be defined by Irwing and Hughes' (2018) criteria: Root Mean Square Error of Approximation (RMSEA) ≤ 0.06, Standardized Root Mean Square Residuals (SRMR) ≤ 0.08, Tucker-Lewis Index (TLI) ≥ 0.95, Comparative Fit Index (CFI) ≥ 0.95. Abad et al. (2011) also suggest guidelines regarding acceptable fit criteria: RMSEA ≤ 0.08, SRMR ≤ 0.09, TLI ≥ 0.90, CFI ≥ 0.90. n = 1,928.

both). This may not be cause for concern, as even small deviations in skew and kurtosis can appear to be statistically significant with a large sample, such as ours (Tabachnick and Fidell, 2014, p. 114). Nevertheless, adopting a robust estimator avoids potential issues with non-normality. To take full advantage of our international and multilingual sample, we adopted a multistep analytic strategy. The first step was to optimize model fit in a regional subset of our English sample. We decided to use the English version as the baseline model as it is the original language of the scale (Pavot et al., 1998) and the largest segment of our sample. The second step was to cross-validate the optimized model in the rest of the English sample (i.e., other regions) and to establish measurement invariance within that sample. The third step was to verify measurement invariance in the translations. We verify Configural (factor structure held constant), Metric (factor loadings held constant), Scalar (intercepts held constant) and Strict (error held constant) invariance. In invariance testing, models are nested such that the constraint of previous steps are included in subsequent steps (i.e., Strict invariance implies equivalent factor structure, factor loadings, intercepts, and error for examined groups). We note that the literature on statistical criteria to determine measurement invariance is not yet fully set (see Putnick and Bornstein, 2016 for a review). The usual cut-off is Cheung and Rensvold (2002) CFI criterion for assessing measurement invariance. That is to say that so long as the CFI for the nested model is not worse by more than 0.01, invariance can be claimed. However, other authors have suggested that certain parameters surrounding the tests of invariance have an important influence on the criteria to be used. For example, Rutkowski and Svetina (2014) argue that when multiple groups are compared, more liberal criteria should be used. For instance, in their analyses containing a range of 10–20 groups, they argue that a .02 reduction for CFI and a .03 augmentation for RMSEA would be evidence of invariance. Overall, then, contextualized interpretations of measurement invariance statistical criteria are preferred rather than specific cut-off points.

RESULTS

Scale Structure and Measurement Invariance

Because we wanted to test measurement invariance, we split our English participants by their continent of origin. We retained regions for which at least 200 participants could be identified in our analyses (Oceania, North America, Europe, and Asia). We decided to test and optimize fit in the Oceania sample, the largest group in the English sample, with the knowledge that our results with this group would be cross-validated using the other regional groups. Because we had strong theoretical expectations regarding the overall shape of the model (multiple previous studies having established a three-factor structure), we opted for a Confirmatory Factor Analysis (CFA) framework.

The first model we tested (Model 1) was therefore a three-factor model with all 15 items. As shown in **Table 3**, fit was mediocre. An examination of the factor loadings for this model showed that item 11 was problematic as it did not meet the 0.6 criterion for factor loadings suggested by Awang (2012), nor did it meet the criteria of 0.4 for the R-squared values suggested by the same author. This item was therefore removed. Model 2 results represent the same model as Model 1, minus item 11. Fit remained mediocre, but was slightly improved. At this point, we made the substantive decision to test a model without items 1 and 6. These items reflect the same concept as item 11, except for the past and present timeframes, respectively. This action was motivated by the fact that we judged it important for the very concept of this questionnaire to have comparable and commensurate scores of life satisfaction across timeframes. Moreover, we determined that it would be for the best if a shorter measure contributed to better fit, as this would favor shorter response time for participants. Model 3 presents the results of a three-factor model with items 11, 1, and 6 removed. Fit was improved but did not yet meet all criteria for excellent fit. We therefore continued to examine potential theoretically based modifications that we could make to the model and concluded that it would be necessary to correlate residuals between items referring to similar concepts across timeframes. The practice of correlating residuals is controversial, but some authors have argued that excluding residual correlations that reflect real shared method variance leads to biased factor estimations (e.g., Cole et al., 2007). In the case of the TSWLS, shared method variance is included by design, with similar items used for all three temporal conditions. Model 4 therefore presents the results of a model with items 11, 1, and 6 removed, and with correlated residuals between similar items (e.g., items 2, 7, and 12; and items 3, 8, and 13). As shown in **Table 3**, fit was quite good for this model. We note that we did test a bifactor model (Model 4 with an added general factor with loadings on all items, and orthogonality between latent variables), but that it caused convergence issues in invariance models (and in one of the translation subsamples). We therefore excluded this analysis from our presentation.

TABLE 4 | Cross-validation of the optimized model (Model 4) in the other English samples.

Model	Robust X2	df	Robust CFI	Robust TLI	Robust RMSEA	SRMR	BIC	AIC
English sample	581.923	39	0.976	0.959	0.059	0.033	152935.782	152615.016
Oceania	289.078	39	0.977	0.961	0.058	0.031	73600.289	73316.513
North America	209.316	39	0.97	0.949	0.068	0.05	36811.396	36563.396
Europe	135.135	39	0.979	0.965	0.056	0.03	30244.005	30006.447
Asia	94.392	39	0.95	0.915	0.081	0.044	8426.337	8254.198
Configural fit	739.193	156	0.974	0.957	0.062	0.037	149418.274	148140.554
Metric fit	784.362	183	0.974	0.962	0.058	0.039	149234.872	148126.262
Scalar fit	952.825	210	0.967	0.959	0.06	0.041	149186.307	148246.808
Strict fit	996.624	246	0.967	0.965	0.056	0.042	149012.399	148298.379

We then proceeded to cross-validate the optimized model (Model 4) in the other English samples, and to verify measurement invariance across the respective groups. As shown in **Table 4**, fit was excellent for all of the concerned groups, thereby confirming the structure found during our optimization procedure. We do note that while all residual correlations were statistically significant in the Oceania sample, that is not the case in all English samples. While this could be explained, at least in part, by differences in sample size, this suggests that correlating all relevant residuals might not be the most parsimonious approach to achieving fit in all circumstances. Nevertheless, because the inclusion of these residual correlations is theoretically motivated, we contend that they should be maintained. In spite of differences in statistical significance for residual correlations, strict invariance is found across English samples, as shown in **Table 4**.

Finally, we verified measurement invariance across translations. As a preliminary step to this analysis, we tested the measurement model for each translation. As shown in **Table 5**, acceptable fit is found for all translations except for Finnish, where the TLI is somewhat low (an examination of modification indices for the Finnish sample suggested that fit might benefit from loading item 12 on all three factors). When we turned to the evaluation of measurement invariance, we found that the reduction in CFI at the Metric invariance step was slightly larger than Cheung and Rensvold's (2002) suggested cut-off (the difference in CFI that we report is 0.011). However, because our analysis did include multiple groups, a more liberal cut-off would be justifiable (Rutkowski and Svetina, 2014). In light of this, and of the proximity of our result to Cheung and Rensvold's cut-off, we judge that Metric invariance is an acceptable conclusion. Scalar and Strict invariance are not found across translations. The final measurement model (Model 4), using the data of the full sample, is illustrated in **Figure 1**.

After obtaining the final model and concluded metric invariance between groups, internal consistency of the 12-item version for the total sample (n = 6,662) is as follow: past life satisfaction $\alpha = 0.85$, $\omega = 0.85$, 95% CI [0.85, 0.86], SE = 0.00, present life satisfaction $\alpha = 0.90$, $\omega = 0.90$, 95% CI [0.89, 0.90], SE = 0.00, future life satisfaction $\alpha = 0.89$, $\omega = 0.89$, 95% CI [0.87, 0.90], SE = 0.00.

As an additional verification in the multilingual samples, we explored whether the 12-item model fit the data better than a 15-item model (with the theoretically based residual correlations permitted in both models). Model fit indices for the 15-item model are provided in the supplemental materials. As can be gleaned from the comparison between the table in the supplemental material and **Table 4**, all fit indices except for one are better in the 12-item model than in the 15-item model (the RMSEA for the Finnish sample is slightly better in the 15-item model). Therefore, while it remains possible that the 12-item model with correlated residuals may not fully optimize fit in all samples, it does improve overall fit over a 15-item model for all languages in the study.

Correlates With Wellbeing and Illbeing

To further explore the psychometric properties of the TSWLS's structural model found above, Pearson correlations between factor scores of past, present, and future dimensions of the 12-item TSWLS were computed with aspects of wellbeing and illbeing for the full sample. As we obtained metric invariance between our translations, we believe it is appropriate to combine the full sample all together as we now know that factor structure and loadings are equivalent between versions. The following **Table 6** shows the results of this analysis.

Interpretation of the correlations is based on Cohen's (1988) guidelines where the strength of the correlation is considered small between $r = 0.10$ and $r = 0.29$, medium between $r = 0.30$ and $r = 0.49$ and large when $r \geq 0.50$. All correlations between the TSWLS and its subscales and wellbeing and illbeing measures were found to be significant at the 0.01 level. As indicated in **Table 6**, all but one of the possible 39 correlations with wellbeing and illbeing were either small, medium or large; with four being large. Also, the strongest correlations were with present, then future, then past life satisfaction.

DISCUSSION

The current study aimed to assess the psychometric properties of the Temporal Satisfaction with Life Scale (TSWLS; Pavot et al., 1998), specifically investigating the scale structure and number of optimal items by testing for measurement invariance between subsets of our sample. Furthermore, we aimed to provide

TABLE 5 | Model fit for the measurement model with each language translation.

Model	Robust X2	df	Robust CFI	Robust TLI	Robust RMSEA	SRMR	BIC	AIC
Full sample	1090.071	39	0.974	0.956	0.059	0.037	294231.349	293877.503
English	581.923	39	0.976	0.959	0.059	0.033	152935.782	152615.016
Hungarian	314.306	39	0.958	0.929	0.079	0.051	42846.796	42589.998
Spanish	125.133	39	0.969	0.947	0.056	0.055	26484.409	26252.816
Finnish	153.727	39	0.932	0.885	0.094	0.069	12791.896	12597.375
Slovene	132.349	39	0.945	0.907	0.091	0.087	10607.401	10420.591
Czech	66.05	39	0.980	0.967	0.053	0.031	8928.617	8749.023
Chinese	49.456	39	0.992	0.986	0.034	0.041	8778.385	8603.488
Configural fit	1414.382	273	0.97	0.949	0.065	0.043	264270.548	261828.306
Metric fit	1895.912	327	0.959	0.942	0.07	0.058	264358.569	262285.742
Scalar fit	3140.663	381	0.928	0.912	0.086	0.069	265332.95	263629.538
Strict fit	3860.815	453	0.911	0.909	0.087	0.068	265801.752	264590.892

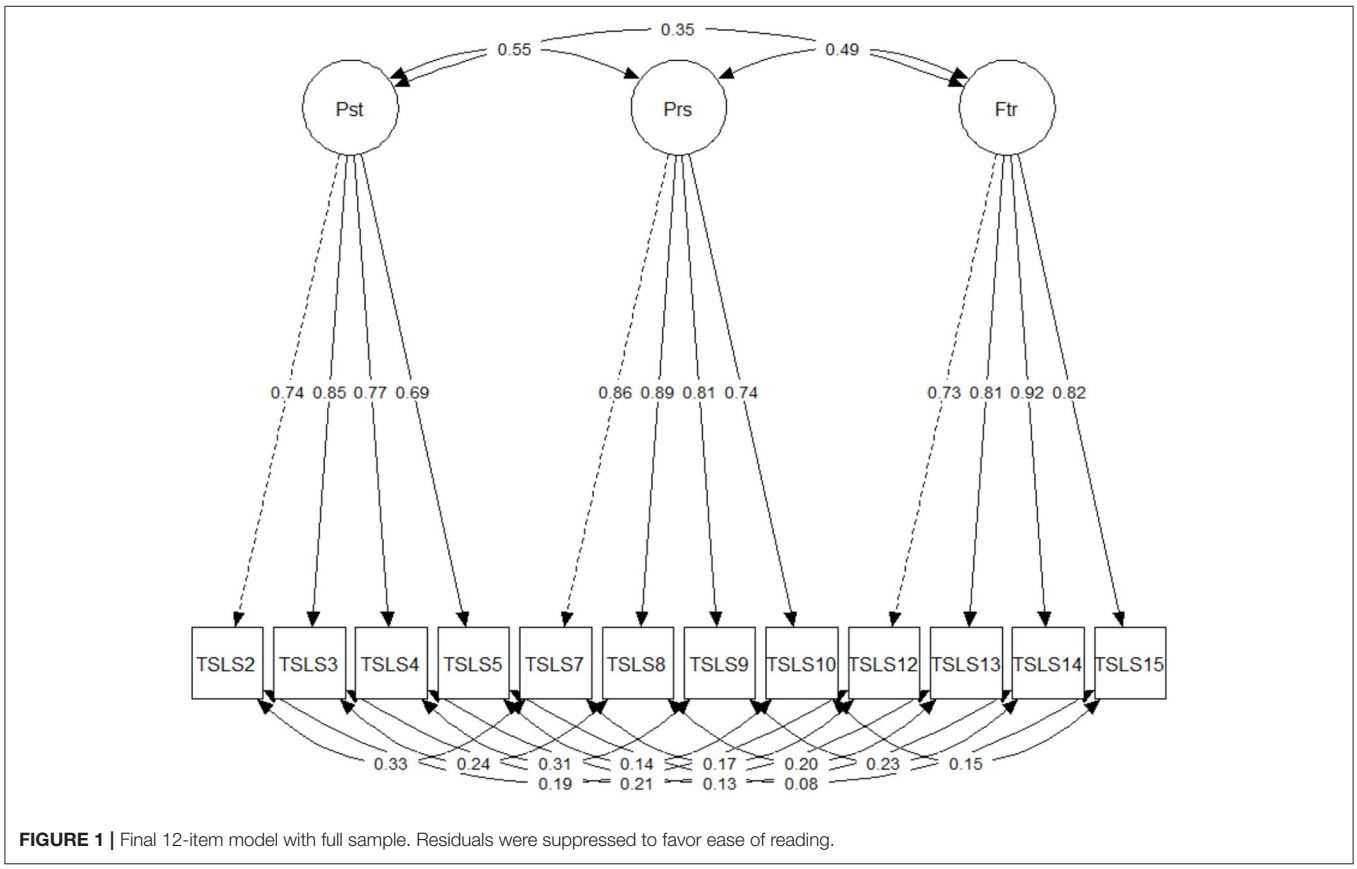

FIGURE 1 | Final 12-item model with full sample. Residuals were suppressed to favor ease of reading.

correlations between temporal dimensions of life satisfaction and aspects of wellbeing and illbeing on a large multicultural sample.

Firstly, we expected to replicate the same three-factor, 15-item structure as the original study by Pavot et al. (1998). This first hypothesis was partially supported by the data as we did find a three-factor structure, but a better fit with 12 items rather than 15. As with past studies, we found item 11 to be quite problematic (McIntosh's, 2001; Ye, 2007; Tomás et al., 2016; Akyurek et al., 2019; Carrillo et al., 2021). After removing item 11, we proceeded to remove items 1 and 6 as well, as they all derive from the same item from the SWLS (Diener et al., 1985). Following the same logic, we also correlated residuals of similar items. Given that the TSWLS past, present, and future dimensions were developed from the same questions (from the SWLS), we believe it is theoretically acceptable and relevant to correlate residuals within this scale.

Secondly, we expected to find strict invariance of the English version of the TSWLS. This hypothesis was supported by the data as factor loadings, factor structure, intercepts and measurement errors were all held constant across participants from Oceania,

TABLE 6 | Correlations between factor scores of the temporal dimensions of the 12-item TSWLS, and wellbeing and illbeing indicators for the full sample.

			Temporal satisfaction with life scale		
Variable	***M***	***SD***	**Past**	**Present**	**Future**
Temporal satisfaction with life scale					
Past (factor scores)	0.00	1.18	1		
Present (factor scores)	0.00	1.35	0.58	1	
Future (factor scores)	0.00	0.96	0.37	0.52	1
Strengths use and current knowledge scale					
Use	26.56	5.38	0.32	0.43	0.40
Knowledge	27.48	4.59	0.28	0.36	0.32
Total	54.04	9.13	0.34	0.44	0.40
Subjective happiness scale					
Subjective happiness	4.84	1.27	0.44	0.59	0.44
Happiness measure					
Emotional wellbeing **(Fordyce)**	7.63	2.18	0.28	0.41	0.30
Gratitude questionnaire					
Gratitude	35.05	5.89	0.32	0.44	0.35
Adult hope scale					
Agency	24.42	4.99	0.40	0.54	0.44
Pathway	24.60	4.61	0.23	0.35	0.38
Total	49.03	8.72	0.35	0.50	0.46
Meaning in life questionnaire					
Presence	25.10	6.85	0.29	0.44	0.39
Search	22.26	7.81	−0.13	−0.21	−0.03
Rumination					
Rumination	25.43	9.08	−0.27	−0.33	−0.20
Center for epidemiological studies depression scale					
Depression	14.69	11.12	−0.36	−0.51	−0.32

n = 6,662.

Correlations were done with the factor scores of the optimized 12-item TSWLS, all languages combined. All correlations were significant at the 0.01 level. Dark gray shade = 0.50 or greater. Light gray shade = 0.30 or greater.

North America, Europe, and Asia. This suggests that the 12-item English version of the TSWLS is equivalent and valid across geographical regions of the world and can be used with any English-speaking individuals regardless of their country of residence.

Thirdly, data supported our expectation to find at least configural invariance between the different translations of the TSWLS. The Finnish translation revealed slight issues with item 12's factor loadings which suggests this version might benefit from more psychometric work. However, we were able to find metric invariance between the translations, but not scalar or strict invariance. In more concrete terms, this means that the translations of the questionnaire function in much the same way (similar factor structure and factor loadings), but that we need more research to disentangle possible differential item functioning across cultures from actual group differences in temporal life satisfaction means (and measurement error). This reflects that life satisfaction, and its temporal aspects might have different base levels in different cultures, but scores between translations and cultures are still generally comparable as they are founded in the same structure.

Fourthly, we expected to find positive correlations between the temporal dimensions of the TSWLS and wellbeing measures as well as negative correlations with illbeing measures, with the better fitting 12-item version. Our second hypothesis is confirmed as all measures correlated in the expected way. In addition to this fact, the correlations presented provide further valuable information for researchers and practitioners regarding the strengths of various correlations with various facets of wellbeing and with different temporal perspectives of life satisfaction. For example, the aspect of self-acceptance has relatively strong correlations with all three temporal dimensions of life satisfaction, whereas one may argue that the presence of meaning aspect is important to present and future life satisfaction and not so much the past. Additionally, some aspects, such as autonomy, are not strongly related to any temporal dimensions of life satisfaction, although may be more strongly associated with other faces of wellbeing beyond life satisfaction. In Pavot et al.'s (1998) article outlining the development of the TSWLS they explain various reasons why the measure was created and when it is beneficial to use rather than the SWLS—for example when taking a developmental focus. Again, these correlations

with wellbeing and illbeing indicators in relation to different temporal perspectives of life satisfaction can aid such work.

As with all research, the present study was subjected to certain limitations. First the sample was mostly female (81.5% female). Second, within the English-speaking participants, the world regions that had enough participants to perform invariance analyses were not so different culturally. More specifically, participants from Oceania, North America and Europe all share a mostly western culture. Invariance of the English version might benefit from invariance analyses between cultures that are more different from one another (i.e., Middle East and North America). Third, from the 15 different TSWLS translations available through the IWS database, only five had enough participants to perform analyses upon. Therefore, the metric invariance found in the current study can only be extended to the Chinese, Finnish, Hungarian, Slovene, and Spanish versions.

Following the results of this study, we suggest future research use a shortened 12-item version of the TSWLS with items 1, 6, and 11 removed. Insofar as language is associated with culture and that cultural differences might explain group differences on temporal life satisfaction, we find that the use of a 12-item, three-factor structure with correlated residuals between related items is a generally appropriate measurement model for the TSWLS.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are not openly available due to ethical constraints, however may be made available from the corresponding author upon reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Open Polytechnic of New Zealand Ethics Committee. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JG: project administration, conceptualization, methodology, investigation, formal analysis, writing—original draft, and writing—review and editing. AJ: data curation, conceptualization, methodology, investigation, formal analysis, writing—original draft, and writing—review and editing. RJ and DL: formal analysis, writing—original draft, and writing—review and editing. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors acknowledge the additional core researchers in the development of the International Wellbeing Study data—Prof. Todd Kashdan, Prof. Paul Jose, Prof. Ormond Simpson, Dr. Kennedy McLachlan, and Dr. Alexander MacKenzie.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.795478/full#supplementary-material

REFERENCES

Abad, F. J., Olea, J., Ponsoda, V., and García, C. (2011). *Medición en ciencias sociales y de la salud [Measuring in social and Health Sciences]*. Madrid: Síntesis.

Akyurek, G., Efe, A., and Aki, E. (2019). Psychometric properties of the temporal satisfaction with life scale among Turkish people. *Psychiatry Clin. Psychopharmacol.* 29, 204–211. doi: 10.1080/24750573.2018.1445898

Awang, Z. (2012). "Validating the measurement model: CFA," in Z. Awang, editor. *A Handbook on SEM* (Bangi: MPWS Publisher), 53–74.

Bertoni, M., and Corazzini, L. (2018). Asymmetric affective forecasting errors and their correlation with subjective well-being. *PLoS ONE*. 13. doi: 10.1371/journal.pone.0192941

Carrillo, A., Etchemendy, E., and Baños, R. M. (2021). Past, present, and future life satisfaction: the role of age, positive and negative mood. *Curr. Psychol.* 40, 629–638. doi: 10.1007/s12144-018-9981-9

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

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd Edn.* Hillside, NJ: Lawrence Erlbaum Associates.

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

Diener, E., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The satisfaction with life scale. *J. Pers. Assess.* 49, 71–75. doi: 10.1207/s15327752jpa4901_13

Fordyce, M. W. (1988). A review of research on The Happiness Measures: a sixty second index of happiness and mental health. *Soc. Indic. Res.* 20, 63–89. doi: 10.1007/BF00302333

Govindji, R., and Linley, P. A. (2007). Strengths use, self-concordance and well-being: implications for strengths coaching and coaching psychologists. *Int. Coaching Psychol. Rev.* 2, 143–153. doi: 10.1037/t,54221-000

Irwing, P., and Hughes, D. J. (2018). "Test development," in *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development., Vols. 1–2,* eds P. T. Irwing, D. Booth, J. Hughes (Chichester: Wiley Blackwell), 3–47. doi: 10.1002/9781118489772.ch,1

Lyubomirsky, S., and Lepper, H. S. (1999). A measure of subjective happiness: preliminary reliability and construct validation. *Soc. Indic. Res.* 46, 137–155. doi: 10.1023/A:1006824100041

McCullough, M. E., Emmons, R. A., and Tsang, J.-A. (2002). The grateful disposition: a conceptual and empirical topography. *J. Pers. Soc. Psychol.* 82, 112–127. doi: 10.1037/0022-3514.82.1.112

McIntosh, C. N. (2001). Report on the construct validity of the temporal satisfaction with life scale. *Soc. Indic. Res.* 54, 37–56. doi: 10.1023/A:1007264829700

Nolen-Hoeksema, S. (1991). Responses to depression and their effects on the duration of depressive episodes. *J. Abnorm. Psychol.* 100, 569–582. doi: 10.1037/0021-843X.100.4.569

Pavot, W. (2014). "Temporal satisfaction with life scale (TSWLS)," in *Encyclopedia of Quality of Life and Well-Being Research*, eds A. C. Michalos (Dordrecht: Springer). doi: 10.1007/978-94-007-0753-5_2993

Pavot, W., Diener, E., and Suh, E. (1998). The temporal satisfaction with life scale. *J. Pers. Assess.* 70, 340–354. doi: 10.1207/s15327752jpa7002_11

Proyer, R. T., Gander, F., Wyss, T., and Ruch, W. (2011). The relation of character strengths to past, present, and future life satisfaction among German-speaking women. *Appl. Psychol. Health Well-Being* 3, 370–384. doi: 10.1111/j.1758-0854.2011.01060.x

Putnick, D. L., and Bornstein, M. H. (2016). Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Dev. Rev.* 41, 71–90. doi: 10.1016/j.dr.2016.06.004

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Appl. Psychol. Measure.* 1, 385–401. doi: 10.1177/014662167700100306

Rutkowski, L., and Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educ. Psychol. Meas.* 74, 31–57. doi: 10.1177/0013164413498257

Snyder, C. R., Harris, C., Anderson, J. R., Holleran, S. A., Irving, L. M., Sigmon, S. T., et al. (1991). The will and the ways: development and validation of an individual-differences measure of hope. *J. Pers. Soc. Psychol.* 60, 570–585. doi: 10.1037/0022-3514.60.4.570

Steger, M. F., Frazier, P., Oishi, S., and Kaler, M. (2006). The meaning in life questionnaire: assessing the presence of and search for meaning in life. *J. Couns. Psychol.* 53, 80–93. doi: 10.1037/0022-0167.53.1.80

Tabachnick, B. G., and Fidell, L. S. (2014). *Using Multivariate Statistics (6th Edn.)*. Boston; Pearson Education, Limited.

