psychology international

MDPI

Article

Psychometric Concerns with the Ego-Resilience Scales in South Africa: Evidence from Classical Test Theory and Item Response Theory

Tyrone B. Pretorius * and Anita Padmanabhanunni *

Department of Psychology, University of the Western Cape, Cape Town 7530, South Africa
* Correspondence: tpretorius@uwc.ac.za (T.B.P.); apadmana@uwc.ac.za (A.P.)

Abstract

The dimensionality of the Ego-Resilience Scale (ER89) remains contested, and little is known about the performance of the original and revised versions of the scale in South Africa. This study examined the psychometric properties of the 14-item ER89 and the 10-item ER89-R in 337 undergraduate students from a university in the Western Cape province of South Africa. Participants completed the ER89, the Sense of Coherence Scale, the Center for Epidemiological Studies Depression Scale, and the trait scale of the State-Trait Anxiety Inventory. The scales were evaluated using classical test theory, confirmatory factor analysis, bifactor modeling, parallel analysis, Mokken scale analysis, and Rasch analysis. For both versions, bifactor models fit better than one-factor models, but the specific factors showed anomalous loadings and accounted for little reliable variance. Parallel analysis and Rasch analysis supported an essentially unidimensional interpretation of both instruments. Although both versions showed acceptable internal consistency and expected associations with sense of coherence, depression, and anxiety, several findings raised concerns about their broader psychometric adequacy. Overall, both scales appeared to reflect a broad general factor of ego-resilience, but their performance in this context was mixed and, in important respects, problematic. Further refinement and validation are needed before either instrument can be recommended for confident use in South African research and practice.

Keywords: dimensionality; ego-resilience; Ego-Resiliency Scale; Mokken analysis; Rasch analysis

Check for updates

Academic Editor: Alexander Robitzsch

Received: 11 May 2026
Revised: 13 June 2026
Accepted: 18 June 2026
Published: 22 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Ego-resiliency refers to the capacity to flexibly modify behavior and self-regulation in response to changing situational demands so as to maintain or restore adaptive functioning (Block & Kremen, 1996). Within Block's framework, it reflects adaptive modulation, resourceful responding, and context-sensitive adjustment. Higher ego-resiliency has been associated with better adaptation and personal achievement across the lifespan (Alessandri et al., 2016; Syed et al., 2020).

Block and Kremen (1996) developed the 14-item Ego-Resiliency Scale (ER89), which has been widely used in research on stress, coping, adjustment, and mental health. For example, it has been used to investigate individuals' capacity to adapt ego control in the context of pandemic-related stressors (Mpanya et al., 2022). It has also been applied in adolescent research to examine the associations between ego-resilience, social support, and depression, highlighting its relevance for understanding protective psychological processes

during development (Kang et al., 2022). In caregiving contexts, the ER89 has been used to explore the relationships between self-esteem, ego-resilience, and caregiver burden among caregivers of children with chronic diseases (Kim & Park, 2023). In addition, the scale has been used in occupational health research, where ego-resilience was found to moderate the relationship between occupational stress and compassion fatigue among nurses (Liu et al., 2024).

Despite its broad application, however, the latent structure of the ER89 remains unresolved. This is a substantive concern given that uncertainty about dimensionality affects how scores are interpreted, whether total scores can be justified, and whether the instrument can be used confidently in research and applied settings. Block and Kremen (1996) and Letzring et al. (2005) treated the scale as unidimensional, relying largely on coefficient alpha and scree plot inspection. Later work argued that this evidence was limited and proposed more complex structures. Fonzi and Menesini (2005) proposed a revised higher-order solution comprising two first-order factors, labeled Resiliency Self-Regulation and Resiliency-Openness, which culminated in the shorter ER89-R. Equally, in their validation study, Alessandri et al. (2007) conceptualized ego-resiliency as a second-order construct defined by two first-order components which they labeled Optimal Regulation and Openness to Life Experience. Farkas and Orosz (2015) proposed an alternative three-component model of ego-resiliency based on an 11-item refined version of the instrument, comprising Active Engagement with the World, Repertoire of Problem-Solving Strategies, and Integrated Performance under Stress. Thus, support has been reported for one-, two-, and three-factor solutions.

This lack of resolution of the latent factor structure of the scale is particularly important in South Africa, where psychometric assessment takes place in multilingual, multicultural, and socioeconomically unequal settings in which equivalence and method bias cannot be assumed (Laher & Cockcroft, 2014). Instruments developed and validated primarily in Western, educated, industrialized contexts may not function equivalently when applied to South African samples, particularly when constructs are shaped by local social, educational, linguistic, and cultural conditions. South African scholars have therefore emphasized the need for local validation of imported measures (Laher, 2024; Oppong et al., 2023). This is especially important for resilience-related constructs, which are increasingly used in South African research and practice to understand adaptation in contexts marked by adversity, inequality, and mental health risk. However, the ER89 and ER89-R have received limited local evaluation, despite the growing use of resilience-related constructs in South African research and practice (Jorgensen & Seedat, 2008; Padmanabhanunni & Pretorius, 2023).

The present study addressed this gap by examining the psychometric properties of the ER89 and ER89-R in a South African student sample. Specifically, we evaluated dimensionality using confirmatory factor analysis (CFA), bifactor indices, parallel analysis, Mokken scale analysis (MSA), and Rasch analysis; examined item and scale performance using classical test theory, MSA, and Rasch indices; and assessed convergent validity through associations with sense of coherence, depression, and trait anxiety. By comparing the original and revised versions across multiple psychometric paradigms, the study sought to determine whether these instruments yield defensible scores in this context. The study contributes to the broader literature by clarifying the dimensionality and score interpretation of the ER89 and ER89-R, while also contributing locally to the validation of resilience-related measures for use in South African research and practice.

2. Materials and Methods

2.1. Participants and Procedures

The participants were 337 undergraduate students at a university in the Western Cape province of South Africa. An electronic questionnaire was developed in Google Forms and distributed by the Registrar's office to a random sample of 1500 students. Invitation emails included the survey link, and reminder emails were sent weekly for four months. During this time, weekly reminder emails were distributed to encourage participation. The extended recruitment period was intended to maximize the response rate; however, it may have coincided with routine academic events such as assessment periods, examinations, or semester recess, which could have influenced students' availability, willingness to participate, or response patterns. It is also possible that repeated reminder emails contributed to survey fatigue among some students, particularly those who chose not to participate. The survey also used forced-response options, which minimized item-level missing data. However, because the survey was anonymous, it was not possible to determine how many students opened the link but did not complete the survey.