Tomás, J. M., Galiana, L., Oliver, A., Sancho, P., and Pinazo, S. (2016). Validating the temporal satisfaction with life scale in Spanish elderly: a bifactor model application. *Behav. Psychol.* 24, 555–567.

Trautwein, U. (2004). Die temporalen Facetten der Lebenszufriedenheit: Eine deutsche Adaptation der Skala von Pavot, Diener und Suh (1998) = The temporal facet of life satisfaction A German adaptation of the scale of Pavot, Diener and Suh (1998). *Diagnostica* 50, 182–192. doi: 10.1026/0012-1924.50.4.182

Treynor, W., Gonzalez, R., and Nolen-Hoeksema, S. (2003). Rumination reconsidered: a psychometric analysis. *Cognit. Ther. Res.* 27, 247–259. doi: 10.1023/A:1023910315561

Ye, S. (2007). Validation of the temporal satisfaction with life scale in a sample of Chinese university students. *Soc. Indic. Res.* 80, 617–628. doi: 10.1007/s11205-006-0010-2

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

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

Copyright © 2022 Guitard, Jarden, Jarden and Lajoie. 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.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 06 May 2022
doi: 10.3389/fpsyg.2022.795770

Further Investigation of the Dimensionality of the Questionnaire for Eudaimonic Well-Being

Amanda Cromhout[1*†], *Lusilda Schutte*[1†], *Marié P. Wissing*[1†] *and Willem D. Schutte*[2†]

[1] *Africa Unit for Transdisciplinary Health Research, Faculty of Health Sciences, North-West University, Potchefstroom, South Africa,* [2] *Centre for Business Mathematics and Informatics, Faculty of Natural and Agricultural Sciences, North-West University, Potchefstroom, South Africa*

OPEN ACCESS

Edited by:
Arianna Costantini, University of Trento, Italy

Reviewed by:
Carmen Ximenez, Autonomous University of Madrid, Spain
Carlos Salavera Bordás, University of Zaragoza, Spain

***Correspondence:**
Amanda Cromhout
cromhout.amanda@gmail.com

†ORCID:
Amanda Cromhout
orcid.org/0000-0002-0008-8212
Lusilda Schutte
orcid.org/0000-0002-2107-3669
Marié P. Wissing
orcid.org/0000-0002-0445-6246
Willem D. Schutte
orcid.org/0000-0001-7551-2009

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

Received: *15 October 2021*
Accepted: *11 April 2022*
Published: *06 May 2022*

Citation:
Cromhout A, Schutte L, Wissing MP and Schutte WD (2022) Further Investigation of the Dimensionality of the Questionnaire for Eudaimonic Well-Being. Front. Psychol. 13:795770. doi: 10.3389/fpsyg.2022.795770

The dimensionality of the Questionnaire for Eudaimonic Well-Being (QEWB) has been a topic of debate and divergent findings in the literature up to date. This study investigated the factor structure and measurement invariance of the QEWB in four culturally diverse South African samples using confirmatory factor analysis (CFA), bifactor CFA, exploratory structural equation modelling (ESEM), and bifactor ESEM. Three student samples completed the English (n = 326), Afrikaans (n = 478), or Setswana (n = 260) version of the QEWB. An adult sample (n = 262) completed the English version. The one-factor structure revealed poor fit for the student samples. Although the four-factor models generally showed slightly better fit than the three-factor models, the latter was preferred for parsimony. The bifactor ESEM model displayed good fit for the student samples, with the general factor and some specific factors attaining sufficient reliability scores, pointing to the potential use of the scale in these samples. Configural invariance between the student samples was supported, but not metric nor scalar invariance. For the adult sample, none of the models displayed good fit and the use of the QEWB in this sample is not recommended. The results point towards the existence of a global eudaimonic well-being factor and, at the same time, the interrelatedness of facets of eudaimonic well-being. It suggests that eudaimonic well-being may be represented by the same items across the three student groups. The influence of developmental phase on the manifestation and measurement of eudaimonic well-being should be explored in future.

Keywords: eudaimonic well-being, factorial validity, reliability, dimensionality, measurement invariance, bifactor ESEM

INTRODUCTION

Eudaimonic well-being (EWB), together with hedonic well-being (HWB), are the main perspectives on well-being (Huta and Waterman, 2014) in the literature on psychosocial well-being, which is core in counselling theory and practice. Whereas HWB is mainly characterised by experiencing increased levels of positive emotions, reduced levels of negative emotions, and increased levels of life satisfaction (Diener, 1984; Waterman, 1993; Diener et al., 2017), EWB is conceptualised differently by different authors and generally includes reference to living or functioning well (see Martela and Sheldon, 2019). For example, Ryff (1989) discerned self-acceptance, personal growth, autonomy, positive relationships, environmental mastery, and purpose in life as elements

of psychological well-being. The mental health continuum model (Keyes, 2002) specifies that the eudaimonic facet of positive mental health includes psychological well-being (as defined by Ryff, 1989) together with social well-being (comprising social coherence, social actualisation, social integration, social acceptance, and social contribution; Keyes, 1998). Martela and Sheldon (2019) indicated that at least 63 elements of EWB are used in about 45 operationalisations thereof. These include elements such as authenticity, emotional stability, mindfulness, optimism, resilience, and self-actualisation. All of these facets may be relevant in the enhancement of mental health, and the availability of valid and reliable measures based on sound theory is necessary to evaluate the outcomes of counselling interventions and growth. The conceptualisation and operationalisation of EWB by Waterman et al. (2010), which is described in the next paragraph, is relevant to this study.

Conceptualisation of Eudaimonic Well-Being Informing the Questionnaire for Eudaimonic Well-Being

Waterman et al. (2010) postulated that EWB should be conceptualised based on the (then) current philosophical understandings of eudaimonic functioning, and discerned six interlinked categories which have strong associations with philosophy and psychology. The categories are: *self-discovery*, which is important for progression towards self-actualisation (and thus for experiencing EWB) and involves identifying who one is; *perceived development of one's best potential*, which involves identifying and actively developing the unique potential that is representative of the best that one can become; *a sense of purpose and meaning in life*, which involves deciding towards which personally meaningful objectives one's talents and skills will be directed; *investment of significant effort in pursuit of excellence*, which refers to individuals' tendency to invest more effort in activities that they find personally meaningful than in other activities; *intense involvement in activities*, which refers to the intensity of the level of involvement in activities that individuals deem personally meaningful compared to their level of involvement in other activities; and *enjoyment of activities as personally expressive*, which refers to individuals' involvement in activities that are expressive of who they are.

This conceptualisation of EWB includes both the objective and subjective elements of EWB (Waterman et al., 2010). The objective elements refer to the behaviours that are related to the pursuit of eudaimonic goals. The subjective elements refer to the experiences of individuals when they are committed to excellence in the actualisation of their personal potential. These subjective experiences of eudaimonia are called "feelings of personal expressiveness" and are typically associated with the pursuit of one's life purpose and the development of one's potential (Waterman et al., 2010, p. 42). Subjective feelings of personal expressiveness are different from subjective well-being (hedonia) in that the latter is a desired outcome in itself, while the former flows from the pursuit of life purpose and the development of potential (Waterman et al., 2010).

In order to test theoretical conceptualisations of EWB, to measure facets and levels of EWB, and to evaluate interventions aimed at enhancing EWB, it is important that psychometrically sound measures are used for this purpose. The Questionnaire for Eudaimonic Well-Being (QEWB; Waterman et al., 2010) is one measure of EWB and has been applied in several recent studies (e.g., Karaś and Chieciuch, 2018; Kimiecik et al., 2019; Sotgiu, 2019).

The Questionnaire for Eudaimonic Well-Being

The Questionnaire for Eudaimonic Well-being (Waterman et al., 2010) measures EWB in terms of the conceptualisation of EWB by Waterman et al. (2010). Although six interlinked categories are discerned in this conceptualisation (Waterman et al., 2010), scale items were not assigned to the specific categories when the scale was constructed (Schutte et al., 2013; Klym-Guba and Karaś, 2018). For the purpose of scale construction, EWB was considered to be a unidimensional construct, where the six categories represent aspects of EWB (Klym-Guba and Karaś, 2018). Waterman et al. (2010) used parcelling and found support for a unifactorial structure in two ethnically diverse American student samples with Cronbach's alpha values of 0.86 and 0.85, respectively. They also found support for convergent, discriminant, construct, and incremental validity. However, Schutte et al. (2013) questioned the use of parcelling and contended that the unidimensionality assumption within parcels was not tested and was likely not to have held. Applying confirmatory factor analysis (CFA) and exploratory factor analysis (EFA) to data from a multicultural South African student group, they found support for a three-factor structure [Sense of Purpose (Cronbach's $\alpha = 0.77$), Purposeful Personal Expressiveness (Cronbach's $\alpha = 0.73$), Effortful Engagement (Cronbach's $\alpha = 0.61$)] and a four-factor structure [Sense of Purpose (Cronbach's $\alpha = 0.77$), Engagement in Rewarding Activities (Cronbach's $\alpha = 0.51$), Living from Beliefs (Cronbach's $\alpha = 0.71$), Effortful Engagement (Cronbach's $\alpha = 0.61$)], thereby pointing towards the multidimensionality of the QEWB. Support was found for convergent and discriminant validity. Schutte et al. (2013) suggested that, although the four-factor solution explained slightly more variance than the three-factor solution, the three-factor solution was preferable in their sample for the sake of parsimony.

Subsequently, both the unidimensionality and the multidimensionality of the QEWB have been supported in recent studies. Applying CFA, Areepattamannil and Hashim (2017) found support for the unidimensionality of the QEWB in an Indian adolescent sample, reporting a Cronbach's alpha value of 0.87. Sotgiu et al. (2019) applied Rasch-analysis to the Italian version of the QEWB in an Italian adult sample and also found support for a unidimensional structure. They reported a separation reliability R of 0.78 and a Cronbach's alpha value of 0.81. Fadda et al. (2017) indicated that a unidimensional structure did not fit the data of the Italian version of the QEWB in Italian student samples. Applying bifactor ESEM to the three- and four-factor solutions found by Schutte et al. (2013), Fadda et al. (2017)

found that the three-factor solution with one general EWB factor revealed superior fit. They reported model-based omega coefficients of composite reliability, namely general EWB factor (ω = 0.90), Sense of Purpose (ω = 0.97), Purposeful Personal Expressiveness (ω = 0.12), and Effortful Engagement (ω = 0.73), with scores on the general EWB factor correlating as expected with scores on measures of life satisfaction and self-esteem. In a subsequent study, Fadda et al. (2020) applied ESEM and bifactor ESEM to the Spanish version of the QEWB in a Spanish student sample, and found that the three-factor bifactor ESEM model outperformed the three-factor ESEM model. They reported sufficient levels of composite reliability with omega values of 0.97 for the general EWB factor, 0.84 for Sense of Purpose, 0.94 for Purposeful Personal Expressiveness, and 0.93 for Effortful Engagement. The general EWB factor correlated positively with a measure of self-esteem, while the specific factors showed no correlation with self-esteem. Applying CFA, EFA, and ESEM to the Polish translation of the QEWB, Klym-Guba and Karaś (2018) found that the three-factor ESEM model, with the three factors as distinguished by Schutte et al. (2013), adequately fitted the data. They reported Cronbach's alpha values for the general EWB factor (α = 0.71 to 0.86), Sense of Purpose (α = 0.79 to 0.87), Purposeful Personal Expressiveness (α = 0.80 to 0.82), and Effortful Engagement (α = 0.63 to 0.71). Ishii et al. (2022) applied ESEM and bifactor ESEM to the Japanese translation of the QEWB in Japanese samples in different age groups (18–29; 30–49; and 50–69) and found that a four-factor ESEM model was most interpretable for the 18- to 29-year group, while a three-factor ESEM model was most interpretable for the 30- to 49-year group and the 50- to 69-year groups. For all groups the models included the Sense of Purpose, Purposeful Personal Expressiveness, and Effortful Engagement factors. Additionally, a "Deep and Meaningful Engagement" factor was discerned for the 18- to 29-age group.

Note that previous validation studies mostly used student samples (Waterman et al., 2010; Schutte et al., 2013; Fadda et al., 2017, 2020), except Areepattamannil and Hashim (2017) who used an adolescent sample and Ishii et al. (2022) who used Japanese adults in various age groups. Although Klym-Guba and Karaś (2018) and Sotgiu et al. (2019) described their samples as adult samples, the mean age of the adult samples used by Klym-Guba and Karaś (2018) was between 20 and 24 years of age across four samples, which is close to the mean age of the student groups used by Fadda et al. (2017, mean age 20 years), Fadda et al. (2020, mean age 20 years), and Schutte et al. (2013, mean age 21 years); and half of the adult sample used by Sotgiu et al. (2019) with a mean age of 28 years, consisted of students. Effectively, the study by Ishii et al. (2022) is the only study, as far as we could establish, that used mature adult samples to investigate the factor structure of the QEWB. The observation that studies exploring the psychometric properties of the QEWB among adults are limited is particularly important since EWB may be experienced differently across developmental phases. For example, Ryff and Keyes (1995) found in an adult sample, divided into young adults (25–29 years), midlife adults (30–64 years), and older adults (65 years and older), that there were differences among the age groups with regard to purpose in life, personal growth, environmental mastery, autonomy, self-acceptance, and personal relationships. Clarke et al. (2000) found that Canadian older adults (65 years or older) were likely to report a decline in their sense of environmental mastery, personal growth, purpose in life, and positive relationships with others with increasing age.

The measurement invariance of the QEWB has been explored in a few earlier studies. Areepattamannil and Hashim (2017) found support for gender invariance of the QEWB for an Indian adolescent sample. Fadda et al. (2020) also found that the Spanish version of the QEWB was gender invariant in a Spanish student sample. Sotgiu et al. (2019) found that the item measures obtained through Rasch analysis were gender invariant, but not age invariant, for the Italian version of the QEWB in an Italian adult sample. Klym-Guba and Karaś (2018) found support for the invariance of the Polish version of the QEWB across four young adult samples. As far as we could establish, no invariance studies investigated cross-cultural invariance of the scale. This is significant, since culture is fundamental to human behaviour, and should be key to theoretical and empirical investigations of psychological constructs (Matsumoto and Yoo, 2006), including eudaimonic well-being.

Besides the possibility that EWB may manifest differently from culture to culture or across sociodemographic different groups, which may influence the psychometric properties of measures of EWB, the statistical analytical procedures used to explore the dimensionality of a scale can also potentially influence the results. This aspect is addressed in the next paragraph.

Measuring Multidimensional Constructs: Application of Exploratory Structural Equation Modelling and Bifactor Modelling

If statistical analyses that do not account for sources of multidimensionality are applied to model multidimensional constructs, it may result in biased parameter estimates (e.g., Morin et al., 2016a; Howard et al., 2018). For example, CFA is based on the independent cluster model (ICM) that assumes that the cross-loadings of items on non-target factors are exactly zero. However, when cross-loadings are constrained to zero, two sources of construct-relevant multidimensionality may not be accounted for, which may lead to biased parameter estimates (Morin et al., 2016b).

Firstly, scale items are rarely related to a single construct (the target factor) when a scale measures conceptually related constructs and will mostly also have construct-relevant associations with the non-target factors (Howard et al., 2018). When these cross-loadings are disregarded it may impact negatively on goodness-of-fit indices since sources of misspecification may be concealed. The discriminant validity of the factors may also be compromised when artificial multicollinearity is created by biased parameter estimates, and the factors are used in prediction (Howard et al., 2018). In order to account for these cross-loadings, exploratory structural equation modelling (ESEM, Asparouhov and Muthén, 2009) can be applied. With ESEM, EFA is incorporated into the structural equation modelling framework, which allows for models to be

specified according to CFA specifications (thus accounting for target factor loadings), while also accounting for cross-loadings (Morin et al., 2016a; Howard et al., 2018).

Secondly, the scale items used to assess multiple dimensions in a psychometric measure could possibly reflect their specific subscales and more global constructs (Morin et al., 2016a). In such instances, hierarchical (or higher-order) CFA is typically applied (Morin et al., 2016a). Higher-order models hypothesise that multiple factors can combine into one or more higher-order factors. The model is specified by allowing each item to load on its specific subscale (i.e., first-order factor) and each first-order factor to load on a higher-order factor (Morin et al., 2016a). The first-order factor fully mediates the associations between the scale items and the higher-order factor (Morin et al., 2016b; Howard et al., 2018). The first-order factor therefore reflects the variance explained by each first-order factor and the variance explained by the higher-order factor (Morin et al., 2016b). In contrast, bifactor models hypothesise that a unitary global factor, that coexists with some specific factors, directly influences the scale items. The variance that is shared by all the scale items is represented by the global factor and the variance that is shared by a specific subset of scale items is represented by the specific factors (Morin et al., 2016b; Howard et al., 2018). The variance that is attributable to the global and specific factors, respectively, can therefore be separated, while simultaneously estimating the direct relations between scale items and the global and specific factors (Morin et al., 2016b; Howard et al., 2018).

Models that allow for the incorporation of cross-loadings and/or a general factor may display superior fit when constructs are conceptually related and/or hierarchically ordered. This is because the estimates of the global factor may be inflated when cross-loadings are not modelled in bifactor CFA models, and estimates of the cross-loadings may be inflated when the global factor is not modelled in EFA models (Morin et al., 2016a; Howard et al., 2018). Therefore, models like ESEM, bifactor CFA, and bifactor ESEM (Jennrich and Bentler, 2011) can be used.

As explicated in the previous section, the dimensionality of the QEWB has been a contentious issue in the literature up to date, with diverse findings being presented in different studies. In attempts to gain more insight into the dimensionality of the scale, ESEM has been applied to Polish (Klym-Guba and Karaś, 2018) and ESEM and bifactor ESEM to Italian (Fadda et al., 2017) and Spanish (Fadda et al., 2020) samples. All of these samples were European and consisted of students or young adults. More recently, ESEM and bifactor ESEM have also been applied to Japanese (Eastern) adult samples (Ishii et al., 2022). Extending the investigations to other cultural and age groups will provide insight into the dimensionality and manifestations of EWB.

The Present Study

Newer analytical approaches, such as ESEM and bifactor ESEM, can provide insight into the dimensionality of a scale – a matter of particular importance for the QEWB for which divergent findings regarding its dimensionality have been presented in the literature. These methods have been applied to data from European (Polish, Italian, and Spanish) student or young adult samples (Fadda et al., 2017, 2020; Klym-Guba and Karaś, 2018), as well as to Eastern (Japanese) adult samples (Ishii et al., 2022). Since culture may largely influence the way in which psychological constructs such as eudaimonic well-being operate and manifest, it would be important to extend explorations to other, particularly non-Western, contexts. Notably, as far as we could establish no studies have investigated the cross-cultural measurement invariance of the scale. In addition, while age and developmental phase may impact how eudaimonic well-being is experienced and expressed, investigations on the psychometric properties of the QEWB have been done mostly on student or young adult samples. In view of these gaps, the aim of the present study was to provide a substantive illustration of various analytical models, namely CFA, bifactor CFA, ESEM, and bifactor ESEM models, to investigate the dimensionality of the QEWB in four culturally diverse South African samples (three student samples, one adult sample) who completed different language versions of the scale and to investigate measurement invariance across samples with adequate baseline fit.

MATERIALS AND METHODS

Research Design and Participants

A quantitative, cross-sectional survey design was used. Three non-probability student samples (N = 1064) from the various campuses of a South African university completed the research battery in English (Sample 1, n = 326), Afrikaans (Sample 2, n = 478), or Setswana (Sample 3, n = 260). Participants could complete the research battery in their home language, or alternatively in the language they were most comfortable with. Participants who indicated "other" as their home language likely spoke one of the other 11 official languages of South Africa. Setswana is an indigenous African language, and participants who completed this version of the scale were most probably of indigenous African heritage. Afrikaans is a language close to Dutch, and taking this together with the demographic profile of the institution where data were gathered into consideration, the cultural heritage of participants who completed the Afrikaans scale version was probably strongly influenced by Western culture. Of the sample who completed the English version of the scale, 18.7% indicated that Setswana was their home language, while 54.9% picked "other." This suggests that the sample was culturally diverse, but with the majority of participants having an African heritage. Sample 4 was a multicultural non-probability adult sample (n = 262) that was recruited with the snowball method across South Africa. The research battery was completed in English.

All samples had to be 18 years of age or older and have at least a Grade 12 level of education. Additionally, Samples 1, 2, and 3 had to be enrolled as students at the university where the data was collected. The socio-demographic information of participants from each sample is presented in **Table 1**.

Measures

Socio-Demographic Questionnaire

Data on socio-demographic variables such as age, gender, home language, and level of education (the latter for Sample 4) were collected.

TABLE 1 | Socio-demographic profile of participants.

Variable	Sample 1	Sample 2	Sample 3	Sample 4
n	326	478	260	262
Gender				
Male	24.5%	35.8%	32.7%	33.2%
Female	75.5%	64.2%	67.3%	66.4%
Missing	0.9%	0%	0%	0.4%
M_{age} (SD_{age})	21.03 (4.08)	19.79 (3.14)	21.59 (4.59)	40.23 (12.19)
Home language				
English	18.4%	0.4%	21.9%	17.2%
Afrikaans	6.7%	99.2%	0.8%	32.4%
Setswana	18.7%	0%	66.5%	18.7%
Other	54.9%	0.4%	9.6%	14.5%
Missing	1.2%	0%	1.2%	17.2%
Education level[a] (Sample 4)				
Secondary	–	–	–	36.3%
Tertiary	–	–	–	32.4%
Post-graduate	–	–	–	29.8%
Missing	–	–	–	1.5%

[a]*Since Samples 1, 2, and 3 consisted of university students, education level was not assessed for this sample.*
M, mean; SD, standard deviation.

The Questionnaire for Eudaimonic Well-Being

The QEWB (Waterman et al., 2010) consists of 21-items and measures EWB as conceptualised by Waterman et al. (2010). We used a seven-point Likert-type scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Refer to the Introduction for detail on the scale development and previous findings on the psychometric properties of the scale.

Ethical Considerations and Procedure

This study was approved by the Health Research Ethics Committee of the North-West University, South Africa (ethics approval number: NWU 00002-07-A2), and formed part of the FORT3 research project [The prevalence of levels of psychosocial health: Dynamics and relationships with biomarkers of (ill) health in South African social contexts; Wissing, 2008/2012]. Participants gave written informed consent, participated voluntarily in the study, and could withdraw from the study without adverse consequences. Data were handled confidentially, and participants received no incentives for participation.

The data of Samples 1, 2, and 3 were collected during 2012, and the data for Sample 4 were gathered during 2011–2014. For Samples 2 and 3, the QEWB was translated from English into Afrikaans and Setswana, respectively, using a research committee approach (Brislin, 1970; Van de Vijver and Humbleton, 1996; Van de Vijver and Leung, 1997). Scale items were checked for cultural appropriateness. The scale was back-translated into English by independent translators (Brislin, 1970). A research committee, that consisted of academics who spoke Afrikaans or Setswana natively and who were fluent in English, compared the back-translated and original English versions of the scale (Van de Vijver and Humbleton, 1996; Van de Vijver and Leung, 1997). A small pilot sample was asked to determine if the scale items of the translated versions were comprehensible and reflected the meaning of the items in a culturally appropriate manner, as well as to evaluate technical aspects such as the clarity of the format and layout of the research battery.

Data Analysis

Stage 1: Descriptive Statistics of Individual Scale Items

IBM SPSS Statistics 25 was used to calculate the mean, standard deviation, and the univariate skewness and kurtosis of each item of the QEWB for all samples. The psych package (v2.1.9; Revelle, 2021) in R4.0.2 (R Core Team, 2021) was used to calculate Mardia's multivariate skewness and kurtosis statistics.

Stage 2: Factorial Validity

All findings reported for factor analysis were based on analyses done using Mplus Version 8.3 (Putnick and Bornstein, 2017), unless otherwise specified. For all samples, the following models were tested: a one-factor CFA model, as well as the following three- and four-factor models: CFA, bifactor CFA, ESEM, and bifactor ESEM. The three- and four-factor models were based on the factors obtained by Schutte et al. (2013) when they performed exploratory factor analysis (EFA) on data from the scale. We used the robust maximum likelihood (MLR) estimator and applied full information likelihood estimation to handle missing data. For the CFA and bifactor CFA models the cross-loadings were constrained to zero, and for the ESEM and bifactor ESEM models cross-loadings were estimated to be close to, but not exactly, zero. We applied oblique target rotation to the ESEM models and orthogonal target rotation to the bifactor ESEM models (Asparouhov and Muthén, 2009). For both oblique and orthogonal rotations, factor variances were set to one, and for the orthogonal rotation, the factor covariances were set to zero (Asparouhov and Muthén, 2009). The following model fit statistics are reported: the χ^2-statistic, comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and the standardised root mean square residual (SRMR). For the χ^2-statistic, higher *p*-values indicate a closer fit between the hypothesised model and perfect fit (Bollen, 1989; Byrne, 2012). CFI and TLI values closer to 0.95 are representative of good model fit (Hu and Bentler, 1999; Byrne, 2012). RMSEA values smaller than 0.05 represent good model fit, while values up to 0.08 represent reasonable model fit (Byrne, 2012). SRMR values of 0.05 or less represent a well-fitting model (Byrne, 2012). The χ^2-statistic is highly sensitive to sample size, therefore the CFI, TLI, RMSEA, and SRMR were used to interpret model fit. If the best-fitting model displayed inadequate fit, this model was used as the model from which areas of local misfit was explored (Byrne, 2012). Model misfit was identified by considering modification indices (MI) and the expected parameter change (EPC) values, where higher MI and EPC values point towards potential model misfit (Byrne, 2012; Whittaker, 2012). Although MI and EPC values were used to identify areas of misspecification, models were only modified if the changes also made sense on substantive grounds (Byrne, 2012; Whittaker, 2012).

Note that the unbiased SRMR fit index (derived by Maydeu-Olivares, 2017) was also calculated for the CFA and bifactor CFA models due to its superiority to other fit statistics (see Ximénez et al., 2022) using the lavResiduals function of the lavaan package (v0.6-10; Rosseel, 2012) in R4.0.2 (R Core Team, 2021). However, since fitting ESEM and bifactor ESEM models using lavaan is still in its infancy, the unbiased SRMR was not calculated for these models. In terms of interpretation, Shi et al. (2018) proposed that the unbiased SRMR divided by the average R^2 of the items (denoted by $\overline{R^2}$) should be less than 0.05 for models with an acceptable fit.

Stage 3: Internal Consistency Reliability

Microsoft Excel was used to calculate model-based omega coefficients of composite reliability, using the formula applied by Sánchez-Oliva et al. (2017). The formula is

$$\omega = \sum (|\lambda_i|)^2 / (\sum [|\lambda_i|]^2 + \sum \delta_{ii}),$$

where the factor loadings are represented by λ_i, and the error variances by δ_{ii} (McDonald, 1970). Calculations were based on parameter estimates obtained from Mplus output. According to Putnick and Bornstein (2017), the guideline that reliability scores larger than 0.70 or 0.80 indicate acceptable reliability is not suitable for bifactor models (see Putnick and Bornstein, 2017, for an explanation). Instead they suggest that omega values larger than 0.50 are indicative of sufficient reliability for bifactor models.

Stage 4: Measurement Invariance

Mplus Version 8.3 (Putnick and Bornstein, 2017) was used to determine invariance across the different language versions of the QEWB in student Samples 1, 2, and 3 (Sample 4 was not included in invariance analyses, since no baseline model with adequate fit could be obtained). We tested for configural, metric, and scalar invariance (Morin et al., 2016a; Putnick and Bornstein, 2017). No equality constraints are applied when testing for configural invariance (Byrne, 2012). If the factor loadings display the same pattern across the groups, configural invariance is supported (Putnick and Bornstein, 2017). For metric and scalar invariance equality constraints are applied. Factor loadings are constrained to be equivalent across the groups for metric invariance, and factor loadings and intercepts in the case of scalar invariance. If metric or scalar invariance is not supported, the non-equivalent factor loadings and intercepts can be released in order to establish support for partial metric or partial scalar invariance (Putnick and Bornstein, 2017). Non-equivalent factor loadings and intercepts can be identified by considering high MI and EPC values (Byrne, 2012). Differences smaller than 0.01 and 0.015 between the CFI and RMSEA values of the nested models, respectively, indicate measurement invariance (Cheung and Rensvold, 2002; Chen, 2007). The likelihood ratio test, which is based on the difference between the χ^2-statistic of the nested models, is highly sensitive to sample size (Cheung and Rensvold, 2002; Chen, 2007). We reported the results of this test but placed more emphasis on other indicators for decision-making.

RESULTS

Stage 1: Descriptive Statistics of Individual Scale Items

For Sample 1 mean values ranged between 3.99 (*SD* = 2.03; item 3) and 6.28 (*SD* = 1.21; item 15), skewness values ranged between −2.35 (item 15) and 0.4 (item 3), and kurtosis values ranged between −1.37 (item 16) and 6.42 (item 15) for the QEWB-English. For Sample 2 mean values ranged between 4.07 (*SD* = 1.70; item 9) and 6.01 (*SD* = 1.30; item 19), skewness values ranged between −1.73 (item 19) and −0.19 (items 9 and 16), and kurtosis values ranged between −0.98 (item 3) and 2.97 (item 19) for the QEWB-Afrikaans. For Sample 3 mean values ranged between 3.25 (*SD* = 2.05; item 3) and 5.78 (*SD* = 1.39; item 18), skewness values ranged between −1.31 (item 5) and 0.15 (item 20), and kurtosis values ranged between −1.36 (item 20) and 1.29 (item 6) for the QEWB-Setswana. For Sample 4 mean values ranged between 3.83 (*SD* = 1.98, item 3) and 5.89 (*SD* = 1.25; item 15), skewness values between −1.38 (item 15) and 0.20 (item 3), and kurtosis values between −1.15 (item 20) and 2.18 (item 15) for the QEWB-English.

There was deviation from normality in Sample 1 as indicated by a few skewness and kurtosis values that were in absolute value larger than 2 and in Samples 2 and 4 as indicated by some kurtosis values that were in absolute value larger than 2 (Bandalos and Finney, 2010). For Sample 3 all skewness and kurtosis values were in absolute value smaller than 2 (Bandalos and Finney, 2010). For all samples, the *p*-values of the test statistics of Mardia's multivariate skewness and kurtosis were small, pointing to deviations from multivariate normality. The descriptive statistics of the individual scale items for all samples are presented in **Supplementary Table 1**, the multivariate skewness and kurtosis values in **Supplementary Table 2**, and the inter-item correlations are presented in **Supplementary Tables 3–6**.

Stage 2: Factorial Validity

The various models tested for the four samples are portrayed in **Figures 1**, **2**: a one-factor CFA (Model 1); and the following three- and four-structure models: CFA (Models 2a and 2b), bifactor CFA (Models 3a and 3b), ESEM (Models 4a and 4b), and bifactor ESEM (Models 5a and 5b). The fit indices are presented in **Table 2**. Model 1 revealed poor fit for all samples. Models 4 showed improved fit indices compared to Models 2, while Models 3 and 5 fitted better than Models 2 and 4. Although the four-factor models yielded slightly improved fit indices compared to the three-factor models, we preferred the three-factor structure for the sake of parsimony. The focus of this section will henceforth be on reporting the detailed results for analyses done using the three-factor structure.

For Samples 1 and 2, Model 5a showed best fit. For Sample 3, Models 4a and 5a, with item 9 ("I can say that I have found my purpose in life") removed, fitted best. Item 9 in the QEWB-Setswana had a negative residual variance[1] which suggested

[1]Before removal of item 9 from Models 4a and 5a, the residual variances were specified to be larger than zero but this resulted in a first-order derivative matrix that was not positive-definite.

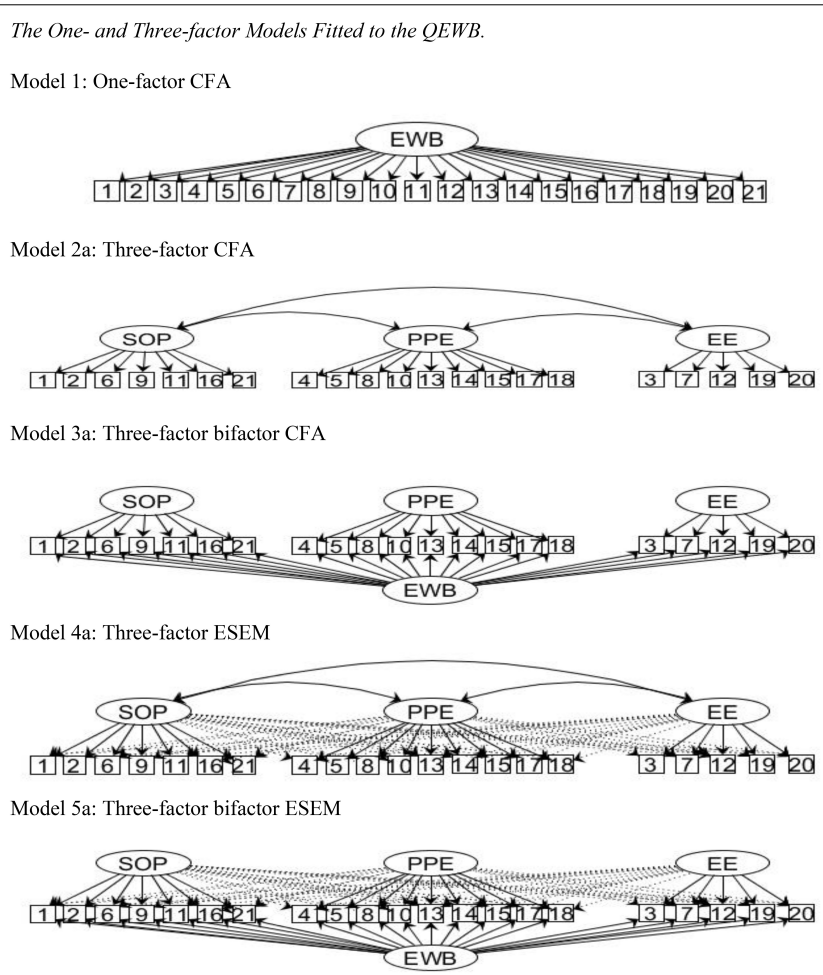

FIGURE 1 | The one- and three-factor models fitted to the QEWB. Model 1: One-factor CFA. Model 2a: Three-factor CFA. Model 3a: Three-factor bifactor CFA. Model 4a: Three-factor ESEM. Model 5a: Three-factor bifactor ESEM. EWB, Eudaimonic Well-being; SOP, Sense of Purpose factor; PPE, Purposeful Personal Expressiveness factor; EE, Effortful Engagement factor.

removal of the item. Item 9 was removed on statistical grounds. Since Model 5a performed well in Samples 1 and 2 and Model 5a, with item 9 removed, performed well across Samples 1, 2, and 3, we selected these models for invariance testing. For Sample 4, all models tested revealed poor fit. Several attempts to find a model with better fit as suggested by high MI's and EPC's, or removing items with negative residual variances, while bearing in mind substantive considerations, did not produce any model with good fit that made substantive sense. We therefore concluded that we could not find support for the validity of the QEWB for Sample 4. The remainder of this section will present further results for Samples 1, 2, and 3.

Next, we examined the factor loadings of the items. The standardised factor loadings for the final preferred models of Samples 1, 2, and 3 (Model 5a for the QEWB-English [Sample 1] and QEWB-Afrikaans [Sample 2], and for the QEWB-Setswana [Sample 3] Model 5a with item 9 removed) are presented in **Table 3**. For Sample 1, all items had statistically significant loadings on the general factor. For the specific factors target factor loadings were generally larger than cross-loadings and were all statistically significant for the SOP and EE factors. For the PPE factor only item 4 had a statistically significant target factor loading, while items 8 and 13 had larger statistically significant cross-loadings on the SOP factor. The SOP and EE factors had target factor loadings that were generally larger than the loadings on the general factor. Target factor loadings on the PPE factor were generally smaller than the loadings on the general factor.

For Sample 2, all items had significant loadings on the general factor. Only item 11 (belonging to the SOP factor) loaded significantly on the SOP factor. Except for items 5 and 8 of the PPE factor, all items loaded significantly on the specific target factor for the PPE and EE factors. All target factor loadings were larger than cross-loadings, except for SOP item 1 that had a larger statistically significant cross-loading on the PPE factor. Although specific factor loadings were mostly larger than 0.3, item loadings on the general factor were mostly larger than specific target factor loadings.

For Sample 3, all items, except items 3, 7, 16, and 20, had statistically significant loadings on the general factor. SOP items

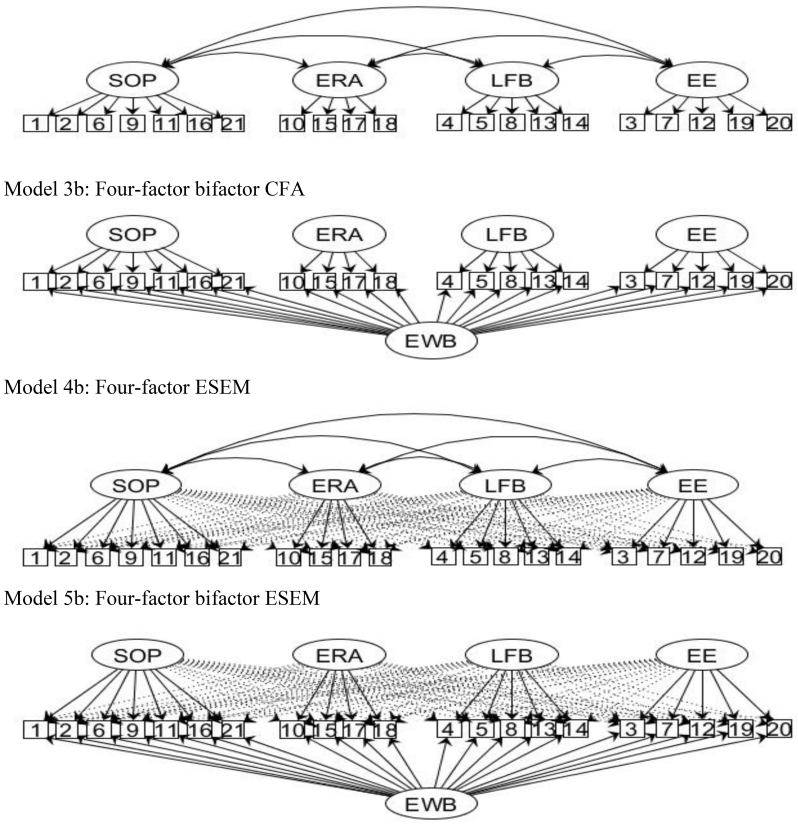

FIGURE 2 | The four-factor models fitted to the QEWB. Model 2b: Four-factor CFA. Model 3b: Four-factor bifactor CFA. Model 4b: Four-factor ESEM. Model 5b: Four-factor bifactor ESEM. EWB, Eudaimonic Well-being; SOP, Sense of Purpose factor; ERA, Engagement in Rewarding Activities; LFB, Living from Beliefs; EE, Effortful Engagement factor.

1 and 16 had statistically significant target factor loadings on the SOP factor, while SOP items 11 and 16 had statistically significant cross-loadings on the EE factor. There were no statistically significant target factor loadings on the PPE factor, but items 4 and 14 had statistically significant cross-loadings on the EE factor. All EE items had statistically significant target factor loadings on the EE factor, with no statistically significant cross-loadings on non-target factors. For the SOP and PPE factors, loadings on the general factor were mostly larger than target factor loadings. For the EE factor target factor loadings were larger than loadings on the general factor. Although only a few target factor loadings, mostly that of the EE subscale, were larger than 0.3, target factor loadings were mostly larger than cross-loadings.

Stage 3: Internal Consistency Reliability of the Final Preferred Models for Samples 1, 2, and 3

Omega coefficients for Samples 1, 2, and 3 are presented in **Table 3**. Support for reliability of scores of the general factor was established for all groups with ω-values higher than 0.50 (Putnick and Bornstein, 2017). Except for the PPE factor of the QEWB-English (Sample 1), the EE factor of the QEWB-Afrikaans (Sample 2), and the SOP and PPE factors of the QEWB-Setswana (Sample 3), support was established for reliability of the specific factor scores for the three student samples.

Stage 4: Measurement Invariance

Model 5a was chosen as the final preferred model for Samples 1, 2, and 3, but item 9 had to be removed for Sample 3 who completed the QEWB-Setswana. To find a baseline model for testing measurement invariance, we first investigated the fit of Model 5a with item 9 removed to data from Samples 1 and 2. Good fit was obtained for Sample 1 (CFI = 0.964; RMSEA = 0.029) and Sample 2 (CFI = 0.969; RMSEA = 0.028). We therefore conducted two sets of measurement invariance tests: First, we tested measurement invariance between Samples 1 and 2 using Model 5a as baseline model. Then we tested measurement invariance between Samples 1, 2, and 3 using Model 5a with item 9 removed as baseline model. The results are presented in **Table 4**.

TABLE 2 | Fit indices for the one-, three-, and four-factor models.

Latent model	χ^2	*df*	*p*	CFI	TLI	RMSEA	90% CI of RMSEA	SRMR	$SRMR_u$	90% CI of $SRMR_u$	$SRMR_u/\overline{R^2}$
Sample 1: QEWB-English (students)											
1-factor											
Model 1	560.472	189	<0.001	0.631	0.590	0.078	(0.070; 0.085)	0.087	0.071	(0.055; 0.087)	0.349
3-factor											
Model 2a	383.762	186	<0.001	0.804	0.778	0.057	(0.049; 0.065)	0.074	0.052	(0.037; 0.067)	0.178
Model 3a[a]	323.438	168	<0.001	0.846	0.807	0.053	(0.045; 0.062)	0.057	0.029	(0.017; 0.042)	0.091
Model 4a	246.033	150	<0.001	0.905	0.866	0.044	(0.034; 0.054)	0.043	N/A	N/A	N/A
Model 5a	180.810	132	0.0031	0.952	0.923	0.034	(0.020; 0.045)	0.036	N/A	N/A	N/A
4-factor											
Model 2b	370.108	183	<0.001	0.814	0.787	0.056	(0.048; 0.064)	0.073	0.049	(0.034; 0.064)	0.156
Model 3b[a]	333.162	168	<0.001	0.836	0.795	0.055	(0.046; 0.064)	0.062	0.026	(0.014; 0.039)	0.078
Model 4b	180.810	132	0.0031	0.952	0.923	0.034	(0.020; 0.045)	0.036	N/A	N/A	N/A
Model 5b	147.231	115	0.0229	0.968	0.942	0.029	(0.012; 0.043)	0.030	N/A	N/A	N/A
Sample 2: QEWB-Afrikaans (students)											
1-factor											
Model 1	599.253	189	<0.001	0.748	0.720	0.067	(0.061; 0.073)	0.068	0.060	(0.05; 0.07)	0.295
3-factor											
Model 2a	392.096	186	<0.001	0.873	0.857	0.048	(0.041; 0.055)	0.059	0.046	(0.037; 0.056)	0.164
Model 3a[a]	274.518	168	<0.001	0.935	0.918	0.036	(0.028; 0.044)	0.043	0.023	(0.016; 0.031)	0.074
Model 4a	219.893	150	<0.001	0.957	0.940	0.031	(0.022; 0.040)	0.031	N/A	N/A	N/A
Model 5a	171.566	132	0.0117	0.976	0.961	0.025	(0.012; 0.035)	0.027	N/A	N/A	N/A
4-factor											
Model 2b[a,b]	378.854	183	<0.001	0.880	0.862	0.047	(0.041; 0.054)	0.057	0.044	(0.035; 0.054)	0.160
Model 3b[a]	No convergence in Mplus								0.024	(0.017; 0.031)	0.079
Model 4b	171.566	132	0.0117	0.976	0.961	0.025	(0.012; 0.035)	0.027	N/A	N/A	N/A
Model 5b	153.891	115	0.0090	0.976	0.956	0.027	(0.014; 0.037)	0.024	N/A	N/A	N/A
Sample 3: QEWB-Setswana (students)											
1-factor											
Model 1	505.915	189	<0.001	0.635	0.594	0.081	(0.072; 0.089)	0.090	0.070	(0.051; 0.089)	0.364
3-factor											
Model 2a	420.470	186	<0.001	0.730	0.695	0.070	(0.061; 0.079)	0.086	0.068	(0.052; 0.084)	0.275
Model 3a[a]	No convergence in Mplus								0.017	(0.002; 0.032)	0.059
Model 4a[c]	146.249	133	0.204	0.983	0.975	0.020	(0.000; 0.037)	0.038	N/A	N/A	N/A
Model 5a[c]	134.157	116	0.119	0.976	0.961	0.025	(0.000; 0.041)	0.031	N/A	N/A	N/A
4-factor											
Model 2b	412.410	183	<0.001	0.736	0.697	0.070	(0.061; 0.079)	0.086	0.063	(0.047; 0.079)	0.242
Model 3b[a]	No convergence in Mplus								0.022	(0.008; 0.036)	0.071
Model 4b[c]	134.157	116	0.119	0.976	0.961	0.025	(0.000; 0.041)	0.031	N/A	N/A	N/A
Model 5b[d]	60.390	73	0.854	1.000	1.043	0.000	(0.000; 0.021)	0.023	N/A	N/A	N/A
Sample 4: QEWB-English (adults)											
1-factor											
Model 1	891.433	189	<0.001	0.400	0.333	0.119	(0.111; 0.127)	0.124	0.100	(0.081; 0.118)	0.510
3-factor											
Model 2a	693.079	186	<0.001	0.567	0.511	0.102	(0.094; 0.110)	0.116	0.094	(0.072; 0.117)	0.327
Model 3a	No convergence in Mplus or lavaan										
Model 4a	324.457	150	<0.001	0.851	0.791	0.067	(0.057; 0.077)	0.050	N/A	N/A	N/A
Model 5a	281.429	132	<0.001	0.872	0.797	0.066	(0.055; 0.076)	0.045	N/A	N/A	N/A
4-factor											
Model 2b	679.200	183	<0.001	0.576	0.513	0.102	(0.094; 0.110)	0.114	0.083	(0.063; 0.103)	0.279
Model 3b	No convergence in Mplus or lavaan										
Model 4b	281.429	132	<0.001	0.872	0.797	0.066	(0.055; 0.076)	0.045	N/A	N/A	N/A

(Continued)

TABLE 2 | (Continued)

Latent model	χ^2	*df*	*p*	CFI	TLI	RMSEA	90% CI of RMSEA	SRMR	$SRMR_u$	90% CI of $SRMR_u$	$SRMR_u/\overline{R^2}$
Model 5b	276.883	115	<0.001	0.862	0.747	0.073	(0.062; 0.084)	0.041	N/A	N/A	N/A

QEWB, Questionnaire for Eudaimonic Well-Being; ***1-factor:*** *Model 1 = confirmatory factor analysis (CFA);* ***3-factor and 4-factor:*** *Model 2 = CFA, Model 3 = bifactor CFA; Model 4 = exploratory structural equation modelling (ESEM); Model 5 = bifactor ESEM; χ^2, Chi-square; df, degrees of freedom; p, probability value; CFI, comparative fit index; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation; 90% CI of RMSEA, 90% confidence interval of the RMSEA; SRMR, standardised root mean square residual; $SRMR_u$, unbiased SRMR calculated using lavaan; 90% CI of $SRMR_u$ = 90% confidence interval of the $SRMR_u$; $SRMR_u/\overline{R^2}$ $SRMR_u$ divided by the average R^2 of the items. N/A indicates that the $SRMR_u$ was not calculated for the ESEM or bifactor ESEM models. All fit statistics were calculated using Mplus Version 8.3 except for the $SRMR_u$ that was calculated using the lavResiduals function of the lavaan package in R4.0.2.*

[a] *Lavaan output in R warns that variance-covariance matrix does not appear to be positive definite.*

[b] *Mplus output warns that Psi matrix is not positive definite.*

[c] *Item 9 removed (negative residual variance).*

[d] *Items 9, 1, 18 removed (negative residual variance).*

TABLE 3 | Standardised factor loadings and omega coefficients for the final preferred 3-factor models of the QEWB (Samples 1, 2, and 3).