A sensitivity analysis was conducted using the R package "simsem" (Pornprasertmanit et al., 2025) in R version 4.3.1 (R Development Core Team, 2020) to determine the minimum sample size required to obtain satisfactory fit indices: root mean square error of approximation (RMSEA) < 0.06, comparative fit index (CFI) > 0.95, Tucker–Lewis index (TLI) > 0.95, and standardized root mean square residual (SRMR) < 0.08, as well as the actual fit indices achieved with 337 participants in the current sample with 80% power. In this regard, we ran 500 CFA simulations of a bifactor model of the ER89 as originally proposed (Alessandri et al., 2007) with 14 items, a total scale and two subscales as conceptualized in the original development study, medium factor loadings (0.60), and 80% power. The results of the sensitivity analysis found that a sample size of 295 would be sufficient to produce the following satisfactory fit indices with 80% power, RMSEA = 0.022, CFI = 0.993, TLI = 0.993, and SRMR = 0.073, while the actual sample of 337 would result in the following satisfactory fit indices with 80% power, RMSEA = 0.021, CFI = 0.994, TLI = 0.994, and SRMR = 0.070. The sample of 337 participants in the current study also meets the recommended requirement of a minimum of 150 participants to obtain item calibration stability within ± 0.5 logits with 99% confidence in Rasch analysis (Linacre, 1994), as well as the minimum of 250 participants recommended for MSA (Watson et al., 2018).

Most participants were women (77.2%) and lived in an urban area (65.4%). The mean age was 21.95 years (SD = 4.68; range = 16–50).

2.2. Instruments

Participants completed a brief demographic questionnaire, the ER89, the 13-item Sense of Coherence Scale (SOC: Antonovsky, 1987), the Center for Epidemiological Studies Depression Scale (CES-D: Radloff, 1977), and the trait scale of the State-Trait Anxiety Inventory (STAI-T: Spielberger, 1985). The latter three instruments were included to examine convergent validity.

The ER89 consists of 14 items scored on a 4-point response scale from 1 (does not apply) to 4 (applies very strongly), with higher scores indicating greater ego-resilience. The ER89-R (Alessandri et al., 2007) is a 10-item version which excludes items 3, 4, 6, and 13 from the original ER89 based on unacceptable factor loadings as part of the published revised version of the instrument.

The SOC-13 assesses the extent to which life is perceived as comprehensible, manageable, and meaningful. It is scored on a 7-point scale, with higher scores reflecting a stronger sense of coherence. Prior work has reported acceptable reliability internationally

and in South Africa, where it has been found to be essentially unidimensional (Eriksson & Mittelmark, 2017; T. B. Pretorius & Padmanabhanunni, 2023).

The CES-D is a 20-item screening measure of depressive symptoms scored from 0 to 3, and the STAI-T is a 20-item measure of general anxiety proneness scored from 1 to 4. Both have shown satisfactory reliability and validity, including in South African samples (Barnes et al., 2002; Padmanabhanunni et al., 2022; T. Pretorius & Padmanabhanunni, 2021; T. B. Pretorius, 1991).

2.3. Ethics

The Humanities and Social Sciences Research Ethics Committee of the University of the Western Cape granted ethical approval (Ethics reference number HS20/5/1, 1 June 2020). The study was conducted in accordance with the Declaration of Helsinki. Participation was voluntary, informed consent was obtained electronically, no incentives were offered, and no personal information was collected.

2.4. Data Analysis

There were no missing data because all items were mandatory in the online survey. Classical test theory (CTT) analyses were conducted in SPSS version 31 (IBM Corporation, Armonk, NY, USA), MSA in R version 4.3.1 using the "mokken" package (Van der Ark, 2007), CFA in Amos version 28 (IBM Corporation, Armonk, NY, USA), and Rasch analysis in Winsteps version 5.8 (Linacre, 2025).

The CTT analyses included interitem correlations, item-total correlations (ITCs), reliability coefficients (alpha, omega, and composite reliability), average variance extracted (AVE), maximum shared variance (MSV), and average shared variance (ASV). Interitem correlations between 0.15 and 0.85 and mean interitem correlations above 0.30 were taken to indicate adequate item relatedness (Clark & Watson, 2019; Hajjar, 2018) and reliability coefficients above 0.70 acceptable reliability (DeVon et al., 2007). The literature recommends ITCs > 0.30 as acceptable (Nunnally & Bernstein, 1994) and >0.50 as good (Hajjar, 2018) in respect of item contribution. As the goal was to cross-culturally validate the ER89 and ER89-R and since a correlation of 0.30 is a medium correlation reflecting that the latent construct only shares 9% of variance with the item, we adopted >0.50 as the threshold for ITCs. AVE values above 0.50 were interpreted as evidence of convergent validity (Cheung et al., 2024), while AVE values exceeding both MSV and ASV were interpreted as evidence of discriminant validity (Almén et al., 2018).

MSA was used to examine scalability, dimensionality, monotonicity, and invariant item ordering (IIO). The automated item selection procedure (AISP) was used to identify scalable item clusters at the default lower bound c = 0.30 as recommended by Mokken (2011) as well as in increments of 0.05 up to 0.55 to examine the stability of the scale. Scale-level H values below 0.30 were treated as unscalable, 0.30 to 0.39 as weak, 0.40 to 0.49 as moderate, and 0.50 or higher as strong; item-level Hi values above 0.30 were considered acceptable (Mokken, 2011). Monotonicity and IIO violations were evaluated with Crit values, with values above 80 indicating serious violations (Crișan et al., 2022). The H^T coefficient was used to assess the accuracy of IIO (Watson et al., 2012), and MSRHO was used as an estimate of reliability.

We used the rating scale model in Rasch analysis since the 4-point response format was the same for all items. In Rasch analysis, dimensionality was assessed by principal components analysis of residuals, with a first contrast eigenvalue below 2 taken to support unidimensionality. Infit and outfit mean square values between 0.5 and 1.5 were regarded as acceptable. Item and person separation and reliability were used to assess how well the scale differentiated item difficulty and respondent standing on the latent trait. A

Wright map was inspected to evaluate the spread of item difficulty, and differential item functioning (DIF) was used to assess invariance across gender, with DIF values below 0.50 indicating negligible bias (Linacre, 2025). We also examined the functioning of the 4-point response format by looking at whether the response categories were used sufficiently, the observed average measures advanced monotonically, infit and outfit statistics were acceptable (should be close to 1 and below 2), and whether the Andrich thresholds were ordered (Linacre, 2025).

Although previous studies have proposed one-, two-, and three-factor structures for the ER89, the present study did not seek to provide an exhaustive comparative test of all published models. These models differ in item composition, scale version, and conceptual specification, which limits their direct comparability. Instead, the CFA and bifactor analyses were used to evaluate whether the ER89 and ER89-R could be interpreted as reflecting a sufficiently strong general ego-resilience factor to justify total-score interpretation in the present sample. CFA was therefore used to evaluate two models of the structure of the ER89 and ER89-R: a one-factor model, in which the instruments are conceptualized as consisting of a single latent factor with 14 (ER89) and 10 (ER89-R) items, respectively, as indicators, and a bifactor model comprising the total score, referred to as the general factor, and two subscales, referred to as specific factors. The extent to which each model fit the data was evaluated with the following fit indices (Gaskin et al., 2022; Hu & Bentler, 1999):

- Chi-squared (χ^2): should be nonsignificant, but this is very seldom achieved as a nonsignificant value would indicate a perfect fit (Jöreskog et al., 2016).
- Relative χ^2 (χ^2/df): between 1 and 3 reflects a good fit, and if it is <5, it would indicate an acceptable fit.
- CFI and TLI: for both these indices, values $\geq$0.95 would indicate a good fit, and $\geq$0.90 would indicate an acceptable fit.
- RMSEA: a value <0.06 indicates a good fit, while a value <0.08 indicates an acceptable fit.
- SRMR: a value of <0.08 indicates a good fit, while a value of <0.10 indicates an acceptable fit.
- Akaike information criterion (AIC): is a model comparison index, and lower AIC values indicate better fitting models.