	Sample 1: QEWB-English bifactor ESEM				Sample 2: QEWB-Afrikaans bifactor ESEM				Sample 3: QEWB-Setswana bifactor ESEM (item 9 out)			
Item	**G**	**SOP**	**PPE**	**EE**	**G**	**SOP**	**PPE**	**EE**	**G**	**SOP**	**PPE**	**EE**
SOP factor												
1	**0.39***	**0.26***	0.01	0.08	**0.55***	**0.02**	0.14*	−0.08	**0.35***	**0.55***	−0.01	−0.26*
2	**0.40***	**0.59***	0.12	−0.04	**0.66***	**0.20**	−0.08	−0.13*	**0.48***	**0.28**	0.13	−0.12
6	**0.53***	**0.43***	−0.15	0.09	**0.62***	**−0.11**	0.14	0.09	**0.58***	**0.04**	0.06	−0.07
9	**0.40***	**0.64***	0.05	−0.10	**0.68***	**0.38**	−0.03	−0.12	**–**	**–**	–	–
11 (R)	**0.25***	**0.46***	0.09	0.26*	**0.35***	**0.71***	0.05	0.29*	**0.34***	**0.23**	−0.07	0.59*
16 (R)	**0.15***	**0.35***	−0.04	0.34*	**0.43***	**0.20**	−0.19*	0.13*	**0.08**	**0.32***	−0.01	0.41*
21	**0.26***	**0.61***	0.04	−0.15	**0.69***	**0.39**	−0.06	−0.11	**0.58***	**0.06**	−0.02	0.02
PPE factor												
4	**0.60***	0.13	**0.58***	−0.03	**0.41***	0.00	**0.18***	0.08	**0.49***	0.13	**0.06**	−0.16*
5	**0.47***	0.05	**0.16**	0.06	**0.33***	−0.04	**0.20**	0.08	**0.37***	−0.09	**−0.17**	0.08
8	**0.58***	−0.11*	**0.00**	−0.05	**0.36***	−0.10	**0.07**	0.05	**0.51***	−0.16	**−0.19**	0.08
10	**0.19***	−0.12	**−0.18**	−0.18	**0.16***	0.02	**0.25***	−0.12	**0.32***	−0.13	**−0.21**	−0.02
13	**0.63***	−0.19*	**0.15**	−0.09	**0.35***	−0.04	**0.39***	0.01	**0.63***	0.04	**−0.04**	0.06
14	**0.48***	0.08	**0.00**	0.00	**0.45***	−0.01	**0.32***	−0.06	**0.52***	0.21*	**−0.15**	−0.26*
15	**0.60***	0.11	**−0.33**	0.09	**0.37***	−0.11	**0.41***	0.03	**0.57***	−0.10	**0.19**	0.05
17	**0.47***	0.10	**−0.10**	−0.05	**0.45***	0.02	**0.42***	0.10	**0.42***	0.13	**0.13**	0.02
18	**0.61***	−0.20*	**−0.31**	−0.03	**0.26***	0.00	**0.62***	−0.06	**0.66***	−0.08	**0.40**	0.03
EE factor												
3 (R)	**0.10**	−0.02	0.13	**0.37***	**0.18***	0.01	0.02	**0.41***	**−0.23**	0.10	0.23	**0.23***
7 (R)	**0.24***	0.13	−0.11	**0.41***	**0.17***	0.05	0.01	**0.36***	**0.21**	−0.05	0.22*	**0.37***
12 (R)	**0.24***	−0.03	−0.01	**0.34***	**0.31***	0.21*	0.17*	**0.33***	**0.30***	0.09	−0.10	**0.57***
19 (R)	**0.23***	−0.05	−0.03	**0.59***	**0.28***	−0.01	−0.07	**0.41***	**0.33***	0.10	−0.06	**0.64***
20 (R)	**0.28***	0.22*	0.03	**0.50***	**0.39***	0.06	−0.09	**0.36***	**0.11**	−0.06	−0.03	**0.45***
Omega coefficients	0.83	0.73	0.37	0.58	0.84	0.54	0.55	0.48	0.83	0.38	0.29	0.60

QEWB, Questionnaire for Eudaimonic Well-Being; ESEM, exploratory structural equation modelling; G, general factor; SOP, Sense of Purpose factor; PPE, Purposeful Personal Expressiveness factor; EE, Effortful Engagement factor; (R), item is reverse scored. Factor loadings on the general factor and target factor loadings on the intended specific factors are indicated in bold. Scale items are available from Waterman et al. (2010).

$^*p < 0.05$.

For Samples 1 and 2 (using Model 5a as baseline model), the configural invariance model fitted the data well (CFI = 0.966; RMSEA = 0.029). When testing for metric invariance, several factor loadings had to be freely estimated in the two groups to reach a point where support for partial metric invariance was indicated by adequately small ΔCFI and ΔRMSEA values. Selection of parameters to free was based on relatively small MI-values (less than 10) and these changes were not substantively justifiable. The subsequent partial scalar invariance model did not converge. We therefore concluded that only support for configural invariance was established.

For Samples 1, 2, and 3 (using Model 5a with item 9 removed as baseline model), the configural invariance model yielded good fit (CFI = 0.959; RMSEA = 0.032). However, full metric invariance was not supported and when testing for partial metric invariance, the first-order derivative product matrix, as well as the

TABLE 4 | Measurement invariance of the QEWB for Samples 1, 2, and 3.

Model	χ^2	*df*	*p*	CFI	RMSEA	Model comparison	χ^2	*df*	*p*	Δ CFI	Δ RMSEA
Samples 1 and 2 (Model 5a)											
Invariance Model 1	353.179	264	<0.001	0.966	0.029	–	–	–	–	–	–
Invariance Model 2A	485.643	332	<0.001	0.941	0.034	2A vs. 1	131.291	68	<0.001	−0.025	0.005
Invariance Model 2B	470.891	328	<0.001	0.945	0.033	2B vs. 1	114.807	64	0.000	−0.021	0.016
Invariance Model 2C	440.588	322	<0.001	0.954	0.030	2C vs. 1	86.199	58	0.010	−0.012	0.001
Invariance Model 2D	417.300	314	<0.001	0.960	0.029	2D vs. 1	64.553	50	0.081	−0.006	0.000
Invariance Model 3	No convergence										
Samples 1, 2, and 3 (Model 5a, item 9 removed)											
Invariance Model 1	472.511	348	<0.001	0.959	0.032	–	–	–	–	–	–
Invariance Model 2A	685.999	476	<0.001	0.930	0.035	2A vs. 1	206.723	128	<0.001	−0.029	0.003
Invariance Model 2B	The first-order derivative product matrix, as well as the latent variable covariance matrix for Sample 3 was not positive definite										

Samples 1 and 2 (Model 5a): *Invariance Model 1 = configural invariance model; Invariance Model 2A = metric invariance model; Invariance Model 2B = partial metric invariance model with the factor loading of item 11 on the Effortful Engagement factor freely estimated in both groups; Invariance Model 2C = partial metric invariance model with the factor loadings of items 11 and 6 on the Effortful Engagement factor freely estimated in both groups; Invariance Model 2D = partial metric invariance model with the factor loadings of items 11 and 6 on the Effortful Engagement factor and item 8 on the Sense of Purpose factor freely estimated in both groups; Invariance Model 3 = partial scalar invariance model;* ***Samples 1, 2, and 3 (Model 5a, item 9 removed):*** *Invariance Model 1 = configural invariance model; Invariance Model 2A = metric invariance model; Model 2B = partial metric invariance model with the factor loading of item 3 on the Effortful Engagement factor freely estimated in all groups;* χ^2*, Chi square; df, degrees of freedom; p, probability value; CFI, comparative fit index; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation; 90% CI, 90% confidence interval of the RMSEA; SRMR, standardised root mean square residual.*

latent variable covariance matrix for Sample 3 was not positive definite. We therefore concluded that only support for configural invariance was established.

DISCUSSION

This study explored the dimensionality of the QEWB in four culturally diverse South African samples (three student samples, one adult sample) who completed different language versions of the scale, demonstrating the performance of the scale when applying different analytic techniques. Measurement invariance was also examined where sufficient baseline fit was obtained. The bifactor ESEM model best fitted the data for all student samples, although item 9 had to be removed from the QEWB-Setswana. Although scale items should ideally be removed on both statistical and substantive grounds (Byrne, 2012), item 9 was removed on statistical grounds only as there were no clear substantive reasons for its removal. Future research may investigate whether this result replicates in other samples. For the student samples, support was established for the reliability of scores on the general EWB factor and some specific factors. None of the tested models fitted the adult sample. Configural invariance was supported between the student samples, but not metric or scalar invariance. Findings will be discussed in the paragraphs below.

Dimensionality of the Questionnaire for Eudaimonic Well-Being

The results for student Samples 1, 2, and 3 are in line with research that supported the multidimensionality of the QEWB (e.g., Schutte et al., 2013; Fadda et al., 2017, 2020; Klym-Guba and Karaś, 2018; Salavera and Usán, 2019; Ishii et al., 2022). The one-factor CFA model showed poor model fit for all the student samples. Although the four-factor structure yielded models with slightly better fit compared to the three-factor models, we selected the more parsimonious three-factor structure for our final preferred models. The three-factor structure was also supported by Klym-Guba and Karaś (2018) who, with the application of ESEM, obtained similar item-factor fit as Schutte et al. (2013), except for items 1 and 6 that loaded on the PPE factor, item 4 that loaded on the SOP factor, and item 10 that was removed.

Our results point towards the multidimensionality of EWB and support the existence of a general EWB factor that coexists with some specific EWB factors. The results further point towards the limitations inherent in CFA and indicate that multidimensional constructs may be represented better by statistical models that account for sources of multidimensionality. Firstly, model fit improved when cross-loadings were modelled (e.g., ESEM models). The cross-loadings were generally small and can therefore be regarded as the influence of the non-target factor on the construct-relevant part of the item (Morin et al., 2016a). Small cross-loadings compared to loadings on target factors point towards the factorial validity of all the language versions of the QEWB for the student samples. Secondly, the improvement in model fit when a general factor was modelled (e.g., bifactor CFA and bifactor ESEM), indicates that a general EWB factor (that directly influences all items of the QEWB) coexists with the specific EWB factors. Together these results indicate that the inclusion of cross-loadings and/or a general factor resulted in improved model fit, thereby pointing towards the existence of a global eudaimonic well-being factor and the interrelatedness of the specific eudaimonic well-being factors.

Measurement Invariance

We established support for configural invariance between Samples 1 and 2 when the three-factor bifactor ESEM model was

applied; and between Samples 1, 2, and 3 when the three-factor bifactor ESEM model, with item 9 removed, was applied. This implies that the same factor structure of the QEWB held across the relevant samples (cf. Lee, 2018) and that latent theoretical constructs are associated with the same items, connoting that the same items can be used to measure the constructs across the groups (Boer et al., 2018). However, neither full nor partial metric or scalar invariance was established across the samples. The samples can therefore not be compared on factor variances and covariances, nor on factor mean scores. These findings are noteworthy because it means that, although factor loadings and factor mean scores cannot be compared, there are similarities in how eudaimonic well-being is experienced and expressed across the more African and more Western cultural groups.

Measuring Eudaimonic Well-Being Across Age Groups

For Sample 4, the adult sample, none of the models tested displayed good fit. This could not be remedied by correlating residual variances of item pairs suggested by high MI and EPC values, nor by setting residual variances to be larger than zero to avoid negative residual variance values. This finding is in contrast with the good psychometric properties of the QEWB found in previous studies (Waterman et al., 2010; Schutte et al., 2013; Areepattamannil and Hashim, 2017; Fadda et al., 2017, 2020; Klym-Guba and Karaś, 2018), as well as in the student samples of the current study. In another study amongst adults, Ishii et al. (2022), that used Japanese adult samples, tested several different models with different factor structures before they selected the four-factor ESEM model for the 18- to 29-years age group and the three-factor ESEM model for the 30- to 49-year age group and the 50- to 69-year age group (see Ishii et al., 2022 for a description). The fit indices were inconsistent across the age groups, but suggested that three to five factors are most appropriate for the QEWB. The other models, including the bifactor models, did not yield interpretable results. The only other study of the QEWB's psychometric properties amongst adults was by Sotgiu et al. (2019) who used a sample in a similar age range (18–60 years) to our adult sample (18–74 years). They applied Rasch analysis and found that a unidimensional factor structure displayed good model fit for the QEWB. However, with a mean age of 28 years for the sample in the study by Sotgiu et al. (2019) and a mean age of 40 years for our adult sample, as well as the fact that Sotgiu et al. (2019) applied Rasch-analysis, which is different to the statistical techniques applied in this study, the comparison of the results between our study and the study done by Sotgiu et al. (2019) with regard to model fit should be made with caution. Overall, it does seem as if the developmental phase of the participants may play a role in the psychometric performance of the QEWB.

The adult sample in this study consisted mainly of participants in young (18–40 years) and middle (40–65 years) adulthood as discerned by Erikson (1997). According to Erikson (1997) young adulthood is the developmental phase during which individuals become less self-directed as they become more concerned with the developmental task of forming intimate and long-term relationships with others. During middle adulthood the main developmental task is to develop generativity, which involves the concern to contribute to others and society by actions such as parenting, volunteering, mentoring, and engaging in productive and meaningful work (Erikson, 1997). Adults in midlife (40–65 years of age) search for meaning in life and may adapt their sense of identity reflecting on their lives so far (Kuther and Burnell, 2019). To the contrary, the student samples in this study (mean age between 19 and 21 years) are emerging adults (18–25 years, Arnett, 2000). During this phase the development of a sense of self is regarded as the main developmental task (Erikson, 1997). Although this stage was initially associated with adolescence (12–18 years, Erikson, 1997), it was later recognised that this stage may last into emerging adulthood (Arnett, 2000). Emerging adulthood is characterised by, *inter alia*, identity explorations (e.g., developing one's identity through the exploration of various life possibilities), instability (e.g., experiencing life changes), self-focus (being focused on oneself while acquiring skills needed for adulthood), feeling in-between (e.g., subjectively experience that one is in a transitional phase of life), and possibilities/optimism (e.g., believing that the future holds possibilities; Arnett, 2004; Arnett and Mitra, 2018). These features were proposed to be more prominent in, but not exclusive to, emerging adulthood (Arnett, 2004; Arnett and Mitra, 2018), and may differ across cultures (Arnett, 2011).

The main developmental tasks associated with each developmental phase may have influenced how the student and adult samples, respectively, responded to the items. For example, items measuring "self-discovery" may have been more relevant in the student groups, while items measuring "sense of purpose and meaning in life" may have been more relevant in the adult group. In this regard, Sotgiu et al. (2019), who applied Rasch-analysis, indicated that certain items of the Italian version of the QEWB were more typical of some age groups than of other. They found that items 3 ("I think it would be ideal if things came easily to me in my life"), 12 ("I can't understand why some people want to work so hard on the things that they do"), and 19 ("If something is really difficult, it probably isn't worth doing") were more typical of emerging adults (18–25 years), item 11 ("As yet, I've not figured out what to do with my life") was more typical of young adults (26–35 years) and middle-aged adults (36–60 years), and items 2 ("I believe I have discovered who I really am"), 19 ("If something is really difficult, it probably isn't worth doing"), and 21 ("I believe I know what I was meant to do in my life") were more typical of middle-aged adults (36–60 years). They argued that EWB seemed to have been cultivated in different ways across the age groups. Whereas emerging adults and young adults seemed to have emphasised hard work and putting effort into difficult activities, middle-aged adults seemed to have emphasised self-knowledge and setting life goals (Sotgiu et al., 2019).

These findings imply that EWB, as operationalised in the QEWB, may operate differently across age groups, which may influence the psychometric properties of the QEWB across different age groups. The findings suggest that practitioners, such as psychologists and counsellors, must consider the developmental phase of clients when considering and assessing clients' levels of eudaimonic well-being. The QEWB shows

potential for use in practice in student samples, but not the current adult sample, to measure and evaluate levels of EWB. Future research is indicated that explores EWB and the measurement thereof from a developmental perspective and more studies are needed to see if the findings replicate.

Limitations and Recommendations

The study provide preliminary support for applying the bifactor ESEM model to QEWB data from student samples. However, the study has limitations. Firstly, the use of non-probability samples limits the generalisation of the results to other student and adult groups. Secondly, the unbiased SRMR fit statistic was calculated for the CFA and bifactor CFA models only, and not the ESEM and bifactor ESEM models since fitting these models with the lavaan package is still in its infancy. The unbiased SRMR has shown superiority to other fit indices (Ximénez et al., 2022) and future research should explore the performance of this fit statistic when ESEM and bifactor ESEM models are applied. Thirdly, the different models were tested in the same samples, and item 9 was removed from the three-factor bifactor-ESEM model for Sample 3, a result that may not hold across samples or populations as model modifications followed a data driven approach and the results may be partly or entirely influenced by idiosyncratic sample characteristics (MacCallum et al., 1992). In this sense, the current study should be conceived as a substantive illustration to explore the performance of the different analytic procedures, rather than a validation study. Future research should test the different analytical models in representative independent samples to determine its validity across samples and populations. Fourthly, while we consider the results for the adult sample to be noteworthy, especially since this study was one of very few to evaluate the performance of the QEWB among adults, the sample size was small and multicultural, and findings may not replicate in other adult samples. Future research should investigate the performance of the QEWB in other larger adult groups to not only determine to what extent the scale is usable in adult samples, but also to better understand how the underlying theoretical construct manifest and operate across age groups. Such research may be done from a developmental perspective where item functioning in various developmental phases are investigated, while cultural/contextual variables are also considered.

CONCLUSION

The dimensionality of the QEWB and its underlying theoretical construct has been a contentious issue in the literature. The current study supports previous findings that EWB is multidimensional, but at the same time represents an overarching higher order construct, and suggests that analytic models that allow for the articulation of this structure are preferred when modelling the QEWB. The study further found support for configural invariance of the scale across three language versions of the scale completed by university students, with the samples representing more African and more Western cultural groups. However, metric and scalar invariance were not achieved. Although factor variances, covariances, and mean scores cannot be compared, the findings imply that there are similarities in how EWB manifests and is expressed across cultural groups. For the adult sample, use of the QEWB in the current sample is not recommended. The QEWB seems to show differential psychometric properties for different developmental phases which points towards the need to validate, and establish the equivalence of, the QEWB in age groups other than emerging adults. This also suggests the broader need for investigation of the manifestation of EWB across different age groups, and suggest that practitioners should take cognisance of possible varying manifestations of EWB in different developmental phases.

DATA AVAILABILITY STATEMENT

The data analysed in this study is subject to the following licenses/restrictions: The datasets generated during and/or analysed during the current study are available from the author MW on reasonable request, subject to ethics approval. Requests to access these datasets should be directed to MW, Marie.Wissing@nwu.ac.za.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Health Research Ethics Committee of the North-West University, South Africa, ethics approval number: NWU 00002-07-A2. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AC, LS, and MW contributed to the design and planning of the study. LS and MW were responsible for the gathering and capturing of the data. AC, LS, and WS attended to the statistical analyses and the interpretation of the results. AC drafted the manuscript, incorporated the suggestions from the co-authors, and prepared the final manuscript for submission. WS drafted selected parts of the Data Analysis and Results sections. LS and MW provided continuous and critical feedback regarding the intellectual content of the document. The final manuscript was read and approved by all authors.

FUNDING

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant Numbers: 106050, 91557, 123266, 120060, and 121948). The Grantholders acknowledge that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF-supported research are that of the authors, and that the NRF accepts no liability whatsoever in this regard. Funding was also provided to the first author in the form of a Doctoral degree scholarship and a bursary from the Faculty of Health Sciences of the North-West University, South Africa.

ACKNOWLEDGMENTS

The authors wish to thank I. P. Khumalo and L. Temane for their assistance with the gathering of the data for Samples 1, 2, and 3 and for their assistance with the translation of the QEWB from English into Setswana. The authors also wish to thank S. Mbowa for her assistance with the interpretation of the scale items of the Setswana version of the QEWB and for inputs from a Batswana cultural perspective. This work is based on an earlier draft in the Ph.D. of the author AC.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.795770/full#supplementary-material

REFERENCES

Areepattamannil, S., and Hashim, J. (2017). The questionnaire for eudaimonic well-being (QEWB): psychometric properties in a non-western adolescent sample. *Personali. Indiv. Diff.* 117, 236–241. doi: 10.1016/j.paid.2017.06.018

Arnett, J. J. (2000). Emerging adulthood: a theory of development through the late teens through the twenties. *Am. Psychol.* 55, 469–480. doi: 10.1037//0003-066X.55.5.469

Arnett, J. J. (2004). *Emerging Adulthood: The Winding Road From The Late Teens Through The Twenties*. Oxford: Oxford University Press.

Arnett, J. J. (2011). "Emerging adulthood(s): The cultural psychology of a new life stage," in *Bridging cultural and developmental psychology: New syntheses in theory, research, and policy*, ed. L. A. Jensen (Oxford University Press), 255–275. doi: 10.1007/s11524-013-9827-6

Arnett, J. J., and Mitra, D. (2018). Are the features of emerging adulthood developmentally distinctive? A comparison of ages 18-60 in the United States. *Emerging Adulthood* 2018, 1–8. doi: 10.1177/2167696818810073

Asparouhov, T., and Muthén, B. O. (2009). Exploratory structural equation modeling. *Structural Equation Modeling* 16, 397–438. doi: 10.1080/10705510903008204

Bandalos, D. L., and Finney, S. J. (2010). "Factor analysis: Exploratory and confirmatory," in *The reviewer's guide to quantitative methods in the social sciences*, eds G. R. Hancock and R. O. Mueller (Routledge), 93–114. doi: 10.1037/10694-008

Boer, D., Hanke, K., and He, J. (2018). On detecting systematic measurement error in cross-cultural research: A review and critical reflection on equivalence and invariance tests. *Journal of Cross-Cultural Psychology* 49, 713–734. doi: 10.1177/0022022117749042

Bollen, K. A. (1989). *Structural equations with latent variables*. Hoboken, NJ: Wiley.

Brislin, R. W. (1970). Back translation for cross-cultural research. *Journal of Cross-Cultural Psychology* 1, 185–216. doi: 10.1177/135910457000100301

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Milton Park: Routledge.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal* 14, 464–504. doi: 10.1080/10705510701301834

Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equat. Mod. Multidis. J.* 9, 233–255. doi: 10.1207/S15328007SEM09025

Clarke, P. J., Marshall, V. W., Ryff, C. D., and Rosenthal, C. J. (2000). Well-being in canadian seniors: findings from the canadian study of health and aging. *Can. J. Aging* 19, 139–159. doi: 10.1017/s0714980800013982

Core Team, R. (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Diener, E. (1984). Subjective well-being. *Psychol. Bull.* 95, 542–575. doi: 10.1037/0033-2909.95.3.542

Diener, E., Heintzelman, S. J., Kushlev, K., Tay, L., Wirtz, D., Lutes, L. D., et al. (2017). Findings all psychologists should know from the new science on subjective well-being. *Can. Psychol.* 58, 87–104. doi: 10.1037/cap0000063

Erikson, E. H. (1997). "The life cycle completed," in *Extended Version with New Chapters on the Ninth Stage of Development*, ed. J. M. Erikson (New York, NY: W. W. Norton & Co).

Fadda, D., Quevedo-Aguado, M. P., Benavente Cuesta, M. H., and Scalas, L. F. (2020). The multidimensional and hierarchical nature of the questionnaire for eudaimonic wellbeing: a bifactor-esem representation in a spanish sample. *Front. Psychol.* 11:422. doi: 10.3389/fpsyg.2020.00422

Fadda, D., Scalas, L. F., Meleddu, M., and Morin, A. J. S. (2017). A bifactor ESEM representation of the questionnaire for eudaimonic well-being. *Personali. Indiv. Diff.* 116, 216–222. doi: 10.1016/j.paid.2017.04.062

Howard, J. L., Gagné, M., Morin, A. J. S., and Forest, J. (2018). Using bifactor exploratory structural equation modeling to test for a continuum structure of motivation. *J. Manage.* 44, 2638–2664. doi: 10.1177/0149206316645653

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

Huta, V., and Waterman, A. S. (2014). Eudaimonia and its distinction from hedonia: developing a classification and terminology for understanding conceptual and operational definitions. *J. Happ. Stud.* 15, 1425–1456. doi: 10.1007/s10902-013-9485-0

Ishii, Y., Sakakibara, R., Kubota, A. K., and Yamaguchi, K. (2022). Reconsidering the structure of the questionnaire for eudaimonic well-being using wide age-range japanese adult sample: an exploratory analysis. *BMC Psychol.* 10:1–10. doi: 10.1186/s40359-021-00707-2

Jennrich, R. I., and Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrica* 76, 537–549. doi: 10.1007/S11336-011-9218-4

Karaś, D., and Chieciuch, J. (2018). The relationship between identity processes and well-being in various life domains. *Personali. Indiv. Diff.* 121, 111–119. doi: 10.1016/j.paid.2017.09.027

Keyes, C. L. M. (1998). Social well-being. *Soc. Psychol. Quart.* 61, 121–140.

Keyes, C. L. M. (2002). The mental health continuum: from languishing to flourishing in life. *J. Health Soc. Behav.* 43, 207–222. doi: 10.2307/3090197

Kimiecik, J., Ward, R. M., and Sohns, E. (2019). Back to basics: tracing health and risk behaviors back to well-being. *Am. J. Health Stud.* 34, 11–22.

Klym-Guba, M., and Karaś, D. (2018). Polish version of the questionnaire for eudaimonic well-being – three factors rather than one. *Health Psychol. Rep.* 6, 273–283. doi: 10.5114/hpr.2018.75684

Kuther, T. L., and Burnell, K. (2019). A life span development perspective on psychosocial development in midlife. *Adult. J.* 18, 27–39. doi: 10.1002/adsp.12067

Lee, S. T. H. (2018). Testing for measurement invariance: does your measure mean the same thing for different participants? *Assoc. Psychol. Sci. Observer* 31, 32–33.

MacCallum, R. C., Roznowski, M., and Necowitz, L. B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychol. Bull.* 111, 490–504. doi: 10.1037/0033-2909.111.3.490

Martela, F., and Sheldon, K. M. (2019). Clarifying the concept of well-being: psychological need satisfaction as the common core connecting eudaimonic and subjective well-being. *Rev. General Psychol.* 23, 458–474. doi: 10.1177/1089268019880886

Matsumoto, D., and Yoo, S. H. (2006). Toward a new generation of cross-cultural research. *Perspect. Psychol. Sci.* 1, 234–250. doi: 10.1111/j.1745-6916.2006.00014.x

Maydeu-Olivares, A. (2017). Assessing the size of model misfit in structural equation models. *Psychometrika* 82, 533–558. doi: 10.1007/s11336-016-9552-7

McDonald, R. P. (1970). Theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *Br. J. Math. Statist. Psychol.* 23, 1–21. doi: 10.1111/j.20448317.1970.tb00432.x

Morin, A. J. S., Arens, A. K., and Marsh, H. W. (2016a). A bifactor exploratory structural equation modeling framework for the identification of distinct

sources of construct-relevant psychometric multidimensionality. *Struct. Equat. Mod. Multidis. J.* 23, 116–139. doi: 10.1080/1070551.2014.961800

Morin, A. J. S., Arens, K., Tran, A., and Caci, H. (2016b). Exploring sources of construct-relevant multidimensionality in psychiatric measurement: a tutorial and illustration using the composite scale of morningness. *Int. J. Meth. Psychiatric Res.* 25, 277–288. doi: 10.1002/mpr.1485

Muthén, L. K., and Muthén, B. O. (1998–2019). *Mplus (Version 8.3)*.

Perreira, T. A., Morin, A. J. S., Hebert, M., Gillet, N., Houle, S. A., and Berta, W. (2018). The short form of the workplace affective commitment multidimensional questionnaire (WACMQ-S): a bifactor ESEM approach among healthcare professionals. *J. Vocat. Behav.* 106, 62–83. doi: 10.1016/j.jvb.2017.12.004

Putnick, D. L., and Bornstein, M. H. (2017). Measurement invariance conventions and reporting: the state of the art and future direction for psychological research. *Dev. Rev.* 41, 71–90. doi: 10.1016/j.dr.2016.06.004

Revelle, W. (2021). *psych: Procedures for Personality and Psychological Research*. Evanston, Illinois: Northwestern University.

Rosseel, R. (2012). lavaan: an R package for structural equation modeling. *J. Statist. Software* 48, 1–36. doi: 10.18637/jss.v048.i02

Ryff, C. D. (1989). Happiness in everything, or is it? Explorations on the meaning of psychological well-being. *J. Personali. Soc. Psychol.* 57, 1069–1081. doi: 10.1037/0022-3514.57.6.1069

Ryff, C. D., and Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *J. Personali. Soc. Psychol.* 69, 719–727. doi: 10.1037//0022-3514.69.4.719

Salavera, C., and Usán, P. (2019). Psychometric properties of the QEWB questionnaire on eudaimonic well-being in teenagers. *Educ. Psychol.* 25, 139–146. doi: 10.5093/psed2019a3

Sánchez-Oliva, D., Morin, A. J. S., Teixeira, P. J., Carracxa, E. V., Palmeira, A. L., and Silva, M. N. (2017). A bifactor exploratory structural equation modeling representation of the structure of the basic psychological needs at work scale. *J. Vocat. Behav.* 98, 173–187. doi: 10.1016/j.jvb.2016.12.001

Schutte, L., Wissing, M. P., and Khumalo, I. P. (2013). Further validation of the questionnaire for eudaimonic well-being (QEWB). *Psychol. Well Being Theory Res. Pract.* 3, 1–22.

Shi, D., Maydeu-Olivares, A., and DiStefano, C. (2018). The relationship between the standardized root mean square residual and model misspecification in factor analysis models. *Multiv. Behav. Res.* 53, 676–694. doi: 10.1080/00273171.2018.1476221

Sotgiu, I. (2019). Gender differences and similarities in autobiographical memory for eudaimonic happy events. *J. Happ. Stud.* 20, 1457–1479. doi: 10.1007/s10902-018-0006-z

Sotgiu, I., Anselmi, P., and Meneghini, A. M. (2019). Investigating the psychometric properties of the questionnaire for eudaimonic well-being: a rasch analysis. *Testing Psychometrics Methodol. Appl. Psychol.* 26, 237–247. doi: 10.4473/TPM26.2.5

Van de Vijver, F. J. R., and Leung, K. (1997). *Methods and Data Analysis For Cross-Cultural Research*. Thousand Oaks, CA: SAGE.

Van de Vijver, F., and Humbleton, R. J. (1996). Translating tests: some practical guidelines. *Eur. Psychol.* 1, 89–99.

Waterman, A. S. (1993). Two conceptions of happiness: contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment. *J. Personali. Soc. Psychol.* 64, 678–691. doi: 10.1037/0022-3514.64.4.678

Waterman, A. S., Schwartz, S. J., Zamboanga, B. L., Ravert, R. D., Williams, M. K., Agocha, V. B., et al. (2010). The questionnaire for eudaimonic well-being: psychometric properties, demographic comparisons, and evidence of validity. *J. Posit. Psychol.* 5, 41–61. doi: 10.1080/17439760903435208

Whittaker, T. A. (2012). Using the modification index and the standardized expected parameter change for model modification. *J. Exp. Educ.* 80, 26–44. doi: 10.1080/00220973.2010.531299

Wissing, M. (2008/2012). *The Prevalence of Levels of Psychosocial Health: Dynamics and Relationships with Biomarkers of (ill) Health in South African Social Contexts (FORT3) [Unpublished Research Protocol]*. Potchefstroom: North-West University.

Ximénez, C., Maydeu-Olivares, A., Shi, D., and Revuelta, J. (2022). Assessing cutoff values of SEM fit indices: advantages of the unbiased SRMR index and its cutoff criterion based on communality. *Struct. Equat. Mod.* doi: 10.1080/10705511.2021.1992596

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

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

Copyright © 2022 Cromhout, Schutte, Wissing and Schutte. 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.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 02 September 2022
DOI 10.3389/fpsyg.2022.775622

Check for updates

OPEN ACCESS

EDITED BY
Rebecca Shankland,
Lumière University Lyon 2, France

REVIEWED BY
Deepa K. Damodaran,
Jain University, India
Katarzyna Danuta Durniat,
University of Wrocław, Poland
Chris C. Martin,
Oglethorpe University, United States

*CORRESPONDENCE
Virág Zábó
zabo.virag@ppk.elte.hu

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

RECEIVED 14 September 2021
ACCEPTED 03 August 2022
PUBLISHED 02 September 2022

CITATION
Zábó V, Oláh A and Vargha A (2022) A new complex mental health test in a positive psychological framework. *Front. Psychol.* 13:775622. doi: 10.3389/fpsyg.2022.775622

COPYRIGHT
© 2022 Zábó, Oláh and Vargha. 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.

A new complex mental health test in a positive psychological framework

Virág Zábó[1,2,3]*, Attila Oláh[1] and András Vargha[1,4]

[1]Positive Psychology Research Group, Institute of Psychology, Faculty of Education and Psychology, Eötvös Loránd University, Budapest, Hungary, [2]Doctoral School of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary, [3]Institute of Research on Adult Education and Knowledge Management, ELTE Eötvös Loránd University, Budapest, Hungary, [4]Person- and Family-oriented Health Science Research Group, Institute of Psychology, Faculty of Humanities and Social Sciences, Károli Gáspár University of the Reformed Church in Hungary, Budapest, Hungary

According to the Maintainable Positive Mental Health Theory (MPMHT), the main pillars of positive mental health are global well-being, efficient coping that enables an individual to maintain positive conditions and functioning, savoring capacity, resilience, and dynamic self-regulation. This study presents the validation of a new five-scale mental health test (MHT), the MHT that operationalizes MPMHT. The methodology comprised two online cross-sectional studies using self-report questionnaires. Participants in Study I (n = 1,736; 448 males, 1,288 females; mean age 51.3 years; SD = 11.6 years) filled in the MHT, the Flow, the Positive emotions, Engagement, Positive Relationship, Meaning, Accomplishment Questionnaire (PERMA-Profiler), and the Flourishing Scale. Participants in Study II (n = 1,083; 233 males, 847 females; mean age 33.9 years; SD = 12.2 years) filled in the MHT, the Shortened Aspiration Index, the short form of the Beck Depression Inventory, the WHO Well-Being Index, the Satisfaction with Life Scale, the Purpose in Life Test, and the Schema Questionnaire–Short Form. Exploratory factor analysis (EFA) identified a five-factor structure with 17 items in Study I that was confirmed with excellent fit measures in confirmatory factor analysis in Study II. Both studies indicated a high level of internal consistency (above 0.70). In each subscale, a minimum part of 44% did not overlap with the set of the other subscales. The content validity of the subscales was confirmed by 10 tests of mental health. We found a positive correlation of the self-regulation and resilience subscales with age, while women showed a higher level of savoring than men at all age levels. When Study I was replicated after 2 weeks and again after 11 months, excellent internal consistency and good test–retest correlation values of the MHT scales were found. The MHT can thus be considered a reliable and valid measurement tool for mental health.

KEYWORDS

happiness, subjective well-being, mental health, mental health test, positive psychology (PP1.0 and PP2.0), positive psychological assessments, maintainable positive mental health theory

10.3389/fpsyg.2022.775622

Introduction

The present study proposes a new theory of mental health, the Maintainable Positive Mental Health Theory (MPMHT). First, we present the theoretical framework of our new positive mental health concept. In the next step, we provide an overview of the models of well-being and mental health to date, and we highlight the gaps between these and our proposed concept. By presenting the five pillars of MPMHT, we aim to shed light on how our new model integrates and complement the accumulated knowledge of the science of mental health measures to date.

The conceptualization of maintainable positive mental health theory

A central topic in positive psychology is the identification of indicators (symptoms) of mental health and the elaboration of models to serve as a theoretical framework when developing a diagnostic system in the realm of positive mental health. Such a diagnostic system could provide a positive alternative to the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), which has long been used in clinical practice. Given that, in identifying mental health, the absence of mental disease is not an adequate definition, it is necessary first to develop a positive concept of mental health.

We argue that the clarification of the concept of well-being, the theoretical and empirical analysis of the relationship between well-being and mental health, and the rehabilitation of classical interpretations of mental health can lead to an integrative concept of the positive realm of mental health. According to our MPMHT, the level of well-being will depend both on the presence or absence of the capacities and psychological resources needed to ensure positive mental health, and on the ability to use these capacities. Such a concept would treat all theoretically and empirically identified components of well-being as the set of symptoms of mental health that reflect the presence and proper functioning of the psychological capacities needed to ensure and maintain positive mental health.

This new approach is in line with the mental health definition of the WHO (Galderisi et al., 2015). The World Health Organization (WHO) defines mental health as "a dynamic state of internal equilibrium which enables individuals to use their abilities in harmony with universal values of society. Basic cognitive and social skills; ability to recognize, express and modulate one's own emotions, as well as empathize with others; flexibility and ability to cope with adverse life events and function in social roles; and harmonious relationship between body and mind represent important components of mental health which contribute, to varying degrees, to the state of internal equilibrium" (Galderisi et al., 2015). In other words, the components of well-being are not the only agents that contribute to positive mental health; resilience, accommodation to changes, and the development of efficient coping capacities (savoring and the ability to establish positive states and handle negative states) are also major contributors.

Maintainable positive mental health theory endorses the view of what Keyes (2007; Keyes et al., 2020) emphasized, which is that mental disease and mental health are two separate dimensions that must be treated as independent continuums. At one pole of mental disease are intensive and frequent occurrences of mental disorders, while at the opposite pole are symptoms of mental disorders that are rarely present or that take insignificant forms. In the case of mental health, one pole represents the rare occurrence and weak appearance of mental health indicators, while the opposite pole represents the high frequency of positive mental health indicators.

An overview of the concepts and the dimensions of previous models of well-being and mental health

Just as mental disorders comprise a theoretically based and empirically demonstrable set of symptoms, positive mental health can also be characterized by a clearly identifiable set of symptoms that can be framed within an appropriate theory and demonstrated and studied empirically.

Mental health as multidimensional well-being

According to one approach in positive psychology, mental health can essentially be defined according to the pillars of multidimensional well-being. Ryff's (1995) multifactor psychological well-being scale integrates the six components of overall well-functioning but does not cover hedonic well-being which is also an important component of mental health. Keyes (2007) 13-dimensions mental health questionnaire does not comprehensively include all the important capabilities (e.g., savoring, self-regulation, resilience) and components (e.g., spiritual well-being) which are essential in achieving a high degree of positive mental health. Diener's and his colleagues' flourishing scale (Diener et al., 2009) only measures psychological well-being, SPANE (Diener et al., 2009) only assess emotions, and Positive emotions, Engagement, positive Relationship, Meaning, Accomplishment (PERMA) (Seligman, 2018) describes the factors of well-being, but none of them cover the competencies and abilities that play an irreplaceable role in mental health.

Mental health as mirror opposite to the symptoms of mental disorders

An alternative approach is to argue that mental health is characterized by symptoms that are the mirror opposite of mental disorders (Huppert and So, 2013; Caprara et al., 2019; Oláh, 2019). For example, the Positivity Scale

(Caprara et al., 2019; Oláh, 2019) measures a combination of high self-esteem, life satisfaction, and optimism as suggested sources of a syndrome of optimal functioning. These models do not include any ability or capacity to guarantee mental health, only the resilience appears in Huppert and So's (2013) flourishing questionnaire.

Mental health as flourishing

Although these approaches differ, 70% of their components, from which the multidimensional domain of mental health is constructed, are more or less identical. They share the common feature of integrating the hedonic and eudemonic approaches in the metaphoric concept of flourishing, which can be characterized by the simultaneous presence of positive feelings and good psychological functioning. Flourishing, as an umbrella concept for the components of well-being, is the positive pole or uppermost zone on the mental health continuum.

Mental health as "hedo-eudemonic" well-being

"Flourishing" is also used in the names of measurement tools that operationalize "hedo-eudemonic" models of well-being, for example Global Well-being Scale (Oláh, 2019; Oláh et al., 2020), Positive Mental Health Scale (Lukat et al., 2016) and Positive Functioning Inventory (Joseph and Maltby, 2014) capture only the hedonic and eudaimonic aspects of mental health. Measurement tools (or diagnostic scales) like Diener's and his colleagues' flourishing scale (Diener et al., 2009), SPANE (Diener et al., 2009) and a version of PERMA extended with overall well-being, negative emotions, loneliness, and physical health components (PERMA Profiler, Butler and Kern, 2016) encompass the major components of well-being and the symptoms of positive mental health in terms of their targets and measured content. Nevertheless, these hedo-eudemonic models do not cover all aspects of well-being (e.g., self-acceptance, spiritual well-being). In a recent review paper that outlined empirical studies covering 99 well-being measurement tools, 196 different components of well-being were identified (Linton et al., 2016).

Classical models of mental health

In the so-called classical models of mental health (Jahoda, 1958; Vaillant, 2003; Vaillant G. E., 2012) the focus is on personality traits that guarantee efficient self-regulation and flexible accommodation, and on psychological resources that foster fulfillment. For instance, the use of the character strengths that leads to human flourishing (Mayerson, 2020) are only predictors of the optimal mental health. The presence of personality traits resources is also regarded as a primary factor in positive mental health in salutogenic theory (Antonovsky, 1991), according to which mental health is maintained by efficient stress management, resilient self-regulation, and flexible accommodation to the continuously changing world (Block and Kremen, 1996; Block and Block, 2014). In theories of this kind, well-being is the consequence, or the result, of mental health and is due to resilient accommodation, capacities that establish a physical, psychological, and environmental balance, and well-functioning health maintenance skills. The PISI (Oláh, 2005) includes skills for self-efficacy and self-regulation, but does not cover global well-being, savoring, and resilience.

Balanced models of mental health

According to a recent model, the life balance and harmony model, the key to well-being is to maintain balance and harmony in all areas of human functioning (Lomas, 2021). In the equilibrium theory, a key factor in well-being is the maintenance of a relative balance between challenges and available resources, with an emphasis on the capacities needed to handle situations that are out of balance (Dodge et al., 2012). Apart from omitting some important elements of well-being, these models do not imply the abilities and capacities. Balance, harmony, and equilibrium could be a principle in the dynamic interaction of the pillars of mental health. Also, according to the capability theory of Amartya Sen (see, e.g., Walker and Unterhalter, 2007), capacities can explain the establishment and maintenance of all possible aspects of well-being. The focus of the model deviates from the essence of positive mental health and this approach can be a potential predictor of positive mental health. Sustainable Mental Health Model (Bohlmeijer and Westerhof, 2021) integrates dysfunctional and functional perspectives of mental health.

Mental health as the sum of the components of well-being

In earlier tools for the measurement of the realm of positive mental health (Bech et al., 1996; Diener et al., 2009; Lamers et al., 2011; Huppert and So, 2013; Butler and Kern, 2016; Lukat et al., 2016; Oláh et al., 2020), the focus has been primarily on the components of well-being, without taking into account all the aspects of mental health referred to in the WHO definition and in classical theories of mental health. Another feature of these tools is that they cannot be considered as the operationalization of a comprehensive mental health model. It is also important to note that the number of published measures of well-being and mental health has decreased significantly since 2010 (Lomas, 2021).

Extracting the pillars of Maintainable Positive Mental Health Theory

The holistic concept of positive mental health must integrate the realm of well-being models, the psychological resources and capacities needed to ensure and maintain positive mental health, and the skills that guarantee their efficient functioning.