Where bifactor models fit better, ancillary bifactor indices were examined. Explained common variance (ECV) was used to estimate variance attributable to the general factor and the specific factors, and omega hierarchical coefficients (ωH) were used to assess whether the total score and subscale scores accounted for reliable variance. Percent of uncontaminated correlations (PUCs) refers to the percentage of correlations that only reflect variance from the general factor. Consistent with prior guidance, ECV values above 0.70, ωH values above 0.80, and ωHS (for subscales) values below 0.50, as well as PUC values greater than 0.70, were taken to support essential unidimensionality (Rodriguez et al., 2016; Schmitt et al., 2018). Where doubts exist, it is recommended that PUC, ECV, and ωH should be considered together, and if PUC < 0.80, ECV > 0.60, and ωH > 0.70, there is some evidence of dimensionality but it is not sufficient to rule out the conclusion that the instrument under consideration should be considered essentially unidimensional (Reise et al., 2013).

Parallel analysis with 1000 simulated datasets was used to determine the number of meaningful factors. In parallel analysis, the eigenvalues in the current dataset are compared with the mean or the 95th percentile of eigenvalues obtained in simulated datasets (n = 1000), and only factors in the current dataset with eigenvalues greater than the mean or the 95th percentile of eigenvalues in the simulated datasets represent meaningful factors (Goretzko, 2025).

To assess convergent validity, Pearson correlations were computed between the ER89 scores and scores on sense of coherence, depression, and anxiety. As the construct of ego-resilience is conceptualized as a protective factor, we predicted that it would be significantly positively related to sense of coherence, which is also a protective factor, and significantly negatively related to indices of psychological distress. We used Cohen's guidelines of Pearson's r = 0.10, 0.30, and 0.50 to interpret observed effect sizes as small, medium, or large, respectively (Cohen, 1992). Because Pearson's r assumes approximate normality, skewness and kurtosis values between −2 and +2 were regarded as indicative of normality (Hair et al., 2021).

3. Results

3.1. Dimensionality

Figure 1 depicts the one-factor and bifactor models tested for the ER89 and ER89-R.

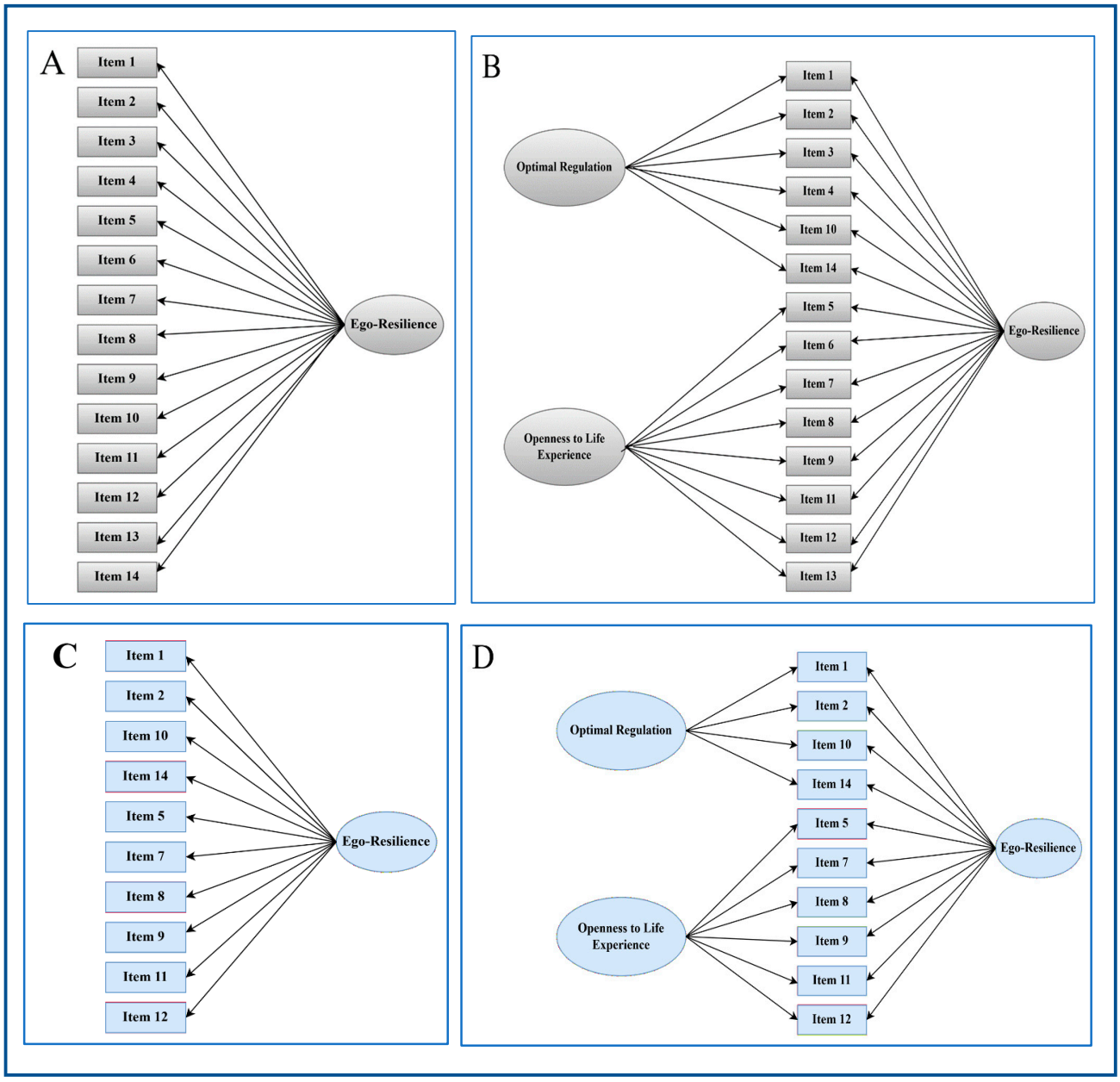

Figure 1. Two models of the structure of the ER89 and ER89-R. (**A**) One-factor model of ER89; (**B**) bifactor model of ER89; (**C**) one-factor model of ER8-R; and (**D**) bifactor model of ER89-R.