Our main goal is to develop a construct of mental health that differs both theoretically and empirically from

other theoretically defined constructs of multidimensional well-being (flourishing, PERMA, etc.) called often as an equivalent or the top-zone of mental health. We assert that mental health is a broader concept than well-being, justified by research based on the classical theories of mental health (Antonovsky, 1991; Vaillant G., 2012) and on our own studies (Oláh, 2021; Nagy et al., 2021), which is also consistent with the mental health definition of the WHO (Galderisi et al., 2015). However, we do not encounter in the literature a measurement tool of mental health that could be considered an operationalizing construct of the WHO description of mental health, and we do not know also a mental health model that would integrate suggestions of scientific approaches for the pillars of positive mental health from classical models up to most recent theories of positive psychology. Keyes' concept of mental health, the Flourishing construct of Diener and Huppert, and the PERMA model of Seligman all construe positive mental health based on well-being components. Building on classical and recent ideas of positive mental health, in the concept of Maintainable Positive Mental Health we would like to clarify that mental health is a function of individual capacities (resilience, creative and executive competencies) by means of which the individual can work up an equilibrium with the outside world, promoting his/her development, creating a steady state for within-person functioning (self-regulation), and an equilibrium of positive and negative emotions (coping, savoring). The existence and efficient functioning of these elements may lead to the global well-being, a multifaceted component of positive mental health. Summarizing, in our suggested definition mental health is a high level of global well-being which goes-together with psychological, social, and spiritual well-functioning, resilience, efficient creative and executive functioning, coping and savoring capacities, all pillars insuring the maintainability of mental health.

The first pillar in our new model is Global Well-being. MPMHT integrates existing well-being theories and the identified dimensions of well-being in the global well-being pillar. Global Well-being means multi-component subjective well-being encompassing emotional states and psychological functioning in the emotional, psychological, social, and spiritual areas in life (Oláh and Kapitány-Fövény, 2012; Oláh, 2016; Oláh et al., 2020). **Table 1** indicates the pillars of Global-Well-being which goes hand in hand with savoring capacity, creative and executive efficiency, self-regulation, and resilience.

The second pillar is Savoring which refers to the ability and capacity which allows individuals to mentally mobilize their joyful memories and experiences to generate mental well-being, reliving them in the present and, furthermore, extending them to future events (Bryant and Veroff, 2007). Savoring is also a prerequisite for MPMHT because it is an ability that can guarantee the achievement and maintenance of positive mental health (Bryan et al., 2022).

TABLE 1 Pillars of global well-being according to maintainable positive mental health theory.

Global well-being

Emotional well-being	Positive functioning		Spiritual well-being
	Psychological well-being	Social well-being	
Positive affect Happiness Life Satisfaction	Self-acceptance Personal growth Environmental mastery Autonomy Positive relations with others	Social acceptance Social actualization Social contribution Social coherence Social integration	Joy of transcendence experience Joy of universality experience Vertical and horizontal responsibility

Thirdly, utilizing Creative and Executive Efficiency competence, an individual becomes able to cope with the difficulties they encounter by mobilizing their various competencies in the difficult, stressful, and challenging situations of life. Furthermore, it indicates how individuals are able to provide successful individual and social problem-solving behavior (Oláh, 2005; Oláh et al., 2020).

Fourthly, the ability to regulate and control emotions, temperament, and negative states and to persist in achieving a given goal plays an important role in mental health (Oláh, 2005; Elliot et al., 2011; Singh and Sharma, 2018). Self-regulation is the capacity of the individual to disregard prominent responses and to regulate affects, cognitions, and behaviors. It is the ability to alter thoughts, feelings, desires, and actions in the perspective of such higher goals and would represent one of the most adaptive variables of human behavior (Vohs and Baumeister, 2004).

The fifth pillar is Resilience. With the ability of resilience, an individual is able to mobilize their mental capacities and resources to maintain positive mental health when they face unexpected, stressful, and difficult situations. The higher the level of resilience, the more quickly the individual is able to recover from a sudden, unexpected stressful situation (Connor and Davidson, 2003; Southwick and Charney, 2018; Verdolini et al., 2021). According to MPMHT, these independent components are together responsible for the mental health of the individual.

Although various psychological constructs exist, the mental health test (MHT) could be the first test to have a five-dimensional complex structure (Global Well-being, Savoring, Creative and Executive Efficiency, Self-regulation, and Resilience), with the aim of covering the wide spectrum of mental health. For these reasons,

the present study aimed to operationalize the MPMHT by preparing the validation of the five-scale MHT on a Hungarian population.

Overview of the present studies

Two studies were carried out to prepare and validate the final version of the MHT. The aim of Study I was to finalize with EFA a set of items for which the five-dimensional statistical model of the MHT yields appropriate fit indices. The main aim of Study II was to confirm the model on a new, independent sample. After this verification of structural validity, both studies were used to check the substantive validity of the MHT.

Study I exploratory factor analysis

Method

Participants and procedure

Participants completed a 64-item online questionnaire posted from mid-January 2020 for 2 months in Facebook groups that are frequently visited by adults of different ages[1] and with different occupations and interests. Ethical approval for the study was granted by the Research Ethics Committee of the local university (permission number: 2019/61). Participation was voluntary and anonymous. Informed consent was obtained but no compensation was given. The valid study sample consisted of 1,736 adult persons (448 males and 1,288 females). Among them 1,540 persons were residents of Hungary, and 196 persons were residents of other countries who filled in the Hungarian-language online questionnaire[2]. The sociodemographic characteristics of the sample are summarized in **Table 2**.

Based on the data in **Table 2**, we can conclude that the Study I sample was sufficiently heterogeneous for us to draw valid conclusions about the MHT from the sample. Although the majority of participants in the sample were women (74.2%), the number of men (448) was also sufficiently large to ensure representative results.

In terms of age, the majority of the participants (38.9%) were middle-aged (36–50 years), although the number of people over 65 (198) was also substantial. There was a strikingly low proportion of young people aged between 18 and 25 (1.9%, 33 people). The sample was also balanced according to the type of city where participants lived, since the number of respondents in all categories was over 300. More than 98% of the participants had a high-school certificate. Regarding marital status, 50.1% of the participants were married, while a small proportion (5.0%) were widowed. The majority of the participants (57.8%) were employed, although there was also a significant proportion of retired participants (20.7%) and entrepreneurs (14.5%). The vast majority of participants in the sample (74.6%) considered their financial situation to be average, although there was also a non-negligible proportion of wealthy people (17.5%). Only a small proportion of respondents declared themselves to be poor (2.0%) or rich (0.6%).

Measures

Nine of the questions in the questionnaire referred to sociodemographic data (gender, age, place of residence, etc., see **Table 2**). One question (Positive experience%) assessed the proportion of the respondent's recent positive experiences (1 = *10% positive experiences and 90% negative experiences . . .* 9 = *90% positive experiences and 10% negative experiences*). Four special questions, scored using a six-point Likert scale, assessed the physical and mental condition of the respondent: (1) Physical condition (My physical state is: 1 = *very bad*, 2 = *bad*, 3 = *acceptable*, 4 = *good*, 5 = *very good*, 6 = *excellent*); (2) General mental state (My general mental state is: 1 = *very bad*, 2 = *bad*, 3 = *acceptable*, 4 = *good*, 5 = *very good*, 6 = *excellent*); (3) General health condition (I am satisfied with my general health: 1 = *strongly disagree*, 2 = *moderately disagree*, 3 = *slightly disagree*, 4 = *slightly agree*, 5 = *moderately agree*, 6 = *strongly agree*); and (4) Physical strength (I feel strong and physically robust: 1 = *strongly disagree*, 2 = *moderately disagree*, 3 = *slightly disagree*, 4 = *slightly agree*, 5 = *moderately agree*, 6 = *strongly agree*).

Mental health test

The basic concept was to develop a mental health test (MHT) in the form of a short questionnaire comprising no more than 20 items to obtain a comprehensive picture of mental health according to MPMHT. The items on the five scales were selected based on the following arguments, considering that the item that most strongly represents the given measuring instrument should be selected (Oláh et al., 2018):

(1) Well-being. This scale is based on the following three areas of subjective well-being: (i) the ratio of positive and negative experiences (Fredrickson, 2009; **Appendix**, item 1); (ii) the subjective quality of the state of mental health and well-being (Lyubomirsky, 2010; **Appendix**, item 14); and (iii) the global level of happiness (Huppert and So, 2013; **Appendix**, item 18).

(2) Savoring. For this scale, three items were selected from the short Hungarian version of the Savoring Beliefs Inventory (Nagy et al., 2021) based on the criteria that they be

1 Seniors Group; Through Engineers' Eyes; Professionals Group; Danube Anglers; Youth of Budapest; Soccer fans; We enjoy cooking; Gardeners; Classical music lovers; Viticulturists and Winemakers Club.

2 Their place of residence could be identified based on the name of the settlement provided by them.

comprehensive, simple, and highly representative of the total score of the questionnaire. The three items selected for the MHT according to these principles (**Appendix**, items 3, 10, and 12) explained the total score of the 10-item short Hungarian form of the Savoring Beliefs Inventory with a variance ratio of $R^2 = 0.81$ based on annual data from the Happiness Map of Hungary between 2015 and 2019, with a sample size of 8,035[3].

(3) Creative and Executive Efficiency. This scale consists of four items (**Appendix**, items 5, 7, 15, and 17) taken from the 16-item short version of the Psychological Immune Competence Inventory (PICI; see Oláh, 2005), using the same names. A fifth item, which was also considered an important component of creative and executive efficiency, was added from the 80-item version of the PICI (Oláh, 2005) (**Appendix**, item 9).

(4) Self-regulation. For this scale, three items were selected from the self-regulation subscale of the 16-item short version of the PICI (Oláh, 2005).

(5) Resilience. Items for this scale in the first version of the MHT (Oláh et al., 2018) were selected from the Resilience subscale of the 16-item short version of the PICI (Oláh, 2005). Taking into account the many different scales of resilience (Block and Kremen, 1996; Connor and Davidson, 2003; Smith et al., 2008; Windle et al., 2011), we were looking for one that is internationally recognized and that has already been translated into several languages. Applying linear regression analysis, four items (**Appendix**, items 4, 6, 11, and 13) of the six-item Brief Resilience Scale (Smith et al., 2008) explained the total score of the six-item test with a variance ratio of $R^2 = 0.95$ on a sub-sample ($n = 8,035$) of the above-mentioned Happiness Map of Hungary.

Based on the above considerations, the MHT questionnaire comprises 18 items (see **Appendix**) divided into five scales of three, three, five, three, and four items respectively. The items are scored according to six-point Likert scales (1 = *does not agree at all*, 6 = *agrees completely*). The Well-being, Savoring, and Creative and Executive Efficiency subscales include positive items only. The Self-regulation subscale contains negative items only. The Resilience subscale consists of two positive and two negative items. The final scores are obtained by averaging the scores obtained for the items in each of the five scales.

Flow

The key element of Csíkszentmihályi's flow construct (immersion in activity, constant interest) was examined based on Magyaródi et al. (2013, 2014), using the following item: "If something really interests me, I am able to do it with pleasure and in depth, even in difficult situations."

Positive emotions, engagement, positive relationship, meaning, accomplishment questionnaire (PERMA-profiler)

The PERMA model was developed by Seligman (2018), building on his further developed earlier concept of authentic happiness. The components of the five-pillar model reinforce one another in creating and maintaining a state of well-being. The 23-item PERMA-Profiler (Butler and Kern, 2016) measures Seligman's model using a 10-point Likert scale (0 = *never/not at all/terrible*; 10 = *always/completely/excellent*). In the case of 15 items, the five basic pillars are measured using three questions each; eight items, comprising three questions each, are used to assess negative emotions (e.g., "How often do you feel anxious?") and health (e.g., "How would you rate your health?"). Among the 23 items one item is used to measure happiness ("All in all, how happy would you say you are?") and one item is used to measure loneliness ("How lonely do you feel in your daily life?"). In terms of the scales that represent the five basic pillars, the Positive Emotion scale focuses on the frequency of experiencing positive emotions (e.g., "How often do you feel joyful?"). The Engagement scale refers to the frequency of experiencing flow and the absorption of cognitive and emotional resources (e.g., "How often do you become absorbed in what you are doing?"). The Meaning scale examines the tendency to

3 https://dailynewshungary.com/happiness-map-of-hungary/

TABLE 2 Sociodemographic characteristics of the participants of Study I ($n = 1,736$, 448 – 25.8% males, 1,288 – 74.2% females).

Age	**18–25 years old: 1.9%**	**26–35 years old: 6.5%**	**36–50 years old: 38.9%**	**51–65 years old: 41.3%**	**66–90 years old: 11.4%**
Number of children	0: 18.5%	1: 23.2%	2: 41.7%	3: 13.0%	3 + : 3.6%
Type of city	Village: 22.4%	Small town: 33.2%	Large town: 25.8%	Capital: 18.6%	
Educational level	Primary: 1.7%	Secondary: 38.2%	College: 35.2%	University: 24.9%	
Marital status	Lives alone: 28.9%	Civil partnership: 16.0%	Married: 50.1%	Widow: 5.0%	
Profession	Employee: 57.8%	Retired person: 20.7%	Entrepreneur: 14.5%	Unemployed: 3.9%	Other: 3.1%
Financial status	Poor: 2.0%	Below average: 5.3%	Average: 74.6%	Wealthy: 17.5%	Rich: 0.6%

TABLE 3 Cronbach's alpha and McDonald's omega values measuring the internal consistency of the MHT scales in Study I.

Scale	Number of items	Study I (n = 1,736)	
		α	ω
Well-being	3	0.84	0.85
Savoring	3	0.85	0.85
Creative and Executive Efficiency	5	0.85	0.85
Self-regulation	3	0.85	0.85
Resilience	4	0.75	0.77
Resilience (without item 6)	3	0.74	0.78

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

set meaningful goals, seek meaning, and serve meaningful goals that transcend the self (e.g., "In general, to what extent do you lead a purposeful and meaningful life?"). The Accomplishment scale provides information about the extent to which one's own successful performance and the joy of experiencing competence contribute to increasing well-being in an individual's life (e.g., "How often do you
achieve the important goals you have set for yourself?"). The Relationships scale measures the extent to which it is true for an individual's life that happiness, love, and good relationships with others are the *sine qua non* of well-being (e.g., "To what extent do you feel loved?"). All the scales in the Hungarian version of the PERMA-Profiler (Oláh, 2016) showed reliable Cronbach's alpha values: Positive Emotions: 0.88; Engagement: 0.57; Relationships: 0.79; Meaning: 0.76; Accomplishment: 0.74; Health: 0.88; and Negative Emotions: 0.77.

Flourishing scale

This eight-item scale (Diener et al., 2009) operationalizes an improved version of Diener's concept of subjective well-being, in which, in addition to life satisfaction and the dominance of positive emotions, the necessity of competence, optimism, contributing to the well-being of others, life purpose, self-esteem, and positive relationships are highlighted. Cronbach's alpha: 0.94.

To summarize, with one exception (the Engagement subscale of the PERMA-Profiler), the scales in the above tests all had excellent reliability, with Cronbach's alpha values of above 0.74.

Results

The internal consistency of all the scales in the MHT was computed with ROPstat (Vargha et al., 2015). **Table 3** shows that the Cronbach's alpha and McDonald's omega values mostly being above 0.80 (DeVellis, 2016; Barbera et al., 2020) were adequate for all scales. One possible way to improve the internal consistency would be to drop one item from the Resilience subscale (item 6: "Stressful events are difficult to bear"). Calculating pairwise correlations of the scales, **Table 4** shows that most of the scales have a moderate positive relationship with Pearson's r of between 0.35 and 0.62 with two exceptions. This is consistent with the theory that different components of mental health are positively related to one another (Ryff and Marshall, 1999).

In the next step we performed a five-factor EFA using Mplus (Muthén and Muthén, 1998–2011) to verify that the 18 items of the MHT form five factors according to the constructed subscales. Due to the strongly non-normal distribution of the scales (see the last two columns of **Table 4**), a robust maximum likelihood estimator (MLR) (Maydeu-Olivares, 2017) was applied with Geomin oblique rotation, allowing correlating factors (Hattori et al., 2017). The rotated factor loading matrix is shown in **Table 5**. In this table, the item indices (i1, i2, etc.) correspond to the item indices in the **Appendix**, supplemented by the abbreviation of their scale (i1W, i2SR, etc.). Based on **Table 5**, we can conclude that each item in the Well-being, Savoring, Creative and Executive Efficiency and Self-regulation scales forms a separate factor with loadings above 0.60, with two exceptions (i12S: 0.40; i9C: 0.41). In the case of the Resilience scale, one of the four items loaded to Self-regulation scale (i6R: 0.50) and another item (i13R) loaded to the Self-regulation scale as well with a 0.38 loading (see **Table 3**).

Study II confirmatory factor analysis

Method

Participants and procedure

Participants completed a 179-item online questionnaire. The questionnaire was published in various groups on Facebook, as the most commonly used social platform. The responses were obtained from the widest possible variations of the groups according to place of residence (e.g., people living in the 2nd district of Budapest, "What I heard in Debrecen," "What I saw in Sopron"); topic (e.g., nature conservationists, vegetarians); sports (e.g., kayak-canoeists, training plans and experiences, novice runners); profession (e.g., job seekers, social workers); spirituality (e.g., atheist–Christian discussion group, daily spiritual quotes, Christian youth); higher education (e.g., law students, teachers); and others (e.g., mathematics for everyone, motivation for everyday life) in order to reach participants from as many social strata as possible.

Ethical approval for the study was granted by the Research Ethics Committee of the local university (permission number: 70/2019/P/ET/2). Participation was voluntary and anonymous. Informed consent was obtained but no compensation was given. The sample, selected by random sampling, consisted of 1,083 individuals (233 males, 847 females, and three individuals

TABLE 4 Intercorrelations and coefficients of the skewness and kurtosis of the MHT scales in Study I (n = 1,736).

Scale	Savoring	CEE	Self-regulation	Resilience	Skewness	Kurtosis
Well-being	0.62***	0.62***	0.35***	0.60***	−0.85***	0.61***
Savoring	1	0.61***	0.19***	0.46***	−0.79***	0.48***
Creative and Executive Efficiency		1	0.24***	0.49***	−0.78***	0.65***
Self-regulation			1	0.45***	−0.50***	−0.45***
Resilience				1	−0.34***	−0.01

***$p < 0.001$.

TABLE 5 Five-factor exploratory factor analysis of the 18 items of the MHT using MLR method and Geomin oblique rotation on the data from Study I (n = 1,736): Five-factor factor weight matrix.

Item	Factor1	Factor2	Factor3	Factor4	Factor5
i1W	**0.67**	0.06	− 0.01	0.03	− 0.01
i14W	**0.85**	− 0.01	0.03	0.01	0.04
i18W	**0.79**	0.00	− 0.01	0.02	0.03
i3S	0.04	**0.72**	0.04	0.01	0.08
i10S	− 0.02	**0.99**	− 0.03	0.00	− 0.04
i12S	0.32	**0.40**	0.09	0.01	0.14
i5C	− 0.03	0.01	**0.78**	0.01	0.08
i7C	− 0.03	0.00	**0.78**	0.04	0.04
i9C	0.30	0.14	**0.41**	− 0.02	0.04
i15C	0.23	− 0.06	**0.65**	0.03	− 0.03
i17C	0.12	0.08	**0.51**	− 0.06	− 0.08
i2SR	0.02	0.07	0.01	**0.84**	− 0.02
i8SR	0.08	0.03	0.00	**0.78**	− 0.08
i16SR	− 0.07	0.00	0.044	**0.75**	0.21
i4R	0.06	0.08	0.19	0.00	0.67
i6R	0.08	− 0.07	− 0.01	0.50	0.25
i11R	0.15	0.065	0.03	− 0.01	**0.69**
i13R	− 0.02	− 0.046	− 0.14	0.38	**0.50**

W, Well-being; S, Savoring; C, Creative and Executive Efficiency; SR, Self-regulation; R, Resilience. Cells with factor loadings greater than 0.35 are highlighted with a bold.

who did not disclose their gender) who completed the 179-item questionnaire online between December 2019 and March 2020. Their sociodemographic characteristics are summarized in Table 6.

Based on the data presented in Table 6, it can be concluded that the Study II sample was slightly less heterogeneous and of slightly different composition than the Study I sample. The majority of the participants in the sample were female (78.2%). On the other hand, there were significantly more young people below the age of 25 (33.1% vs. 2%), while the proportion of those over the age of 50 (1.2% vs. 52.7%) was substantially lower. There were more university students (22.2%) and fewer retired participants (1.4%, as part of the "other" category). As in Study I, the vast majority of participants in Study II (67.1%) also considered their financial status to be average. In terms of religiosity, 68.9% declared themselves to be religious in their own way and 31.1% declared themselves to be atheists, while 3.4% said that religiosity was important (or very important) to them.

Measures

Eight of the questions from among all the items were related to sociodemographic data (gender, age, etc., see Table 6). Three questions were specifically related to the subject's physical and mental well-being: (1) Subjective health status ("Overall, how would you rate your health status?" with response options: 1 = *very bad*, 2 = *bad*, 3 = *average*, 4 = *good*, 5 = *excellent*); (2) Subjective satisfaction with life ("Overall, how satisfied are you with your daily life?" with endpoints for response options: 1 = *completely dissatisfied* and 10 = *completely satisfied*); and (3) Subjective happiness ("All in all, how happy would you say you are?" with endpoints for response options: 1 = *completely unhappy* and 10 = *completely happy*).

Mental health test

A detailed description of the measure can be found above and in the Appendix.

Shortened aspiration index

The original questionnaire (Kasser and Ryan, 1996) was designed to measure attitudes toward general life goals representing intrinsic (growth, affiliation, and community contribution), extrinsic (wealth, fame, and physical appearance), and health-related motivations. The shortened, 14-item Hungarian version, which contains no reversed items, measures life goals on a seven-point Likert scale (Martos et al., 2006). The Cronbach's alpha values obtained were 0.80 (Extrinsic Aspiration scale); 0.74 (Intrinsic Aspiration scale); and 0.55 (Health Aspiration scale).

Short form of the Beck Depression Inventory

This inventory (Beck et al., 1997; Rózsa et al., 2001) is used to measure the inverse symptoms of mental health. The nine items measure the following symptoms of depression using a four-point Likert scale: social withdrawal, indecisiveness, sleep disorder (parasomnia), fatigue, excessive anxiety about physical symptoms, incapacity, pessimism, lack of satisfaction and joy, and self-blame. For example, the

TABLE 6 Sociodemographic characteristics of the participants of Study II (n = 1,083).

Gender	Male: 21.5%	Female: 78.2%	No data: 0.3%		
Age	18–25 years old: 33.2%	26–35 years old: 26.2%	36–50 years old: 3.4%	51–65 years old: 8.9%	66–90 years old: 1.3%
Educational level	Primary: 2.5%	Secondary: 21.2%	Ongoing higher education: 24.3%	University: 52.0%	
Profession	Employee: 66.2%	Student: 22.2%	Entrepreneur: 3.3%	Unemployed: 3.0%	Other: 5.2%
Monthly income	Less than 411 EUR: 33.8%	411–822 EUR: 41.7%	More than 822 EUR: 24.5%		
Subjective financial status	Below average: 15.4%	Average: 67.1%	Above average: 17.5%		
Religiosity	Atheist: 31.1%	Religious in their own way: 47.4%	Religious: 21.5%		
Importance of religiosity	Not important at all: 37.0%	Quite important: 32.6%	Very important: 2.1%	It affects all my actions: 1.3%	

item "I am very worried about physical problems and it's hard to think of much else" assesses excessive anxiety about physical symptoms. Higher scores obtained by averaging the scales indicate more depressive symptoms. Cronbach's alpha: 0.86.

World Health Organization well-being index

This five-item scale (Bech et al., 1996; Susánszky et al., 2006) provides information about the respondent's general well-being based on the previous 2-week period using a four-point Likert scale. Higher scores indicate more positive well-being. Cronbach's alpha: 0.87.

Satisfaction with life scale

This scale (Diener et al., 1985; Martos et al., 2014) measures global satisfaction with life using a seven-point Likert scale for five items such as: "If I could live my life over, I would change almost nothing." A higher score on the scale indicates a higher degree of life satisfaction. Cronbach's alpha: 0.87.