Table 1 presents the fit indices obtained with CFA for the four models and shows that for both the ER89 and the ER89-R, the bifactor model demonstrated a better fit to the data than the one-factor model and yielded lower AIC values. The one-factor models showed a mixed pattern of fit, with acceptable RMSEA and SRMR values but poorer CFI and TLI values. By contrast, the bifactor models produced a consistently better fit, with no poor fit indices for the ER89 and a slightly improved fit for the ER89-R.

Table 1. Fit indices for two models of the ER89 and two models of the ER89-R.

Fit Index	Good Fit Criteria	ER89		ER89-R	
		One-Factor	Bifactor	One-Factor	Bifactor
χ^2(df)		204.39 (77)	136.71 (63)	92.91 (35)	59.39 (26)
p-value	Nonsignificant	<0.001	<0.001	<0.001	<0.001
χ^2/df	Between 1 and 3	2.65	2.17	2.66	2.28
CFI	≥0.95	0.88	0.92	0.88	0.93
TLI	≥0.95	0.84	0.90	0.84	0.88
RMSEA [90% CI]	<0.06	0.07 [0.06, 0.08]	0.06 [0.05, 0.07]	0.07 [0.05, 0.09]	0.06 [0.04, 0.08]
SRMR	<0.08	0.06	0.05	0.06	0.05
AIC	Lower levels	260.39	220.71	132.91	117.39

Note. CFI = comparative fit index, TLI = Tucker–Lewis index, RMSEA = root mean square error of approximation, SRMR = standardized root mean square residual, and AIC = Akaike information criterion.

Table 2 displays the factor loadings of the two models of the ER89 and the ER89-R. Although the bifactor models fit better, the factor loadings in Table 2 were problematic. In the one-factor model, six ER89 items and seven ER89-R items loaded below 0.50. A similar pattern appeared for the general factor in the bifactor models, where some loadings were as low as 0.24 for the ER89 and 0.27 for the ER89-R. The specific factors were especially unstable, with negative, zero, and nonsignificant loadings in both scales.

Table 2. The factor loadings of two models of the ER89 and the ER89-R.

Item	ER89				ER89-R			
	One-Factor	Bifactor			One-Factor	Bifactor		
		G-Factor	S-Factor 1	S-Factor 2		G-Factor	S-Factor 1	S-Factor 2
1. I am generous with my friends	0.37 **	0.35 **	0.18		0.40 **	0.42 **	1.32	
2. I recover quickly from being startled	0.52 **	0.51 **	0.12		0.52 **	0.55 **	0.00	
3. I enjoy new and unusual situations †	0.61 **	0.63 **	−0.15					
4. I like making a favorable impression †	0.54 **	0.53 **	0.00					
5. I enjoy trying new foods	0.27 **	0.37 **		0.38 **	0.39 **	0.27 **		−0.37
6. Regarded as a very energetic person †	0.45 **	0.58 **		−0.07				
7. I like to take different paths to places	0.35 **	0.66 **		0.28 *	0.63 **	0.50 **		−0.48
8. I am more curious than most	0.59 **	0.49 **		0.02	0.48 **	0.37 **		−0.32
9. Most people are likeable	0.63 **	0.44 **		−0.21 *	0.41 **	0.43 **		−0.03
10. I think carefully before acting	0.49 **	0.24 **	0.52		0.31 **	0.36 **	0.04	
11. I like to do new and different things	0.45 **	0.59 **		0.13	0.60 **	0.47 **		−0.42
12. Life full of things keeps me interested	0.59 **	0.53 **		−0.25 *	0.49 **	0.52 **		0.00
13. Myself as a pretty strong personality †	0.53 **	0.64 **		−0.32 *				
14. I overcome anger reasonably quickly	0.63 **	0.44 **	0.34		0.48 **	0.58 **	−0.07	

Notes. G-factor = general factor (ego-resilience total score), S-factor 1 = specific factor 1 (optimal regulation), S-factor 2 = specific factor 2 (openness to life experience), and † = items not included in the ER89-R. * $p < 0.05$; ** $p < 0.001$.

Ancillary bifactor indices indicated that the general factor dominated the item variance. For the ER89 and ER89-R, respectively, the general factor explained 80.1% and 74.3% of the common variance, whereas the specific factors accounted for much smaller proportions. The omega hierarchical for the general factor was 0.82 for the ER89 and 0.72 for the ER89-R, while omega hierarchical subscale coefficients were very low for all specific factors. PUC was 0.53 for both versions of the ER89. Since PUC < 0.80, and ECV and ωH of the general

factor was greater than 0.60 and 0.70, respectively, these findings indicate that the total score captures the bulk of reliable variance, whereas the subscales do not.

Parallel analysis supported a one-factor solution for both versions of the scale, as only the first observed eigenvalue exceeded both the mean and the 95th percentile of the simulated eigenvalues. Rasch analysis yielded the same conclusion: the first contrast eigenvalue from the residual PCA was below 2 for both the ER89 (1.74) and ER89-R (1.67), indicating that a single dimension was sufficient to account for residual item variance.

Only MSA suggested more than one scale. At lower bound c = 0.30, AISP identified items 1, 5, and 9 as unscalable in both versions; grouped items 2, 7, 8, 11, and 12 together in both scales; and grouped items 10 and 14 into a second scale. In the 14-item ER89, items 3, 4, 6, and 13 formed part of the first cluster. For both versions of the ER89, some lower bounds were found to be greater than the maximum item-pair scalability coefficients; thus, the maximum lower bound for the ER89 was 0.50 and for the ER89-R it was 0.40. At lower bound c = 0.50, 12 of the 14 items of the ER89 were found to be unscalable, while for the ER89-R at the maximum lower bound c = 0.40, six items were found to be unscalable, two grouped on one scale and two on another. The results of the AISP at the various lower bounds for the ER89 and the ER89-R are included as Table A1 in Appendix A.

Overall, however, the weight of evidence across methods supported an essentially unidimensional interpretation of both instruments.

3.2. Item Performance

Table 3 presents the CTT, Rasch, and MSA item-level indices. Interitem correlations were generally low, ranging from 0.05 to 0.47 for the ER89 and from 0.05 to 0.40 for the ER89-R. Several correlations were nonsignificant and many fell below the recommended lower bound of 0.15. Mean interitem correlations were also low (0.25 for the ER89 and 0.22 for the ER89-R), and most item-total correlations fell below 0.50.

Table 3. The CTT, Rasch, and Mokken indices for the items of the ER89 and ER89-R.

Index	Items													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Classical Test Theory Indices														
Item 1 †	—	0.23 ***	0.17 ***	0.28 ***	0.18 ***	0.20 ***	0.22 ***	0.18 ***	0.22 ***	0.20 ***	0.24 ***	0.15 **	0.15 **	0.16 **
Item 2	0.23 ***	—	0.41 ***	0.27 ***	0.17 **	0.27 ***	0.35 ***	0.26 ***	0.17 **	0.15 **	0.23 ***	0.26 ***	0.27 ***	0.36 ***
Item 3			—	0.32 ***	0.20 ***	0.31 ***	0.47 ***	0.31 ***	0.20 ***	0.06	0.42 ***	0.35 ***	0.34 ***	0.22 ***
Item 4				—	0.22 ***	0.41 ***	0.26 ***	0.25 ***	0.35 ***	0.13 *	0.23 ***	0.23 ***	0.40 ***	0.18 ***
Item 5	0.18 ***	0.17 **			—	0.18 ***	0.34 ***	0.12 *	0.10	0.05	0.31 ***	0.07	0.13 *	0.19 ***
Item 6						—	0.39 ***	0.31 ***	0.25 ***	0.09	0.32 ***	0.31 ***	0.43 ***	0.20 ***
Item 7	0.22 ***	0.35 ***			0.34 ***		—	0.36 ***	0.19 ***	0.07	0.40 ***	0.29 ***	0.33 ***	0.27 ***
Item 8	0.18 ***	0.26 ***			0.12 *		0.36 ***	—	0.23 ***	0.07	0.35 ***	0.22 ***	0.31 ***	0.12 *
Item 9	0.22 ***	0.17 **			0.10		0.19 ***	0.23 ***	—	0.17 *	0.24 ***	0.26 ***	0.35 ***	0.25 ***
Item 10	0.20 ***	0.15 **			0.05		0.07	0.07	0.17 **	—	0.21 ***	0.22 ***	0.21 ***	0.28 ***
Item 11	0.24 ***	0.23 ***			0.31 ***		0.40 ***	0.35 ***	0.24 ***	0.21 ***	—	0.29 ***	0.35 ***	0.22 ***
Item 12	0.15 **	0.26 ***			0.07		0.29 ***	0.22 ***	0.26 ***	0.22 ***	0.29 ***	—	0.42 ***	0.31 ***
Item 13													—	0.34 ***
Item 14	0.16 **	0.36 ***			0.19 ***		0.27 ***	0.11 *	0.25 ***	0.28 ***	0.22 ***	0.31 ***		—
ER89 and ER89_R: Mean	3.52	2.90	2.65	3.04	3.01	2.87	2.87	3.09	2.93	3.25	3.07	2.44	2.97	2.74
ER89 and ER89_R: SD	0.76	0.80	0.91	0.81	0.97	0.96	1.01	0.83	0.75	0.77	0.84	0.85	0.95	0.97
ER89: ITC	0.35	0.48	0.53	0.49	0.31	0.51	0.56	0.43	0.41	0.26	0.54	0.47	0.57	0.42
ER89-R: ITC	0.35	0.45			0.31		0.52	0.38	0.36	0.28	0.51	0.41		0.43
Rasch Indices														
ER89: Infit MnSq	1.32	0.79	0.84	0.84	1.45	1.04	1.07	0.97	0.80	1.15	0.85	0.83	0.97	1.16
ER89-R: Infit MnSq	1.20	0.86	0.83	0.85	1.42	1.01	1.05	1.04	0.85	1.21	0.84	0.86	1.04	1.14
ER89: Outfit MnSq	1.25	0.78			1.37		1.07	0.98	0.81	1.07	0.84	0.85		1.08
ER89-R: Outfit MnSq	1.12	0.81			1.34		1.04	1.02	0.87	1.09	0.82	0.87		1.05
ER89: DIF Gender	0.00	0.37	0.00	0.10	0.44	0.06	0.29	0.03	0.34	0.30	0.00	0.12	0.08	0.33
ER89-R: DIF Gender	0.02	0.36			0.45		0.29	0.00	0.36	0.32	0.00	0.11		0.32
Mokken Indices														
ER89: H_i	0.24	0.30	0.34	0.30	0.20	0.32	0.34	0.27	0.27	0.17	0.33	0.31	0.34	0.27
ER89-R: H_i	0.24	0.29			0.20		0.31	0.24	0.24	0.18	0.32	0.28		0.27
ER89: SE of H_i	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.03	0.03	0.03	0.03
ER89-R: SE of H_i	0.04	0.03			0.03		0.03	0.03	0.03	0.03	0.03	0.04		0.03
ER89: Crit for monotonicity	25	0	18	0	42	0	0	0	25	0	0	0	0	0
ER89-R: Crit for monotonicity	0	0			29	0	0	0	0	0	0	0		0
ER89: Crit for IIO	0	93	0	63	127	70	112	40	127	30	67	0	138	27
ER89-R: Crit for IIO	0	55			53		90	67	73	65	103	0		0

Note. ER89 interitem correlations above the diagonal; ER89-R below the diagonal. † For item descriptions, see Table 2. ITC = item-total correlations; IIO = invariant item ordering. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

The item means in Table 3 indicate that item 1 was the easiest to endorse and item 12 the most difficult, a pattern also evident in the Wright map (Figure 2). Despite the weak CTT item relationships, Rasch fit statistics were acceptable: infit and outfit mean square

values fell within the recommended range for both scales, and all DIF values by gender were below 0.50.

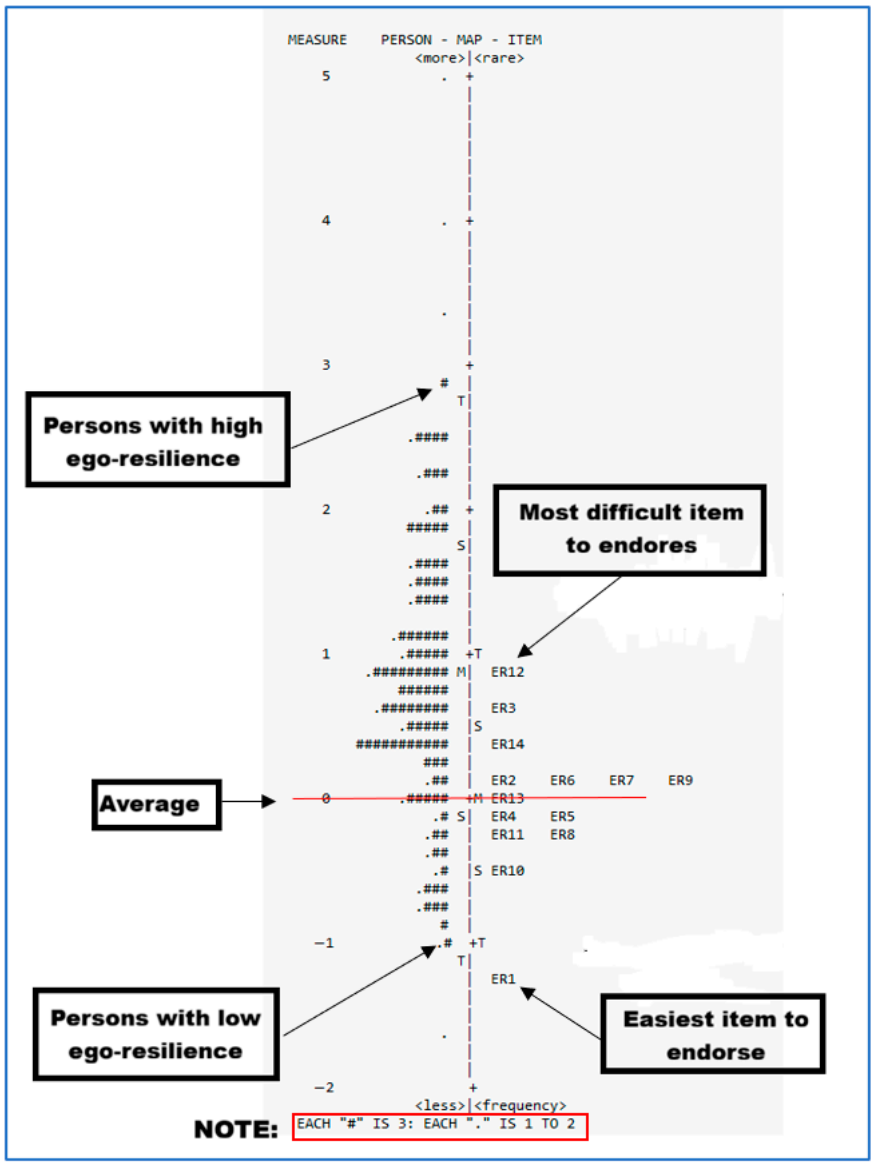

Figure 2. A Wright map with persons plotted on the left and item difficulty on the right.

The MSA results were less favorable. Item scalability coefficients were very low, with most items at or below 0.30 in both versions. Monotonicity violations were few and minor, but IIO violations were more frequent, and several items (items 2, 5, 79, and 13 in the vase of the ER89; 7 and 11 in the case of the ER89-R) showed Crit values above 80, indicating serious departures from invariant item ordering.

3.3. Performance of the ER89 and ER89-R at the Scale Level

At the scale level (Table 4), internal consistency estimates for both the ER89 and ER89-R were acceptable, with alpha, omega, and composite reliability all above 0.70. The MSV and ASV excluded the correlation between the two versions of the ER89 and MSV is based on the correlation between ego-resilience and anxiety, while the ASV is based on the correlations between ego-resilience and all other variables are used to determine convergent validity. As can be seen in Table 4, AVE was below 0.50 for both scales, and for the ER89 it was equal to MSV, which does not support strong convergent or discriminant validity.

Table 4. The CTT, Rasch, and Mokken indices for the ER89 and ER89-R at the scale level.

Index	Threshold	ER89	ER89-R
Classical test theory indices			
Alpha	>0.70	0.82	0.74
Omega	>0.70	0.82	0.74
Composite reliability	>0.70	0.86	0.81
Average variance extracted (AVE)	>0.50	0.31	0.30
Maximum shared variance (MSV)	<AVE	0.31	0.27
Average shared variance (ASV)	<AVE	0.18	0.16
Rasch indices			
Item reliability	>0.80	0.97	0.98
Item separation index	>3	6.08	6.73
Person reliability	>0.80	0.79	0.70
Person separation index	>2	1.96	1.53
Mokken indices			
H	>0.50	0.29	0.26
H^T	>0.50	0.14	0.17
MS_{RHO}	>0.70	0.83	0.74

Note. H = scalability coefficient to assess the strength of the scale; H^T = coefficient to assess invariant item ordering.

Rasch analysis showed excellent item separation and item reliability for both scales, indicating adequate spread in item difficulty. Person separation and person reliability, however, were slightly below recommended thresholds for the ER89 and clearly below threshold for the ER89-R, suggesting limited ability to distinguish among respondents along the latent trait continuum.

The MSA scale coefficients were also poor. H values were below 0.30 for both versions, indicating unscalable scales, and H^T values were far below 0.50, indicating weak invariant item ordering at the scale level.

3.4. Performance of the Response Format

In both versions of the ER89 the category structure supported the functioning of the four-point response format. All four response categories were sufficiently endorsed, with category frequencies ranging from 7% to 39% in the case of the ER89 and 7% to 38% in the case of the ER89-R. Observed average measures increased monotonically across categories for both versions of the ER89, indicating that higher response categories corresponded to higher levels of the latent trait. Category infit and outfit mean square values were close to 1.00 and within acceptable limits. The Andrich thresholds advanced in the expected order from −1.28 to −0.08 to 1.35 logits for the ER89 and from −1.33 to −0.05 to 1.38 logits for the ER89-R, supporting ordered threshold functioning. Overall, the four-point rating scale demonstrated adequate category functioning.

3.5. Convergent Validity

The skewness and kurtosis values for all study variables fell within acceptable limits, indicating approximate normality (Table 5). Both versions of the ER89 correlated in the expected directions with the external variables: they were positively associated with sense of coherence and negatively associated with depression and anxiety. The correlations with sense of coherence and depression were of moderate magnitude, whereas the correlations with anxiety were large.

Table 5. Intercorrelations, descriptive statistics, distribution indices and reliabilities.

Scale/Variables	1	2	3	4	5
1. ER89	—				
2. ER89-R	0.96 **	—			
3. Sense of Coherence	0.45 **	0.42 **	—		
4. Depression	−0.42 **	−0.42 **	−0.71 **	—	
5. Anxiety	−0.56 **	−0.52 **	−0.77 **	0.79 **	—
Mean	41.36	29.82	50.06	27.50	48.12
SD	6.75	4.70	12.68	13.36	10.53
Skewness	−0.33	−0.37	−0.06	0.02	0.25
Kurtosis	−0.42	−0.31	−0.09	−0.94	−0.17
α	0.82	0.74	0.81	0.92	0.90
ω	0.82	0.74	0.81	0.92	0.90

Note. ER89 = Ego-Resilience Scale-89; ER89-R = Revised Ego-Resilience Scale-89. ** $p < 0.001$.

4. Discussion

This study examined the psychometric properties of the ER89 and ER89-R in a South African student sample using complementary methods. The clearest finding was that both instruments are best understood as reflecting a broad general factor rather than reliable subscales. Although bifactor models fit better than one-factor models, the specific factors showed anomalous and negative loadings and negligible reliable variance once the general factor was taken into account. The most likely explanation for the anomalous and negative loadings is that a dominant general factor accounted for nearly all of the item variance, with the result that the specific factors "vanish empirically" (Eid et al., 2017, p. 544). Parallel analysis and Rasch analysis independently converged on the same conclusion, and the bifactor indices likewise indicated essential unidimensionality, particularly for the ER89.

At the same time, the findings do not support an unqualified endorsement of either scale. Some results were encouraging. Internal consistency estimates were acceptable across methods, Rasch item fit statistics were within the recommended range, DIF suggested that men and women understood the items similarly, and the scales correlated with sense of coherence, depression, and anxiety in theoretically expected directions. These findings provide some support for the interpretation of ER89 scores as indicators of a broad resilience-related construct.

However, the broader psychometric pattern was mixed and often problematic. Several indicators pointed to weak item functioning and limited construct coherence. Interitem correlations and item-total correlations were generally low, AVE values were below recommended thresholds, and most Hi coefficients were poor. At the scale level, the H coefficients suggested that the items did not form strong Mokken scales, while the H^T coefficients indicated weak invariant item ordering. Person separation and person reliability were also suboptimal, particularly for the ER89-R, suggesting that the scales do not distinguish especially well between respondents with different levels of ego-resilience. In addition, AVE was equal to MSV for the ER89, raising further concerns about discriminant validity.

These findings have practical implications. First, the subscale scores for Optimal Regulation and Openness to Life Experience should not be interpreted as distinct, psychometrically defensible dimensions in this context. Second, the ER89-R does not appear to offer a clear advantage over the original ER89. The two versions were extremely highly correlated, yet the shorter scale did not show better psychometric performance. Shortening the instrument therefore did not resolve the underlying measurement concerns. At present, both scales are better regarded as provisional research indicators of general ego-resilience than as stand-alone instruments for individual screening, profiling, or intervention decisions.

Several limitations should be noted. The study was conducted at a single university, and all data were based on self-report. Although the sample was predominantly female,

this should be interpreted in light of broader higher education enrolment patterns, where women often outnumber men (García-Holgado & García-Peñalvo, 2022). This imbalance may reflect patterns of self-selection into voluntary survey research, but it also means that the performance of the ER89 and ER89-R among male students could not be examined with the same degree of confidence. Although this does not in itself undermine the reliability of the collected data, it limits the extent to which the findings can be assumed to apply equally across gender groups.

The study is also limited by the low response rate. Given the voluntary nature of participation, students who completed the survey may have differed systematically from those who did not, and selection bias cannot be ruled out. In addition, because the survey was anonymous and did not collect personal identifiers, in accordance with South Africa's Protection of Personal Information Act, it was not possible to determine whether any participant completed the survey more than once. These factors limit the representativeness of the sample and require cautious generalization of the findings. The use of forced-response items reduced missing data but also meant that missingness patterns could not be examined. It is possible that some participants may have provided less considered responses when required to answer all items; however, this possibility could not be directly assessed in the present study. The cross-sectional design also meant that temporal stability, sensitivity to change, and predictive utility could not be examined. Future studies should test revised item sets, examine performance in more diverse South African samples, and evaluate test–retest reliability and predictive validity over time.

5. Conclusions

This study evaluated the psychometric properties of the ER89 and ER89-R in a South African student sample using multiple complementary methods. Both versions of the scale appear to reflect a broad general factor, and the proposed subscales do not justify separate interpretation. Although some findings were favorable, the overall psychometric profile was mixed and, in important respects, problematic. Neither version can therefore be regarded as fully satisfactory for confident use in this context without further refinement and validation. Future refinement should begin with a qualitative item review of items that performed inconsistently across methods, followed by the development or revision of items and cross-validation of the revised scale in an independent South African sample.

Author Contributions: Conceptualization, T.B.P. and A.P.; Formal analysis, T.B.P.; Investigation, T.B.P. and A.P.; Methodology, T.B.P.; Project administration, A.P.; Supervision, A.P.; Writing—original draft, T.B.P. and A.P.; Writing—review and editing, T.B.P. and A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Humanities and Social Science Research Ethics Committee of the University of the Western Cape (Ethics reference number HS20/5/1, 1 June 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original data presented in the study are openly available in FigShare at https://doi.org/10.6084/m9.figshare.32242464.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. The results of the automated item selection procedure at various different lower bounds.

Item	ER89					ER89-R		
	Lower Bounds					Lower Bounds		
	0.30	0.35	0.40	0.45	0.50	0.30	0.35	0.40
Item 1	0	0	0	0	0	0	0	0
Item 2	1	1	3	0	0	1	2	2
Item 3	1	1	1	1	1			
Item 4	1	2	0	0	0			
Item 5	0	0	0	0	0	0	0	0
Item 6	1	1	2	0	0			
Item 7	1	1	1	1	1	1	1	1
Item 8	1	1	0	0	0	1	1	0
Item 9	0	2	0	0	0	0	0	0
Item 10	2	0	0	0	0			
Item 11	1	1	1	1	0	2	0	0
Item 12	1	1	2	2	0	1	1	1
Item 13	1	1	2	2	0	1	0	0
Item 14	2	0	3	0	0	2	2	2

References

Alessandri, G., Eisenberg, N., Vecchione, M., Caprara, G. V., & Milioni, M. (2016). Ego-resiliency development from late adolescence to emerging adulthood: A ten-year longitudinal study. *Journal of Adolescence*, *50*, 91–102. [CrossRef] [PubMed]

Alessandri, G., Vecchio, G. M., Steca, P., Caprara, M. G., & Caprara, G. V. (2007). A revised version of Kremen and Block's Ego Resiliency Scale in an Italian sample. *Testing, Psychometrics, Methodology in Applied Psychology*, *14*(3–4), 165–183.

Almén, N., Lundberg, H., Sundin, Ö., & Jansson, B. (2018). The reliability and factorial validity of the Swedish version of the Recovery Experience Questionnaire. *Nordic Psychology*, *70*(4), 324–333. [CrossRef]

Antonovsky, A. (1987). *Unraveling the mystery of health: How people manage stress and stay well*. Jossey-Bass.

Barnes, L. L. B., Harp, D., & Jung, W. S. (2002). Reliability generalization of scores on the spielberger state-trait anxiety inventory. *Educational and Psychological Measurement*, *62*(4), 603–618. [CrossRef]

Block, J., & Kremen, A. M. (1996). IQ and ego-resiliency: Conceptual and empirical connections and separateness. *Journal of Personality and Social Psychology*, *70*(2), 349. [CrossRef] [PubMed]

Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management*, *41*(2), 745–783. [CrossRef]

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, *31*(12), 1412–1427. [CrossRef] [PubMed]

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. [CrossRef] [PubMed]

Crișan, D. R., Tendeiro, J. N., & Meijer, R. R. (2022). The Crit coefficient in Mokken scale analysis: A simulation study and an application in quality-of-life research. *Quality of Life Research*, *31*(1), 49–59. [CrossRef] [PubMed]

DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship*, *39*(2), 155–164. [CrossRef] [PubMed]

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in *G*-factor models: Explanations and alternatives. *Psychological Methods*, *22*(3), 541–562. [CrossRef] [PubMed]

Eriksson, M., & Mittelmark, M. (2017). The sense of coherence and its measurement. In M. Mittelmark, S. Sagy, G. Bauer, J. Pelikan, B. Lindstrom, & G. Espnes (Eds.), *The handbook of salutogenesis* (pp. 97–103). Springer Open.

Farkas, D., & Orosz, G. (2015). Ego-resiliency reloaded: A three-component model of general resiliency. *PLoS ONE*, *10*(3), e0120883. [CrossRef] [PubMed]

Fonzi, A., & Menesini, E. (2005). Strategie di coping e caratteristiche di resilienza in adolescenza. *Psicologia Clinica dello Sviluppo*, *9*(3), 437–456. [CrossRef]

García-Holgado, A., & García-Peñalvo, F. J. (2022). A model for bridging the gender gap in STEM in higher education institutions. In F. J. García-Peñalvo, A. García-Holgado, A. Dominguez, & J. Pascual (Eds.), *Women in STEM in higher education: Good practices of attraction, access and retainment in higher education* (pp. 1–19). Springer Nature Singapore. [CrossRef]

Gaskin, J., Lim, J., & Steed, J. (2022). *"Model fit summary", AMOS plugin*. Available online: http://statwiki.gaskination.com (accessed on 8 February 2025).

Goretzko, D. (2025). How many factors to retain in exploratory factor analysis? A critical overview of factor retention methods. *Psychological Methods*. Advance online publication. [CrossRef]

Hair, J. F., Jr., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage Publications.

Hajjar, S. (2018). Statistical analysis: Internal-consistency reliability and construct validity. *International Journal of Quantitative and Qualitative Research Methods*, *6*(1), 46–57.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. [CrossRef]

Jorgensen, I. E., & Seedat, S. (2008). Factor structure of the Connor-Davidson resilience scale in South African adolescents. *International Journal of Adolescent Medicine and Health*, *20*(1), 23–32. [CrossRef]

Jöreskog, K. G., Olsson, U. H., & Wallentin, F. Y. (2016). Confirmatory factor analysis (CFA). In K. G. Jöreskog, U. H. Olsson, & F. Y. Wallentin (Eds.), *Multivariate analysis with LISREL* (pp. 283–339). Springer. [CrossRef]

Kang, M., Min, A., & Min, H. (2022). Predicting multicultural adolescents' depression in South Korea: Roles of ego-resilience and social support. *Journal of Pediatric Nursing*, *65*, e93–e98. [CrossRef] [PubMed]

Kim, J. Y., & Park, B. K. (2023). Relationships among self-esteem, ego-resilience, and caregiver burden among families of children with hematologic and oncologic disease: A cross-sectional study. *Heliyon*, *9*(12). [CrossRef] [PubMed]

Laher, S. (2024). Advancing an agenda for psychological assessment in South Africa. *South African Journal of Psychology*, *54*(4), 515–530. [CrossRef]

Laher, S., & Cockcroft, K. (2014). Psychological assessment in post-apartheid South Africa: The way forward. *South African Journal of Psychology*, *44*(3), 303–314. [CrossRef]

Letzring, T. D., Block, J., & Funder, D. C. (2005). Ego-control and ego-resiliency: Generalization of self-report scales based on personality descriptions from acquaintances, clinicians, and the self. *Journal of Research in Personality*, *39*(4), 395–422. [CrossRef]

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurements Transactions*, *7*, 328.

Linacre, J. M. (2025). *A user's guide to Winsteps version 5.10.4*. Winsteps.com. Available online: https://www.winsteps.com/winman/copyright.htm (accessed on 11 December 2025).

Liu, D., Xie, S., Jing, J., Niyomsilp, E., Xie, L., Nie, X., & Liang, Y. (2024). The effect of perceived organizational support and ego-resilience on the relationship between occupational stressors and compassion fatigue in COVID-19 frontline nurses: A cross-sectional study in Sichuan, China. *BMC Nursing*, *23*(1), 817. [CrossRef] [PubMed]

Mokken, R. J. (2011). *A theory and procedure of scale analysis*. De Gruyter Mouton.

Mpanya, J. M., Mukendi, A. K., Seker, M., & Turhan, T. (2022). Examination of ego resilience. *Anadolu Akademi Sosyal Bilimler Dergisi*, *4*(2), 123–130.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw Hill.

Oppong, S., Appiah, R., Hapunda, G., & Kheswa, J. G. (2023). Editorial: Contextualizing psychological assessment in Africa: COVID-19 and beyond. *Frontiers in Psychology*, *14*, 1150387. [CrossRef] [PubMed]

Padmanabhanunni, A., & Pretorius, T. B. (2023). The resilience of South African school teachers in the time of COVID-19: Coping with risk of infection, loneliness, and anxiety. *International Journal of Environmental Research and Public Health*, *20*(4), 3462. [CrossRef] [PubMed]

Padmanabhanunni, A., Pretorius, T. B., Stiegler, N., & Bouchard, J.-P. (2022). A serial model of the interrelationship between perceived vulnerability to disease, fear of COVID-19, and psychological distress among teachers in South Africa. *Annales Médico Psychologiques*, *180*(1), 23–28. [CrossRef] [PubMed]

Pornprasertmanit, S., Miller, P., Schoemann, A. M., & Jorgensen, T. D. (2025). *simsem: SIMulated structural equation modeling. R package version 0.5-17*. Available online: https://cran.r-project.org/web/packages/simsem/ (accessed on 11 December 2025).

Pretorius, T., & Padmanabhanunni, A. (2021). A looming mental health pandemic in the time of COVID-19? Role of fortitude in the interrelationship between loneliness, anxiety, and life satisfaction among young adults. *South African Journal of Psychology*, *51*, 256–268. [CrossRef]

Pretorius, T. B. (1991). Cross-cultural application of the Center for Epidemiological Studies Depression Scale: A study of black South African students. *Psychological Reports*, *69*(Suppl. S3), 1179–1185. [CrossRef]

Pretorius, T. B., & Padmanabhanunni, A. (2023). Is less more? Reevaluating the psychometric properties of the sense of coherence-13 and a revised seven-item version in South Africa using classical theory and item response theory. *Psych*, *5*(4), 1030–1045. [CrossRef]

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385–401. [CrossRef]

R Development Core Team. (2020). *R: A language and environment for statistical computing*.. R Foundation for Statistical Computing. Available online: http://www.R-project.org/ (accessed on 12 December 2023).

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, *73*(1), 5–26. [CrossRef]

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223–237. [CrossRef] [PubMed]

Schmitt, T. A., Sass, D. A., Chappelle, W., & Thompson, W. (2018). Selecting the "best" factor structure and moving measurement validation forward: An illustration. *Journal of Personality Assessment*, *100*(4), 345–362. [CrossRef] [PubMed]

Spielberger, C. D. (1985). Assessment of state and trait anxiety: Conceptual and methodological issues. *Southern Psychologist*, *2*(4), 6–16.

Syed, M., Eriksson, P. L., Frisén, A., Hwang, C. P., & Lamb, M. E. (2020). Personality development from age 2 to 33: Stability and change in ego resiliency and ego control and associations with adult adaptation. *Developmental Psychology*, *56*(4), 815. [CrossRef] [PubMed]

Van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software*, *20*, 1–19. [CrossRef]

Watson, R., Egberink, I. J. L., Kirke, L., Tendeiro, J. N., & Doyle, F. (2018). What are the minimal sample size requirements for Mokken scaling? An empirical example with the Warwick-Edinburgh Mental Well-Being Scale. *Health Psychology & Behavioral Medicine*, *6*(1), 203–213. [CrossRef]

Watson, R., van der Ark, L. A., Lin, L.-C., Fieo, R., Deary, I. J., & Meijer, R. R. (2012). Item response theory: How Mokken scaling can be used in clinical practice: Item response theory. *Journal of Clinical Nursing*, *21*(19pt20), 2736–2746. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.