Purpose in life test

Even though the name of the questionnaire (Crumbaugh and Maholick, 1964; Konkolÿ Thege and Martos, 2006) emphasizes life goals only, this 20-item measure assesses life meaning according to Viktor Frankl's concept, using a seven-point Likert scale. The higher the score, the more the respondent experiences their life as meaningful. Cronbach's alpha: 0.87.

Schema questionnaire–short form

Early maladaptive schemas are thought to be important in connection with mental health as they indicate factors such as Failure to Achieve ("Most people are more talented than I am") and Emotional Inhibition ("I don't want people to know about my emotional inhibition personal flaws"). The 95-item questionnaire (Welburn et al., 2002; Unoka et al., 2004) assesses Young's schemas on six-point Likert scales. The 19 subscales, comprising five items each, are organized into five schema domains: (1) Disconnection/Rejection; (2) Impaired Autonomy/Performance; (3) Other-directedness; (4) Impaired Limits; and (5) Over-vigilance/Inhibition. Subscale scores are obtained by averaging the scale scores. The Cronbach's alpha values for the subscales were between 0.76 and 0.94.

To summarize, with one exception (the two-item Health subscale of the Shortened Aspiration Index), the scales used in the above tests all had excellent reliability, with Cronbach's alpha values of above 0.74.

Results

Confirmatory factor analysis (CFA) was performed with several settings with Mplus (Muthén and Muthén, 1998–2011). The most important results are summarized in **Table 7**. Row 1 of **Table 7** shows the adequacy measures of the EFA model of the MHT. The root mean square error of approximation (RMSEA) (Steiger and Lind, 1980) and standardized root mean square residual (SRMR) (Hu and Bentler, 1999) are absolute fit indices which indicate acceptable fit if their values are less than 0.06 and 0.05 respectively (Maydeu-Olivares et al., 2019). The model has an acceptable fit since the values of the two indicators do not exceed 0.06 (except in one case (RMSEA of Model 1: 0.06), and this is true for all models in **Table 7**. In the case of RMSEA, the theoretical value should optimally not exceed 0.05 (Steiger and Lind, 1980). A 90% interval estimate is included in the CI_{90}(RMSEA) column. This is good if the value 0.05 is included in it (Browne and Cudeck, 1993). The p-value of the test that the theoretical value is not greater than 0.05 can be found in the pClose column. This is good if it is greater than 0.05, or, in other words, if it is not significant. Adequacy indicators also include two relative fit indicators, the comparative fit index (CFI) and the Tucker-Lewis Index (TLI). A good fit is indicated when the

TABLE 7 The main model fit indices in exploratory factor analysis and confirmatory factor analysis of the five-factor model of the MHT.

Chi-square	AIC BIC	RMSEA	CI_{90} (RMSEA)	pClose	CFI TLI	SRMR
Model 1 Exploratory factor analysis on Study I sample (n = 1,736)						
460.963*** (df = 61)	80591.0 81186.1	0.061	[0.056; 0.067]	<0.001	0.966 0.924	0.017
Model 2 Confirmatory factor analysis on Study II sample (n = 1, 083)						
497.180*** (df = 109)	56140.6 56444.9	0.057	[0.052; 0.063]	0. 009	0.934 0.918	0.057
Model 3 Confirmatory factor analysis on Study II sample (n = 1, 083).						
323.529*** (df = 107)	55890.6 56204.8	0.043	[0.038; 0.049]	0. 980	0.963 0.953	0.043
Model 4 Second order confirmatory factor analysis on Study II sample (n = 1, 083).						
351.044*** (df = 112)	55917.9 56207.2	0.044	(0.039; 0.050)	0.960	0.960 0.951	0.048

***$p < 0.001$. All analyses refer to tests done without item i6R. In case of Model 3 and Model 4, we allow the residual terms of items i3S and i10S to correlate within the Savoring scale, and we allow the residual term of item i9C to correlate with latent factor 1 (Well-being).

CFI value reaches 0.95 and the TLI value is not much lower, but definitely higher than 0.90 (Hu and Bentler, 1999). The AIC and BIC are measures of comparative fit and their lower values are favorable (Kenny, 2015).

In CFA we chose a robust method for model fitting (maximum likelihood mean variance, MLMV), which, in the case of CFA, provides a good alternative to the traditional ML method requiring multidimensional normality (Gao et al., 2020). The CFA results testing Model 2 indicated a high modification index (79) for the residual correlation between i3S and i10S, and also a high modification index (102) for the correlation between i9C and latent factor 1 (Well-being). The latter is in accordance with the corresponding factor loading exceeding the 0.40 level in EFA (see **Table 5**). In order to improve model fit, we built these correlations in the subsequent CFA models. The improvement is justified by the more favorable fit values in the Model 3 row (RMSEA, SRMR < 0.05; pClose, CFI, TLI > 0.95), having also decreasing AIC and BIC values.

A very important result is that testing our five-factor model on an independent Study II sample, the values obtained indicated good fit for all indicators. Furthermore, when we built in the model a second order factor of the five scales (see row 4), we still had good fit values. To summarize, the five-factor model of the 17-item MHT with the original five scales was confirmed by good fit indices. This result was achieved by allowing one within factor correlation of the residuals (between i3S and i10S), and one cross-loading (between i9C and the latent factor of Well-being) in the final model. In **Table 7** all chi-square tests reject the null hypothesis of exact fit, in large samples, like in ours, this often occurs even when the postulated model is only trivially false (Shi et al., 2018).

The internal consistency of all the scales in the MHT was computed with ROPstat (Vargha et al., 2015). **Table 8** indicates that the Cronbach's alpha and McDonald's omega values mostly being above 0.71 (DeVellis, 2016; Barbera et al., 2020) were adequate for all scales.

Substantive validity

Discriminant validity

In order to check the discriminant validity of the MHT scales, we computed for each scale the proportion of variance explained by the other four scales using multivariate linear regression (Lehmann, 1988) on the pooled sample of the two studies (n = 2,819). The explained proportions of variance were measured by the appropriate R^2 values. The obtained R^2 values for the Well-being, Savoring, Creative and Executive Efficiency, Self-regulation, and Resilience scales were 0.55, 0.44, 0.41, 0.17, and 0.44 respectively. The unexplained proportions of variance of the five scales were therefore 44.6, 56.4, 58.8, 82.7, and 55.6%, the smallest (44.6%) belonging to Well-being and the largest (82.7%) belonging to Self-regulation. This result, which confirms the discriminant validity of all five scales, shows that each scale has a unique part of at least 44% that is not covered by the other four. Well-being is the only scale where the unexplained proportion of variance is less than 50%, indicating that this scale plays a central role within the five-pillar construct of mental health.

TABLE 8 Cronbach's alpha and McDonald's omega values measuring the internal consistency of the MHT scales in Study II.

Scale	Number of items	Study II (n = 1,083)	
		α	ω
Well-being	3	0.899	0.879
Savoring	3	0.718	0.896
Creative and Executive Efficiency	5	0.768	0.894
Self-regulation	3	0.709	0.914
Resilience (without item 6)	3	0.861	0.873

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

External and content validity of the mental health test

The content validity of the MHT and the characteristics of the five scales were analyzed with Pearson's correlations, using different mental health measurement variables from the two studies. Tables 9, 10 indicate these results. Regarding the classification of the strength of the Pearson's correlations, we followed Cohen's convention (Cohen, 1988, pp. 79–80). According to this, a correlation is said to be weak if the absolute value of r does not reach 0.1, and to be medium, strong, or very strong if the absolute value of r reaches 0.3, 0.5, or 0.7 respectively.

Based on correlations exceeding 0.70 in Tables 9, 10, we can conclude that the three-item Well-being scale measures something very similar as other, traditional tests of well-being such as Diener's Flourishing Scale, the PERMA-Profiler, the WHO Well-Being Index, and the Satisfaction with Life Scale. The Positive Emotions subscale of the PERMA-Profiler explains the biggest proportion of variance among the subscales. In accordance with the construct content of the Well-being scale, strong correlations were found with the meaning in life variable (Purpose in Life Test) and the Meaning subscale of the PERMA-Profiler, which confirms the significant role of experiencing the meaning of life in terms of well-being. All these results provide ample evidence of the criterion

TABLE 9 Correlations of the MHT subscales with the validating mental health variables of Study I (n = 1,736).

Variable	Well-being	Savoring	Creative and Executive Efficiency	Self-regulation	Resilience
Diener	0.728**	0.580**	0.671**	0.344**	0.506**
P-Positive Emotions	0.774**	0.503**	0.520**	0.432**	0.567**
P-Engagement	0.507**	0.386**	0.443**	0.266**	0.372**
P-Positive Relationships	0.585**	0.408**	0.389**	0.258**	0.367**
P-Meaning	0.665**	0.472**	0.548**	0.299**	0.458**
P-Accomplishment	0.580**	0.414**	0.551**	0.279**	0.434**
P-Happiness	0.752**	0.484**	0.460**	0.349**	0.511**
P-Health	0.482**	0.308**	0.338**	0.271**	0.383**
P-Negative Emotions	− 0.616**	− 0.376**	− 0.396**	− 0.540**	− 0.543**
P-Loneliness	− 0.459**	− 0.265**	− 0.265**	− 0.284**	− 0.342**
PERMA	0.743**	0.520**	0.579**	0.367**	0.523**
Positive Experience%	0.484**	0.252**	0.262**	0.296**	0.329**
Physical Condition	0.435**	0.265**	0.319**	0.264**	0.340**
General Mental State	0.720**	0.450**	0.469**	0.431**	0.557**
General Health Condition	0.465**	0.292**	0.320**	0.284**	0.372**
Physical Strength	0.448**	0.296**	0.353**	0.240**	0.365**
Worry	− 0.345**	− 0.148**	− 0.154**	− 0.304**	− 0.357**
Nervous	− 0.407**	− 0.195**	− 0.215**	− 0.394**	− 0.381**
Tense	− 0.422**	− 0.203**	− 0.215**	− 0.384**	− 0.390**
Restless	− 0.347**	− 0.165**	− 0.190**	− 0.247**	− 0.302**
Flow	0.541**	0.521**	0.646**	0.275**	0.429**

df = 1,538; *$p < 0.05$. **$p < 0.01$. The letter P at the beginning of the variables indicates that it is a subscale of the PERMA-Profiler.

TABLE 10 Correlations of the MHT subscales with the validating mental health variables of Study II.

Variable	Well-being	Savoring	Creative and Executive Efficiency	Self-regulation	Resilience
Extrinsic Aspiration	− 0.042	0.077*	0.092**	− 0.208**	− 0.085**
Intrinsic Aspiration	0.168**	0.210**	0.241**	0.040	0.044
Health Aspiration	0.157**	0.218**	0.120**	− 0.078*	0.088**
Beck Depression	− 0.720**	− 0.495**	− 0.482**	− 0.239**	− 0.528**
WHO	0.700**	0.476**	0.439**	0.215**	0.449**
SWLS	0.735**	0.503**	0.449**	0.184**	0.440**
Purpose in Life	0.767**	0.569**	0.607**	0.294**	0.563**
Young Schemas	− 0.618**	− 0.452**	− 0.431**	− 0.345**	− 0.532**

df = 1,083; *$p < 0.05$. **$p < 0.01$.

validity of the MHT's Well-being scale. Convergent validity is indicated by negative emotional states (the Negative Emotions subscale of the PERMA-Profiler, and the "nervous" and "tense" questionnaire items), symptoms of depression (the short form of the Beck Depression Inventory), and maladaptive schemas (Young's Schema Questionnaire). The significant explanatory power of flow and physical well-being, as well as general health, can also be interpreted as strengthening convergent validity (Carlson and Herdman, 2012).

Regarding the Savoring scale, convergent validity is evidenced by correlations above 0.50 with variables from other MHTs (Diener's Flourishing Scale, the PERMA-Profiler, Flow, the Satisfaction with Life Scale, the Purpose in Life Test). Convergent validity is further enhanced by moderate, negative correlations with negative emotional states (Beck Depression Inventory and the Negative Emotions subscale of the PERMA-Profiler).

The strong correlation results for the Creative and Executive Efficiency scale with other tests and special questionnaire items of mental health (Diener's Flourishing Scale, the PERMA-Profiler, the flow items, and the Purpose in Life Test) indicate the fulfillment of convergent validity.

In the case of the convergent validity of the Self-regulation scale, we refer to the mostly medium-level significant correlations with the subscales of the PERMA-Profiler and with general mental state. In line with the meaning of the Self-regulation scale in terms of emotionality and mood control, we can interpret the medium to strong significant correlations with the Negative Emotions scale of the PERMA-Profiler, Young's Schema Questionnaire, and the "nervous" and the "tense" questionnaire items as contributing to the criterion validity.

As fulfillment of the convergent validity of the Resilience scale, we refer here to its strong correlations with the Diener's Flourishing Scale, the general mental state questionnaire item, the Purpose in Life Test, and the scales of the PERMA-Profiler, especially with the Positive Emotions and Happiness subscales. Convergent validity is indicated by strong and medium-level negative correlations with negative emotional states (the Negative Emotions subscale of the PERMA-Profiler, and the "nervous" and "tense" questionnaire items), symptoms of depression (Beck Depression Inventory), and maladaptive schemas (Young's Schema Questionnaire), indicating that a high level of resilience provides some kind of protection against negative emotions and attitudes. This is also confirmed by the medium to strong correlations with scales indicating mental well-being (the Meaning and Accomplishment scales of the PERMA-Profiler, the WHO Well-Being Questionnaire, the flow items, and the Satisfaction with Life Scale). Further evidence is provided by the medium to strong correlation with flow, since certain behavioral competencies are essential for flow, which can be considered as an investment in the individual's coping system.

Results with sociodemographic indicators

An examination of the relationship between the five scales of the MHT discussed in our study and the sociodemographic indicators yields many significant results, although these relationships are typically rather weak. For example, for 92% of the correlations examined, Spearman's r is $< |0.20|$. A detailed description of these results is beyond the scope of the present paper; thus, we refer to the most striking results only.

The effect of gender and age was analyzed in the pooled sample of the two studies for individuals with valid gender values (n = 2,817). Gender means differed significantly in the case of the Savoring scale (males: M = 4.33; females: M = 4.66), with standard deviations around 1. The gender difference (females giving higher values) was significant in the two-sample t-test [$t(2814) = 6.90$; $p < 0.001$], in the Mann–Whitney U test ($Z = 6.50$; $p < 0.001$), and in the robust Brunner-Munzel rank test [BM(1128) = 6.58; $p < 0.001$]. Although the Cohen's d (d = 0.30, 95% CI [0.216, 0.391]) and eta-squared ($\eta^2 = 0.017$) effect size measures were rather weak, they were already at an interpretable level. **Figure 1** shows that the dominance of females was manifested in all age categories.

In the same sample, a more significant effect of age was observed for the Self-regulation (see **Figure 1**) and Resilience. Using a two-way analysis of variance (ANOVA), the main effect of age was significant for both scales: Self-regulation: $F(4; 2806) = 53.24$; $p < 0.001$; Resilience: $F(4; 2806) = 32.91$; $p < 0.001$, with partial eta-squared effect size values of 0.07 for the former and 0.05 for the latter, apparently stronger than the effect of gender. In the case of self-regulation, the gender × age interaction effect was also significant [$F(4; 2806) = 4.38$; $p = 0.002$]. **Figure 2** indicates that it was due to the fact that the level of self-regulation increased linearly with women's age, while in the case of men it showed an increasing trend only in the 26–50 age range. This was also confirmed by the fact that Spearman's r between self-regulation and age was significantly lower in case of men ($r = 0.27$, $p < 0.001$) than in case of women ($r = 0.38$, $p < 0.001$). The curvilinear effect of Age on Self-Regulation in the male subsample (n = 681) was examined also with polynomial regression analysis, where powers of standardized age were entered consecutively into the regression model up to power 5. The last power that significantly increased the R-square value was the cubic term [$R^2 = 0.08$; R^2_increase = 0.01; $F_{increase}(1,677) = 9.13$, $p = 0.003$]. This cubic curvilinear effect of Age on Self-Regulation in the male subsample is, however, negligible relative to the substantially stronger linear relationship that can be seen in the female subsample (n = 2,135), where no power increased significantly the linear effect of age ($R^2 = 0.14$).

The effects of the other sociodemographic variables were assessed separately in the two studies. Regarding level of

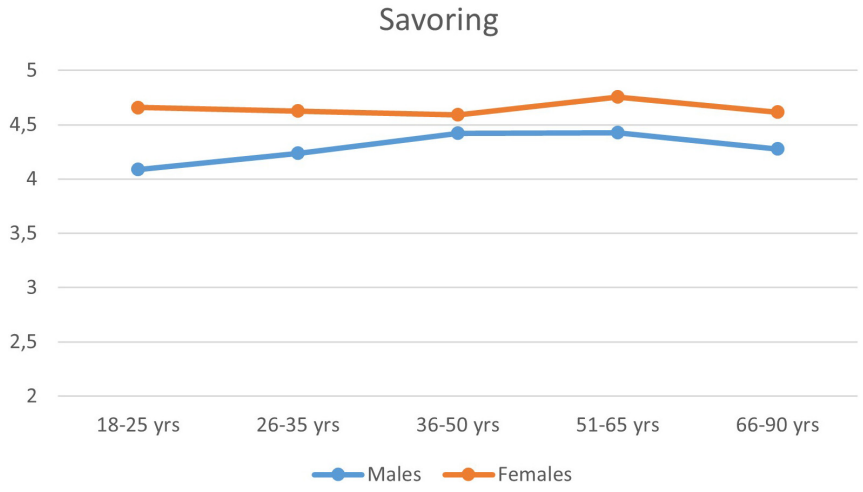

FIGURE 1
Savoring means in gender by age categories (n = 2,817).

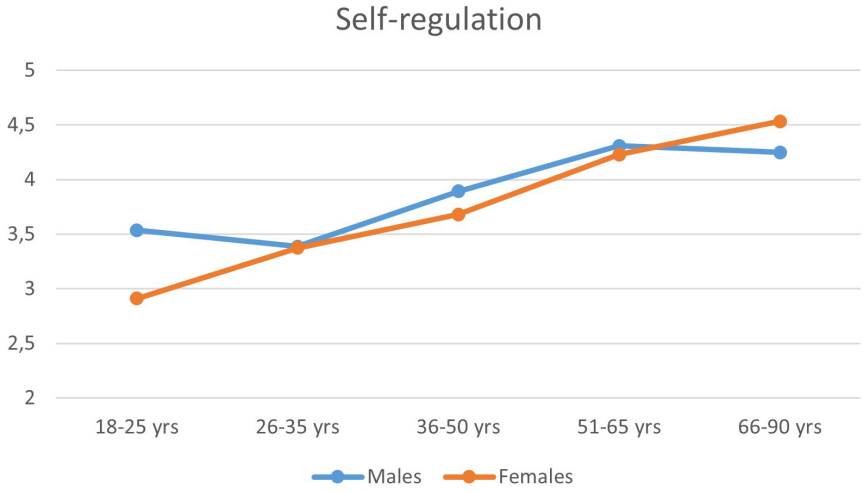

FIGURE 2
The mean of the Self-regulation scale by gender and age zone (n = 2,817).

education, the strongest positive correlation was with the Creative and Executive Efficiency scale (r = 0.16 in sample I, r = 0.13 in sample II, $p < 0.001$). Subjective financial status showed a stronger relationship in both samples. In Sample I, subjective financial status was positively related to Well-being, with r = 0.21, and to Creative and Executive Efficiency, with r = 0.12 (both significant at $p < 0.001$ level). However, in Sample II subjective financial status also had a significant positive relationship with Well-being (r = 0.23), Resilience (r = 0.20), and Creative and Executive Efficiency (r = 0.17), all at $p < 0.001$ level. A slightly stronger but similar pattern was obtained in this sample for correlations with monthly income (Well-being: r = 0.17, Resilience: r = 0.27, Creative and Executive Efficiency: r = 0.25). Financial status therefore is positively related to well-being, resilience, and coping with difficult situations. Finally, the importance of religion in Sample II had a weak but significant ($p < 0.001$) positive relationship with the Well-being (r = 0.14) and Savoring (r = 0.13) scales.

Temporal stability of the mental health test

In Study I, participants could voluntarily provide their email addresses for a possible replication of the online investigation. Participants who provided this information were asked to fill in the same questionnaire 2 weeks (Time 2) and 11 months (Time 3) after the first investigation (Time 1). This enabled

TABLE 11 Test–retest correlations of the MHT scales in Study I for replication after 2 weeks (Time 2; n = 581) and 11 months (Time 3; n = 270).

Scale	Time 1 vs. Time 2	Time 1 vs. Time 3
Well-being	0.762	0.642
Savoring	0.774	0.623
Creative and Executive Efficiency	0.799	0.652
Self-regulation	0.838	0.709
Resilience	0.784	0.697
MHT Total	0.882	0.755

df = 579; 268; $^{*}p < 0.05$, $^{**}p < 0.01$.

us to compute test–retest correlations to measure the temporal stability of the MHT scales. In major categories of financial status (below average, average, above average) the differences among Time 1, Time 2, and Time 3 percentages were never greater than 2 percentage points. Regarding age the Time 1 mean of Time 2 retention sample (52.3 years) differed only slightly from that of Time 1 sample (51.3 years). Those Ss filling in the questionnaires both at Time 1 and at Time 3 were slightly older at Time 1 than those who filled in the questionnaires only at Time 1 (54.4 vs. 50.6 years; $t(1734) = 5.18$, $p < 0.001$; Cohen's $d = 0.33$, 95% CI [0.203, 0.451]).

At Time 2, we had 648 usable cases and Cronbach's alpha values varied between 0.79 and 0.87, whereas at Time 3 we had 304 usable cases and Cronbach's alpha values varied between 0.79 and 0.90. It can therefore be concluded that the internal consistency of the MHT scales shows excellent temporal stability for all scales at Time 2 and Time 3. The test–retest correlations between the Time 1 and Time 2, and between the Time 1 and Time3 MHT scale values are summarized in **Table 11** and reflect an excellent level of temporal stability for all scales at Time 2, and a good level of temporal stability at Time 3, almost 1 year after the initial investigation.

Discussion

The objective of this paper was to conceptualize Maintainable Positive Mental Health Theory and to develop and validate the MHT that operationalizes this model. The theoretical basis of the MPMHT is a mental health construct in line with the mental health definition of the WHO (Galderisi et al., 2015). MPMHT emphasizes that the measurement of mental health must go beyond operationalizations that define the concept in terms of observable characteristics of well-being (e.g., Butler and Kern, 2016; Lukat et al., 2016), or characteristics that are listed as mirror opposites of mental disorders (e.g., Huppert and So, 2013; Caprara et al., 2019; Oláh, 2019).

Our results support the conceptual definition of MPMHT that refers to a degree of global well-being that goes hand in hand with good emotional, psychological, social, and spiritual functioning, resilience, coping, and savoring capacity, as well as mental health sustainability on an ongoing basis, with development and flexible adaptation to changing conditions guaranteed by competencies and personality factors.

Although various psychological tests exist, with different positive psychological constructs (see Sections "Introduction," "Measures," and "Measures"), the MHT is the first test to have a five-dimensional complex structure (Well-being, Savoring, Creative and Executive Efficiency, Self-regulation, and Resilience), with the aim of covering the wide spectrum of mental health.

The most important finding in our online cross-sectional studies is that the five-dimensional structural validity of the 17-item MHT was verified by EFA using a large sample (Study I, n = 1,736). In turn, using CFA, the five-factor MHT model with the original five scales was confirmed by excellent fit indices in another large and independent sample (Study II, n = 1,083). In addition, the internal consistency and temporal stability of the scales was proven. By analyzing the discriminant validity using multivariate linear regression, we were able to conclude that all five scales have a significant individual part of at least 44% that is not covered by the other four scales. The scale with the highest unique variance of 82.7% is Self-regulation, although the unique variances of the Savoring and Creative and Executive Efficiency scales also exceed 55%. The Well-being scale is the only one for which the unexplained proportion of variance is less than 50%. This suggests that, among the five components of mental health, well-being plays a central role. To expand and confirm the individual meaning of the scales, several tests and individual questionnaire items were used in correlation analyses, and sociodemographic variables in ANOVAs.

The results support that the Well-being scale confidently measures subjective well-being, which itself comprises several components (biological, psychological, social, and spiritual). The Well-being scale correlates at a high level with related tests such as Diener's Flourishing Scale, the PERMA-Profiler, the WHO Well-Being Questionnaire, the Satisfaction with Life Scale, and the Purpose in Life Test, thus it can be considered as a very similar test. An important result is that the existence of meaningful goals and the absence of depressive symptoms and maladaptive schemas are an integral part of the construct measured by the Well-being scale (see **Table 10**). Well-being is related to the individual's subjective financial status, although this relationship is weak. Well-being has the strongest relationship with the other components of mental health.

The Savoring scale measures how individuals are able to mentally mobilize their previous positive, joyful memories and experiences to generate mental well-being, reliving them in the present and, furthermore, extending them to future events. This ability appears to be more prevalent among women of all ages than men (see **Figure 1**). The results of

the validity studies show that the Savoring scale is closely related to the main indicators of all the MHTs involved: Diener's Flourishing Scale, the PERMA-Profiler, Flow, the WHO Well-Being Questionnaire, the Satisfaction with Life Scale, and the Purpose in Life Test (see **Tables 9**, **10**), which confirms the validity of its construct. The Creative and Executive Efficiency scale measures how individuals are able to cope with the difficulties they encounter by mobilizing their various competencies in the difficult, stressful, and challenging situations of life. Furthermore, it measures how individuals are able to provide successful individual and social problem-solving behavior. The Creative and Executive Efficiency scale shows strong correlations with other tests of mental health (the Flourishing Scale, the PERMA-Profiler, the Purpose in Life Test), which indicates the fulfillment of convergent validity (see **Tables 9**, **10**). It is also related to the competence component, in that a person who achieves a high value on this scale also has a better chance of reaching flow experience (see **Table 9**). Further results of the validity studies show that the Creative and Executive Efficiency scale is a negative predictor of negative emotions, symptoms of depression, and maladaptive schemas.

The Self-regulation scale provides information about a person's ability to regulate and control emotions, temperament, and negative states. Positive correlations between the Self-regulation scale and general mental state, and negative correlations with negative emotional states were consistently obtained, indicating the general good functioning of the individual (see **Tables 9**, **10**). The construct measured by the Self-regulation scale correlates weakly with age. Self-regulation is sometimes more successful in the case of older people, and this positive age effect is more dominant in women (see **Figure 2**).

The Resilience scale measures the level of mental capacities and resources that can be mobilized when a person faces unexpected, stressful, and difficult situations. The higher the level of resilience, the more quickly the individual is able to recover from a sudden, unexpected stressful situation. Our results confirm that a good capacity to experience flow, meaningful life goals, lack of negative emotional states, satisfaction with life, and the absence of depressive symptoms and maladaptive schemas contribute to a higher level of resilience (see **Tables 9**, **10**). Like self-regulation, resilience has a positive but weak relationship with age.

One of the key findings of our study is that there are competencies behind the different components of mental health. This result implies that these competences could be trained, improved, and strengthened by their nature. This is most obvious in the case of creative and executive efficiency, although, based on the positive correlation with age, it must also be true of self-regulation and resilience. Savoring also implies mental ability that can be improved with cognitive techniques. As a result, the level of experienced subjective well-being, satisfaction with life, and, in ordinary terms, happiness can be significantly increased. Although, this positive prognosis can be confirmed by subsequent research which aim to demonstrate how these competencies can be developed.

One limitation of our studies is the online recruitment of participants: despite the large size of the samples, they cannot be considered representative. The recruitment technique may greatly influence several aspects (e.g., age distribution; see **Tables 2**, **6**). Our analyses are based solely on data from verbal questionnaires, while the MHT scales are based, among other things, on specific behaviors, mental operations, attitudes, and so on. For this reason, to confirm empirical validity it would also be very important to verify the validity of the scales with other types of psychological variables (e.g., rating scales, direct observations, clinical symptoms, specific data measuring physical condition, sociometric ranking, etc.). By way of example, it would be useful to compare mental health as measured by the MHT with the variables of 360-degree studies (Mahar and Strobert, 2010). The meaning of the scales should also be confirmed by examining various clinical cases with identified psychiatric diagnoses.

Conclusion

In sum, we believe our findings show in its present, 17-item form, designed for adults, the MHT can provide a comprehensive picture of mental health in terms of MPMHT. The MHT has a number of advantages over existing measures of well-being and mental health, making it a preferable measurement device for use in future research. Applying the MPMHT approach in consultations, in skill-improvement sessions, in relationship training, in group sessions, and in behavioral therapy could help to improve the level of mental health in psychologically healthy people. Future research should examine the MHT in the clinical population and apply MPMHT in positive clinical psychology.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Research Ethics Committee of the Faculty of Education and Psychology of ELTE

(permission number: 2019/61) and the Research Ethics Committee of the Institute of Psychology of Károli Gáspár University (permission number: 70/2019/P/ET/2). The patients/participants provided their written informed consent to participate in this study.

Author contributions

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

Funding

The present study was supported by a research grant obtained from the Faculty of Humanities, Károli Gáspár University of the Reformed Church in Hungary (Person- and Family-oriented Health Study, Grant No. 2018/20643B800) and by the National Research, Development, Innovation Office of Hungary (Grant No. K 116965), and the ÚNKP-21-3 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund.

Acknowledgments

The authors appreciate the reviewers' and the editors' numerous valuable comments and suggestions, which greatly helped the finalization process of the manuscript.

Conflict of interest

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

Publisher's note

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

References

Antonovsky, A. (1991). "The structural resources of salutogenic strengths," in *Personality and stress: Individual differences in the stress process*, eds C. L. Cooper and R. Payne (Hoboken, NJ: Wiley), 67–104.

Barbera, J., Naibert, N., Komperda, R., and Pentecost, T. C. (2020). Clarity on cronbach's alpha use. *J. Chem. Educ.* 98, 257–258. doi: 10.1021/acs.jchemed.0c00183

Bech, P., Gudex, C., and Johansen, K. S. (1996). The WHO (Ten) well-being index: Validation in diabetes. *Psychother. Psychosom.* 65, 183–190. doi: 10.1159/000289073

Beck, A. T., Guth, D., Steer, R. A., and Ball, R. (1997). Screening for major depression disorders in medical inpatients with the beck depression inventory for primary care. *Behav. Res. Ther.* 35, 785–791. doi: 10.1016/S0005-7967(97)00025-9

Block, J. H., and Block, J. (2014). "The role of ego-control and ego-resiliency in the organization of behavior," in *Development of cognition, affect, and social relations*, ed. W. A. Collins (Milton Park: Psychology Press), 49–112.

Block, J., and Kremen, A. M. (1996). IQ and ego-resiliency: Conceptual and empirical connections and separateness. *J. Pers. Soc. Psychol.* 70, 349–371. doi: 10.1037/0022-3514.70.2.349

Bohlmeijer, E. T., and Westerhof, G. J. (2021). "A new model for sustainable mental health: Integrating well-being into psychological treatment," in *Making an impact on mental health: The applications of psychological research*, eds J. N. Kirby and P. Gilbert (Milton Park: Routledge/Taylor & Francis Group), 153–188.

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 (Thousand Oaks, CA: Sage), 136–162.

Bryan, M. A., Mallik, D., Cochran, G., and Lundahl, B. (2022). Mindfulness and savoring: A commentary on savoring strategies and their implications for addiction treatment. *Substace Use and Misuse* 57, 822–826. doi: 10.1080/10826084.2022.2046090

Bryant, F. B., and Veroff, J. (2007). *Savoring: A new model of positive experience*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Butler, J., and Kern, M. L. (2016). The PERMA-Profiler: A brief multidimensional measure of flourishing. *Int. J. Wellbeing* 6, 1–48.

Caprara, G. V., Alessandri, G., and Caprara, M. (2019). Associations of positive orientation with health and psychosocial adaptation: A review of findings and perspectives. *Asian J. Soc. Psychol.* 22, 126–132.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd Edn. Mahwah, NJ: Lawrence Erlbaum.

Connor, K. M., and Davidson, J. R. (2003). Development of a new resilience scale: The Connor–Davidson Resilience Scale (CD–RISC). *Depress. Anxiety* 18, 76–82. doi: 10.1002/da.10113

Crumbaugh, J. C., and Maholick, L. T. (1964). An experimental study in existentialism: The psychometric approach to Frankl's concept of noogenic neurosis. *J. Clin. Psychol.* 20, 200–207. doi: 10.1002/1097-4679(196404)20:2<200::AID-JCLP2270200203<3.0.CO;2-U

DeVellis, R. F. (2016). *Scale development: Theory and applications*, 4th Edn. Thousand Oaks, CA: Sage Publications.

Diener, E. D., Emmons, R. A., Larsen, R. J., and Griffin, S. (1985). The satisfaction with life scale. *J. Pers. Assess.* 49, 71–75. doi: 10.1207/s15327752jpa4901_13

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., et al. (2009). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Soc. Indic. Res.* 97, 143–156. doi: 10.1007/s11205-009-9493-y

Dodge, R., Daly, A., Huyton, J., and Sanders, L. (2012). The challenge of defining wellbeing. *Int. J. Wellbeing* 2, 222–235. doi: 10.5502/ijw.v2i3

Elliot, A. J., Thrash, T. M., and Murayama, K. (2011). A longitudinal analysis of self–regulation and well–being: Avoidance personal goals, avoidance coping, stress generation, and subjective well–being. *J. Pers.* 79, 643–674. doi: 10.1111/j.1467-6494.2011.00694.x

Fredrickson, B. (2009). *Positivity: Top-notch research reveals the 3-to-1 ratio that will change your life.* New York, NY: Harmony Press.

Galderisi, S., Heinz, A., Kastrup, M., Beezhold, J., and Sartorius, N. (2015). Toward a new definition of mental health. *World Psychiatry* 14, 231–233. doi: 10.1002/wps.20231

Gao, C., Shi, D., and Maydeu-Olivares, A. (2020). Estimating the maximum likelihood root mean square error of approximation (RMSEA) with non-normal data: A Monte-Carlo study. *Struct. Equ. Model.* 27, 192–201. doi: 10.1080/10705511.2019.1637741

Hattori, M., Zhang, G., and Preacher, K. J. (2017). Multiple local solutions and geomin rotation. *Multivariate Behav. Res.* 52, 720–731. doi: 10.1080/00273171.2017.1361312

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

Huppert, F., and So, T. C. (2013). Flourishing across Europe: Application of a new conceptual framework for defining well-being. *Soc. Indic. Res.* 110, 837–861. doi: 10.1007/s11205-011-9966-7

Jahoda, M. (1958). *Current concepts of positive mental health*. New York, NY: Basic Books, doi: 10.1037/11258-000

Joseph, S., and Maltby, J. (2014). Positive functioning inventory: Initial validation of a 12-item self-report measure of well-being. *Psychol. Well-Being* 4, 1–13. doi: 10.1186/s13612-014-0015-6

Kasser, T., and Ryan, R. M. (1996). Further examining the American dream: Differential correlates of intrinsic and extrinsic goals. *Pers. Soc. Psychol. Bull.* 22, 280–287. doi: 10.1177/0146167296223006

Kenny, D. A. (2015). *Measuring model fit.* Available online at: https://davidakenny.net/cm/fit.htm (accessed May 30, 2022).

Keyes, C. L. M. (2007). Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health. *Am. Psychol.* 62, 95–108. doi: 10.1037/0003-066x.62.2.95

Keyes, C. L. M., Sohail, M. M., Molokwu, N. J., Parnell, H., Amanya, C., Kaza, V. G. K., et al. (2020). How would you describe a mentally healthy person? A cross-cultural qualitative study of caregivers of orphans and separated children. *J. Happiness Stud.* 22, 1719–1743. doi: 10.1007/s10902-020-00293-x

Konkolÿ Thege, B., and Martos, T. (2006). Purpose in Life Test – Characteristics of the Hungarian Adaptation. *Mentálhigiéné és Pszichoszomatika* 7, 153–169. doi: 10.1556/mental.7.2006.3.1

Lamers, S. M., Westerhof, G. J., Bohlmeijer, E. T., ten Klooster, P. M., and Keyes, C. L. (2011). Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF). *J. Clin. Psychol.* 67, 99–110.

Lehmann, D. R. (1988). An Alternative Procedure for Assessing Convergent and Discriminant Validity. *Appl. Psychol. Meas.* 12, 411–423. doi: 10.1177/014662168801200409

Linton, M. J., Dieppe, P., and Medina-Lara, A. (2016). Review of 99 self-report measures for assessing well-being in adults: Exploring dimensions of well-being and developments over time. *BMJ open* 6, e010641.

Lomas, T. (2021). Life balance and harmony. *Int. J.* 11, 50–68. doi: 10.5502/ijw.v11i1.1477

Lomas, T., Waters, L., Williams, P., Oades, L. G., and Kern, M. L. (2020). Third wave positive psychology: Broadening towards complexity. *J. Posit. Psychol.* 16, 660–674. doi: 10.1080/17439760.2020.1805501

Lukat, J., Margraf, J., Lutz, R., van der Veld, W. M., and Becker, E. S. (2016). Psychometric properties of the Positive Mental Health Scale (PMH-scale). *BMC Psychol.* 4:8. doi: 10.1186/s40359-016-0111-x

Lyubomirsky, S. (2010). *The how of happiness: A practical guide to getting the life you want*. London: Piatkus.

Magyaródi, T., Nagy, H., Soltész, P., Mózes, T., and Oláh, A. (2013). Psychometric properties of a newly established flow state questionnaire. *J. Happiness Wellbeing* 1 (Budapest), 85–96.

Magyaródi, T., Nagy, H., Soltész, P., Mózes, T., and Oláh, A. (2014). "Psychometric properties of a new flow questionnaire," in *Flow, emotional intelligence and psychological immunity: Empirical studies in positive psychological perspective*, eds A. Oláh and H. Nagy (ELTE Eötvös Kiadó), 21–42.

Mahar, J. A., and Strobert, B. (2010). The use of 360-degree feedback compared to traditional evaluative feedback for the professional growth of teachers in K-12 education. *Plan. Chang.* 41, 147–160.

Martos, T., Sallay, V., Désfalvi, J., Szabó, T., and Ittzés, A. (2014). Psychometric characteristics of the Hungarian version of the Satisfaction with Life Scale (SWLS-H). *Mentálhigiéné és Pszichoszomatika* 15, 289–303. doi: 10.1556/mental.15.2014.3.9

Martos, T., Szabó, G., and Rózsa, S. (2006). Psychometric characteristics of the shortened Aspiration Index in Hungarian sample. *J. Mental Health Psychosom.* 7, 171–191. doi: 10.1556/Mental.7.2006.3.2

Maydeu-Olivares, A. (2017). Maximum Likelihood Estimation of Structural Equation Models for Continuous Data: Standard Errors and Goodness of Fit. *Struct. Equ. Model.* 24, 383–394. doi: 10.1080/10705511.2016.1269606

Maydeu-Olivares, A., Shi, D., and Rosseel, Y. (2019). Instrumental variables two-stage least squares (2SLS) vs. maximum likelihood structural equation modeling of causal effects in linear regression models. *Struct. Equ. Model.* 26, 876–892. doi: 10.1080/10705511.2019.1607740

Mayerson, N. H. (2020). The character strengths response: An urgent call to action. *Front. Psychol.* 11:2106. doi: 10.3389/fpsyg.2020.02106

Muthén, L. K., and Muthén, B. O. (1998–2011). *Mplus user's guide*, 6th Edn. Los Angeles, CA: Muthén & Muthén.

Nagy, H., Magyaródi, T., Vargha, A., and Oláh, A. (2021). Development of the short hungarian version of the savoring beliefs inventory. *J. Mental Health Psychosom.* 23, 95–111. doi: 10.1556/0406.23.2022.003

Oláh, A. (2005). *Anxiety, coping and flow. Empirical studies in interactional perspective*. Budapest: Trefort Press.

Oláh, A. (2016). "A new model and a new inventory of well-being," in *Complete Book of Abstracts of the Session of the 8th European Conference on Positive Psychology* (Budapest: Akadémia Kiadó), 55.

Oláh, A. (2019). Positivity is an up—to—date predictor of well—functioning instead of a eudaemon. *Asian J. Soc. Psychol.* 22, 143–145. doi: 10.1111/ajsp.12362

Oláh, A. (2021). "Health protective and health promoting resources in personality: A framework for the measurement of the psychological immune system," in *Anxiety, coping and flow*, ed. A. Oláh (Budapest: Akadémia Press), 191–225.

Oláh, A., and Kapitány-Fövény, M. (2012). Ten years of positive psychology. *Magyar Pszichológiai Szemle* 67, 19–45. doi: 10.1556/mpszle.67.2012.1.3

Oláh, A., Nagy, H., Magyaródi, T., Török, R., Csengődi, M., and Vargha, A. (2018). "Developing a new test of mental health, MHT, with five sub-scales," in *9th European conference on positive psychology, book of abstracts*, eds C. Bognár and K. Móricz (Budapest: Akadémiai Kiadó), 408–409.

Oláh, A., Vargha, A., Csengődi, M., Bagdi, B., and Diósi, T. (2020). *Magyarország Boldogságtérképe*. Available online at: http://boldogsagprogram.hu/magyarorszag-boldogsagterkepe-2020/ (accessed May 30, 2022)

Rózsa, S., Szádóczky, E., and Füredi, J. (2001). Psychometric characteristics of the short form of the beck depression inventory in hungarian sample. *Psychiatr. Hung.* 16, 384–402.

Ryff, C. D. (1995). Psychological well-being in adult life. *Curr. Dir. Psychol. Sci.* 4, 99–104. doi: 10.1111/1467-8721.ep10772395

Ryff, C. D., and Marshall, W. V. (1999). *The self and society in aging processes*. Location Salmon Tower Building, NY: Spring Publishing Company.

Seligman, M. (2018). PERMA and the building blocks of well-being. *J. Posit. Psychol.* 13, 333–335. doi: 10.1080/17439760.2018.1437466

Shi, D., Lee, T., and Maydeu-Olivares, A. (2018). Understanding the Model Size Effect on SEM Fit Indices. *Educ. Psychol. Meas.* 79, 310–334. doi: 10.1177/0013164418783530

Singh, S., and Sharma, N. R. (2018). Self-regulation as a correlate of psychological well-being. *Indian J. Health Wellbeing* 9, 441–444.

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., and Bernard, J. (2008). The brief resilience scale: Assessing the ability to bounce back. *Int. J. Behav. Med.* 15, 194–200. doi: 10.1080/10705500802222972

Southwick, S. M., and Charney, D. S. (2018). *Resilience: The science of mastering life's greatest challenges.* Cambridge University Press.

Steiger, J. H., and Lind, J. (1980). *Statistically-based tests for the number of common factors [Paper presentation]. The annual spring meeting of the psychometric society*. Iowa City, IA.

Susánszky, É, Konkolÿ Thege, B., Stauder, A., and Kopp, M. (2006). Validation of the short (5-item) version of the WHO Well-Being questionnaire based on a Hungarian representative health survey (Hungarostudy 2002. *J. Ment. Health Psychosom.* 7, 247–255.

Unoka, Zs, Rózsa, S., Fábián, Á, Mervó, B., and Simon, L. (2004). Young's Schema Questionnaire: Study of psychometric characteristics of a test assessing early maladaptive schemas. *Psychiatria Hungarica* 19, 235–243.

Vaillant, G. (2012). *Triumph of experience: The men of the Harvard grant study*. Cambridge, MA: The Belknap Press of Harvard University Press.

Vaillant, G. E. (2003). Mental health. *Am. J. Psychiatry* 160, 1373–1384. doi: 10.1176/appi.ajp.160.8.1373

Vaillant, G. E. (2012). Positive mental health: Is there a cross-cultural definition? *World Psychiatry* 11, 93–99. doi: 10.1016/j.wpsyc.2012.05.006

Vargha, A., Torma, B., and Bergman, L. R. (2015). ROPstat: A general statistical package useful for conducting person-oriented analysis. *J. Pers. Oriented Res.* 1, 87–98. doi: 10.17505/jpor.2015.09

Verdolini, N., Amoretti, S., Montejo, L., García-Rizo, C., Hogg, B., Mezquida, G., et al. (2021). Resilience and mental health during the COVID-19 pandemic. *J. Affect. Disord.* 283, 156–164. doi: 10.1016/j.jad.2021.01.055

Vohs, K. D., and Baumeister, R. F. (2004). *Understanding self-regulation. Handbook of self-regulation, 19.* Availble online at: http://medicine.kaums.ac.ir/UploadedFiles/400/Group/motaeat%20etiad/book/Handbook%20of%20Self%20Regulation%20Research%20Theory%20and%20Applications.pdf#page=19 (accessed 30 May, 2022).

Walker, M., and Unterhalter, E. (2007). *Amartya Sen's capability approach and social justice in education*. London: Palgrave Macmillan.

Welburn, K., Coristine, M., Dagg, P., Pontefract, A., and Jordan, S. (2002). The Schema Questionnaire—Short Form: Factor analysis and relationship between schemas and symptoms. *Cogn. Ther. Res.* 26, 519–530. doi: 10.1023/A:1016231902020

Windle, G., Bennett, K. M., and Noyes, J. (2011). A methodological review of resilience measurement scales. *Health Qual. Life Outcomes* 9:8. doi: 10.1186/1477-7525-9-8

10.3389/fpsyg.2022.775622

Appendix

The Mental Health Test (MHT)

The following statements are designed to provide information about your perceptions of wellness. Please consider each statement carefully and thoughtfully, then enter an X to indicate the response option with which you most agree. There are no right or wrong answers.

	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
(1) Joy is present more than sorrow in my everyday life./W						
(2) I easily become impatient./SR						
(3) It's easy for me to revive the joy from pleasant memories./S						
(4) I tend to bounce back quickly after hard times./R						
(5) I often have ideas that are taken further by others./C						
(6) I have a hard time making it through stressful events./R						
(7) Others describe me as a problem solver./C						
(8) I am impulsive: I act first and think second./SR						
(9) I can successfully achieve targets which I set for myself./C						
(10) I like to store memories of fun times that I go through so that I can recall them later./S						
(11) It does not take me long to recover from a stressful event./R						
(12) I can make myself feel good by imagining what a happy time that is about to happen will be like./S						
(13) I tend to take a long time to get over set-backs in my life./R						
(14) My general psychological state is good./W						
(15) I am good at work that needs new and original ideas./C						
(16) I easily become impatient./SR						
(17) I often know what people are thinking and feeling./C						
(18) How do you feel about your life as a whole?/W (1: Very bad, 6: Very good)						

Scaling guide

Well-being (W): The average of the scores for items 1, 14, and 18.
Savoring (S): The average of the scores for items 3, 10, and 12.
Creative and Executive Efficiency (C): The average of the scores for items 5, 7, 9, 15, and 17.
Self-regulation (SR): The average of the scores for items 2, 8, and 16, after inverting all three items with the 7 – x transformation (x is the original, 7 – x is the inverted score).
Resilience (R): The average of the scores for items 4, 11, and 13, after inverting item 13 with the 7 – x transformation (x is the original, 7 – x is the inverted score).

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership