Authenticity Scale: Validity and Reliability Evidence in a Sample from Brazil and Portugal

Vinicius Coscioni[1,*], Marco Antônio Pereira Teixeira[2], and Maria Paula Paixão[3]

1 University of Coimbra, Faculty of Psychology and Educational Sciences, CINEICC; Universidade Federal do Rio Grande do Sul, Porto Alegre/RS (Brazil)
2 Universidade Federal do Rio Grande do Sul, Porto Alegre/RS (Brazil)
3 University of Coimbra, Faculty of Psychology and Educational Sciences, CINEICC (Portugal)

Título: Escala de Autenticidad: Evidencia de Validez y Confiabilidad en una Muestra de Brasil y Portugal.
Resumen: Este estudio presenta distintos tipos de evidencias de validez y confiabilidad de la Escala de Autenticidad (AS) en una muestra de Brasil y Portugal. El estudio consiste en una encuesta con 1.077 brasileños y 622 portugueses. Se testó el modelo con tres factores correlacionados (autoalienación, vivir auténtico y aceptación de la influencia externa), el modelo unidimensional y el modelo bifactorial. Se retuvo el modelo con tres factores correlacionados, con las tres subescalas alcanzando confiabilidad moderada a buena. Análisis factorial confirmatorio multigrupo sugirió invariancia escalar para cultura, género, edad, educación, ocupación y preocupación e impacto relacionados con Covid. Los ítems fueron evaluados por *graded response model* (GRM), sugiriendo que las tres subescalas no discriminan las personas con altos rasgos de autenticidad. GRM y estadísticas descriptivas indican que la escala de puntuación es inapropiada, particularmente para la subescala vivir auténtico, que es afectada por efecto techo. Las asociaciones con presencia de sentido mostraron evidencia adicional de validez. A pesar de las limitaciones, la AS es una medida adecuada para evaluar la autenticidad en diferentes grupos. Se discuten posibles modificaciones para el aprimoramiento de la AS.
Palabras clave: Autenticidad. Validez. Confiabilidad. Análisis factorial. Invariancia de medida. Modelo de respuesta graduada. Sentido de la vida.

Abstract: This study introduces distinct types of validity and reliability evidence of the Authenticity Scale (AS) in a sample from Brazil and Portugal. It consists of an online survey with 1,077 Brazilian citizens and 622 Portuguese citizens. The study tested the model with three correlated factors (self-alienation, authentic living, and accepting external influence), the unidimensional model, and the bi-factor model. The model with three correlated factors was retained, with the three subscales demonstrating moderate to good reliability. Multigroup confirmatory factor analysis suggested scalar invariance across culture, gender, age, education, occupation, and Covid-related concern and impact. The items were assessed by graded response model (GRM), which suggested that the three subscales are not able to distinguish respondents with high authenticity traits. GRM and descriptive statistics indicated that the rating-scale is inappropriate, particularly for authentic living subscale, which is affected by ceiling effect. Associations with presence of meaning showed additional validity evidence. Despite the limitations, the AS is an effective measure to assess authenticity across different groups. Potential modifications for the improvement of the AS are discussed.
Keywords: Authenticity. Validity. Reliability. Factor analysis. Measurement invariance. Graded response model. Meaning in life.

Introduction

Knowing yourself and acting accordingly is an important personal feature often named as authenticity. As human beings live their lives, they develop a sense of self of which they become aware to a certain degree. Conversely, becoming aware of the self does not imply acting accordingly as people may be influenced by others' expectations (Barrett-Lennard, 1998). Individual differences in authenticity are key factors of personal well-being (Wood et al., 2008) and thus have become of central interest in humanistic psychology (Rogers, 1961), psychodynamics (Winnicott, 1965), and positive psychology (Gable & Haidt, 2005). Empirical psychology has developed measures to assess dispositional levels of authenticity, such as the Authenticity Inventory (Kernis & Goldman, 2005) and the Authenticity Scale (AS; Wood et al., 2008). However, the former has faced criticism for its extensive item count (Nartova-Bochaver et al., 2021) and inconsistent validity evidence (Grégoire et al., 2014). The AS has been extensively used and adapted across various cultures. This study introduces distinct types of validity and reliability evidence of the AS in a sample from Brazil and Portugal.

The Person-Centered Model of Authenticity

The AS is grounded on the person-centered model of authenticity. This is a model grounded on Rogers (1961), an exponent of humanistic psychology and one of the first theorists to discuss the notion of authenticity in psychology and counseling. Based on the person-centered approach of Rogers (1961), Barrett-Lennard (1998) defined authenticity as a tripartite construct designating "consistency between the three levels of (a) a person's primary experience, (b) their symbolized awareness, and (c) their outward behavior and communication" (p. 82). According to Wood et al. (2008), this model provides a widely accepted construct definition and dimensionality framework of authenticity. Thus, it well-suited as the theoretical background for the development of the AS.

The first dimension is referred to as self-alienation, which "involves the inevitable mismatch between the conscious awareness and actual experience" (Wood et al., 2008, p. 386). One's actual experience encompasses one's true self, which differs to some extent from one's conscious awareness. High rates of self-alienation can contribute to psychopathologies, whereas congruence between the true self and conscious awareness reflects the subjective experience of being connected to oneself.

The second and third dimensions are named authentic living and accepting external influence, respectively. Authentic living "involves behaving and expressing emotions in

* Correspondence address [Dirección para correspondencia]:
Vinicius coscioni. E-mail: viniciuscoscioni@gmail.com
(Article received: 04/04/2023; revised: 17/06/2023; accepted: 20/06/2023)

such a way that is consistent with the conscious awareness of physiological states, emotions, beliefs, and cognitions" (Wood et al., 2008, p. 386). In turn, accepting external influence "involves the extent to which one accepts the influence of other people and the belief that one has to conform to the expectations of others" (Wood et al., 2008, p. 386). Both dimensions entail the degree to which one acts in line with one's perceived self. Authentic living refers to behaving in accordance with one's values and beliefs. Conversely, accepting external influences represents the opposite, i.e., the introjection of others' expectations.

The Construction and Adaptations of the AS

The AS was originally developed in the UK and has three factors with four items each (Wood et al., 2008). The 12 items are responded to in a 7-point scale ranging from 1 (does not describe me at all) to 7 (describes me very well); intermediate scale points are not anchored. The factors assess the three dimensions of the person-centered model of authenticity. The first version with 25 items was tested by exploratory factor analysis (EFA) resulting in three factors. A brief version containing the four items with the highest loadings per factor was tested in a confirmatory factor analysis (CFA) with a second-order factor. The model presented excellent fit and reliability, as tested by alpha and test-retest correlations. The AS showed discriminant validity from the Big Five personality. Associations with well-being (self-esteem, life satisfaction, positive affect, psychological well-being, and gratitude) and ill-being (anxiety, stress, and negative affect) showed convergent validity.

The AS has been adapted to Iranian (Shamsi et al., 2012), Turkish (İlhan & Özdemir, 2013), French (Canada; Grégoire et al., 2014), Italian (Di Fabio, 2014), Swedish (Vainio & Daukantaitė, 2016), Serbian (Grijak, 2017), Portuguese (Balbino et al., 2018), Ukrainian (Zlyvkov et al., 2019), Chinese (Xia et al., 2022), Russian (Nartova-Bochaver et al., 2021), and Sinhala (Zoysa et al., 2021). In Russia, the authentic living subscale was rephrased to assume a reserve-coded form. For example, item 1 "I think it is better to be yourself than to be popular" was rephrased as "I think it is better to be popular than to be yourself." In Sweden, item 12 ("I feel alienated from myself") was rephrased to a suitable sentence in Swedish, back-translated as "I feel like a stranger to myself."

EFA was implemented in Iran (Shamsi et al., 2012), Canada (Grégoire et al., 2014), Portugal (Balbino et al., 2018), China (Xia et al., 2022), Russia (Nartova-Bochaver et al., 2021), and Sri Lanka (Zoysa et al., 2021). Item 1 yielded loadings below .50 in Canada and Sri Lanka, being removed from the latter. In Russia, item 1 migrated to accepting external influence. In addition, items 4 and 11 had loadings below .50 in Canada and Russia, respectively.

CFA were implemented in all adaptations but to Iranian (Shamsi et al., 2012), Swedish (Vainio & Daukantaitė, 2016), and Ukrainian (Zlyvkov et al., 2019). In Turkey (İlhan &

Özdemir, 2013), Canada (Grégoire et al., 2014), Italy (Di Fabio, 2014), and China (Xia et al., 2022), the original model with a second-order factor was selected. In Canada, item 1 had loadings below .50, whereas items 1, 5, 8 and 9 had low loadings in Turkey. We did not have access to the factor loadings in the Italian form, and they were not reported in the adaptation to China. The Russian form also selected a second-order model yet with item 1 loading onto accepting external influence due to the results of EFA. In addition, item 4 was removed due to its high correlations to items 5 and 10. In Portugal (Balbino et al., 2018), the 3-factor solution with no second-order factors was selected, and all loadings were above .50. The 3-factor solution was also selected in Sri Lanka (Zoysa et al., 2021) yet item 1 was removed due to the results of EFA. In addition, item 8 was correlated to items 5 and 6. In Serbia, the bifactor model was selected yet the factor loadings were not reported (Grijak, 2017).

Even though most studies tested the second-order model, CFA does not properly estimate second-order factors composed of only three first-order factors. In such a situation, the model has zero degrees of freedom and thus the fit indices may not be computed (Kline, 2011). Fit indices extracted in such circumstances pertain to the estimation of only the first-order factors. This is the reason why fit indices of second-order models and correlated-factor models are the same when the second-order factor has only three factors.

Regarding the bifactor models tested in Serbia, no justification for using such a factor structure was provided. A bifactor model include a general factor onto which all items load. This general factor reflects what is common among all items in a scale. In addition to this general factor, a bifactor model includes orthogonal specific factors. Hence, each item is an indicator of both the general factor and the orthogonal specific factors. The specific factors are not equivalent to first-order factors in a high-order structure. Rather, they represent the common variance of a subgroup of items once the common variance between all items has been partitioned out through the general factor (Dunn & McCray, 2020). Therefore, the use of bifactor models must embrace a rationale for a general factor underlining all items in a scale.

In all adaptations, reliability was assessed through alpha, with values ranging from 0.62 to 0.95. Other coefficients were extracted in Iran (Shamsi et al., 2012) and Russia (Nartova-Bochaver et al., 2021), with Spearman-Brown coefficient and omega indicating good reliability, respectively. Test-retest correlations were assessed in Canada (Grégoire et al., 2014), Serbia (Grijak, 2017) and China (Xia et al., 2022). In Serbia, low correlations were observed, which might be related to a larger interval between data collections.

Multigroup CFA have tested the invariance of the AS across different groups. In the construction study (Wood et al., 2008) and in the adaptions to China (Xia et al., 2022) and Sri Lanka (Zoysa et al., 2021), configural and metric invariance models were tested, whereas in Russia (Nartova-Bochaver et al., 2021) scalar invariance models were additionally tested. Metric invariance was observed across gender

in the construction study and in the adaptations to Sinhala and Chinese; scalar invariance was observed in Russia. Metric invariance across ethnicity and occupation was observed in the construction study and in China, respectively. Metric invariance across age was observed in China and Russia. In Russia, scalar invariance across age was not observed, and a partial model with three freed item intercepts showed good fit. Lastly, in Russia, metric invariance across depression rates was observed, although scalar models did not demonstrate good fit.

Most adaptations assessed relations to other measures but to Iranian (Shamsi et al., 2012), Portuguese (Balbino et al., 2018), and Sinhala (Zoysa et al., 2021). The adaptations included measures of well-being (psychological well-being, subjective well-being, mental well-being, life satisfaction, harmony in life, sense of coherence, need satisfaction, self-esteem, and positive affect) and ill-being (psychological distress, anxiety, negative affect, depression, and stress). The Canadian adaptation (Grégoire et al., 2014) demonstrated discriminant validity from the Big Five personality.

To our knowledge, the AS has only been assessed by item response theory (IRT) in Ukraine (Zlyvkov et al., 2019). Three polytomous models were applied (i.e., graded response model, generalized partial credit model, and graded ratings scale model), with a primary focus on comparing the models. The item parameters were not reported.

Research Goals

This study introduces distinct types of validity and reliability evidence of the AS in Brazil and Portugal. To our knowledge, this is the first study assessing the psychometric properties of the AS in Brazil. Although the psychometric properties of the AS have been previously assessed in Portugal, this paper introduces a new version of the AS with a modification to one item (for more details, refer to the 'Measures' section). Moreover, conducting a study with samples from both Brazil and Portugal enables cross-cultural comparisons. The specific goals are: (a) to test the internal structure and reliability of the AS; (b) to test the invariance of the AS across culture, gender, age, education, occupation, and Covid-related concern and impact; (c) to assess the AS following an IRT approach; (d) to assess potential ceiling and floor effects; and (e) to identify convergent validity evidence based on the relation to presence of meaning. Moderate correlations to presence of meaning are expected. In addition, the magnitudes of the correlations to self-alienation are expected to be higher as this is the authenticity dimension mostly related to meaning in life.

Method

Participants and Procedures

Data collection took place as part of the project 'Future Time Orientation and Life Project: A theoretical and trans-

cultural approach from a psychosocial perspective.' The dataset used in this paper is a subset composed of 1,786 participants who responded to the AS in Brazilian or European Portuguese. Data collection was entirely online, from March to December 2020, via LimeSurvey platform. Incomplete answers ($n = 2$, 0.1%), foreigners ($n = 23$, 1.3%), and participants who selected the same response category for scales with reverse items ($n = 43$, 2.4%) were eliminated. Two participants exhibited unusual patterns of response (i.e., Mahalanobis distance per degree of freedom above 4.0) in multiple scales and were therefore eliminated. We examined participants whose responses in the AS had a Mahalanobis distance per degree of freedom above 2.5. Altogether, 19 participants were eliminated as they selected only extreme responses across all items. As the AS is anchored only in the extreme points, those participants might have not understood how to use the rating-scale.

Therefore, 1,699 participants were analyzed. The overall sample had an age range of 18 to 72 years, with $M = 31.1$ ($SD = 11.60$). In Brazil ($n = 1,077$), ages ranged from 18 to 72 years old, with $M = 32.3$ ($SD = 11.80$); while in Portugal ($n = 622$) ages ranged from 18 to 72 years, with $M = 29.0$ ($SD = 10.94$). As seen in Table 1, participants were predominantly Caucasian, female, workers, and had a college degree.

Table 1

Descriptive Statistics: Sociodemographic Features, and Covid-Related Concern and Impact

	Entire sample ($n = 1,699$)		Brazil ($n = 1,077$)		Portugal ($n = 622$)	
	n	%	n	%	n	%
Gender						
Women	1,123	66.1	701	65.1	422	67.8
Men	540	31.8	355	33.0	185	29.7
TGNC	20	1.2	14	1.3	6	1.0
No answer	16	0.9	7	0.6	9	1.4
Educational degree						
Some elementary school	2	0.1	2	0.2	0	0.0
Elementary school	42	2.5	16	1.5	26	4.2
High school	632	37.2	416	38.6	216	34.7
College degree	1,023	60.2	643	59.7	380	61.1
Occupation[1]						
Student	800	47.1	504	46.8	296	47.6
Grant holder	177	10.4	147	13.6	30	4.8
Worker	838	49.3	539	5.0	299	48.1
Unemployed	164	9.7	109	10.1	55	8.8
Retired	54	3.2	46	4.3	8	1.3
Other	89	5.2	73	6.8	16	2.6
Race/Ethnicity[1]						
Black	194	11.4	184	17.1	10	1.6
Brown[2]	39	2.3	39	3.6	0	0.0
Asian	11	0.6	8	0.7	3	0.5
Caucasian	1,199	70.6	649	60.3	550	88.4
Latin/Hispanic	288	17.0	220	20.4	68	10.9
Indian	2	0.1	2	0.2	0	0.0
Native American	44	2.6	44	4.1	0	0.0
Middle Eastern	9	0.5	6	0.6	3	0.5
Pacific Islander	1	0.1	1	0.1	0	0.0
Other	27	1.6	18	1.7	9	1.4
No answer	47	2.8	38	3.5	9	1.5

	Entire sample (*n* = 1,699)		Brazil (*n* = 1,077)		Portugal (*n* = 622)	
	n	%	*n*	%	*n*	%
Covid-related concern						
Not at all	44	2.6	18	1.7	26	4.2
Slightly	258	15.2	125	11.6	133	21.4
Moderately	466	27.4	224	20.8	241	38.7
Very much	517	30.4	323	30.0	194	31.2
Extremely	157	9.4	132	12.3	27	4.3
No answer	256	15.1	255	23.7	1	0.2
Covid-related impact						
Not at all	553	32.5	258	24.0	295	47.4
Slightly	401	23.6	232	21.5	169	27.2
Moderately	294	17.3	179	16.6	115	18.5
Very much	156	9.2	117	10.9	39	6.3
Extremely	39	2.3	36	3.3	3	0.5
No answer	256	15.1	255	23.7	1	0.2

Notes. TGNC = Transgender and gender non-conforming people, [1]The question allowed for multiple answers, [2]The category was not originally included in the study but was commonly reported by participants in the 'other' field.

Data collection occurred during the Covid-pandemic and two questions assessed (a) the level of concern regarding the pandemic; and (b) the extent to which the pandemic impacted the answers in the survey. As seen in Table 1, most participants expressed moderate to high levels of concern, but most indicated that the pandemic had little to no impact on their answers.

Measures

Authenticity Scale (Appendix 1)

The forms used in this study differ slightly from that adapted by Balbino et al. (2018). The Brazilian form was created with two items being slightly modified considering syntactic particularities in Brazil. Item 12 ("I feel alienated from myself") was modified in both forms to enhance comprehension. Taking inspiration from the Swedish form (Vainio & Daukantaitė, 2016), a statement that aligns better with daily language was created and is back-translated as "I feel like a stranger to myself."

Meaning in Life Questionnaire

The Meaning in Life Questionnaire (MLQ; Steger et al., 2006) measures presence of and search for meaning in life. However, in this study, only presence of meaning was used. It consists of a 5-item subscale responded to in a 7-point rating scale ranging from 'totally false' to 'totally true.' The MLQ has been adapted in Brazil (Damásio & Koller, 2015) and Portugal (Portugal, 2017). The internal structure was tested in our samples by CFA with maximum likelihood robust (MLR) estimator. The single-factor model yielded good fit in Brazil, $\chi^2 = 29.6(5)$, $p > .001$, CFI = .986, TLI = .972, RMSEA [90% C. I.] = .068 [.050; .087], SRMR = .019, and Portugal, $\chi^2 = 16.9(5)$, $p = .005$, CFI = .988, TLI = .977,

RMSEA [90% C. I.] = .062 [.036; .089], SRMR = .019.[2] Reliability analysis yielded excellent results both in Brazil, $a = .90$ and $\omega = .90$, and Portugal, $a = .91$ and $\omega = .91$.

Data Analysis

The AS was tested by CFA considering three models: the 3-factor solution with correlated factors, the unidimensional model, and the bifactor model. In accordance with Barrett-Lennard (1998), authenticity is a tripartite construct. Therefore, we anticipated that the 3-factor solution would provide the best fit. The bi-factor model was also tested to compare our findings with the Serbian version (Grijak, 2017), which supported the bi-factor model. Additionally, the unidimensional model was tested to further support that the 3-factor solution best fit the data. MLR estimator was chosen because data violated multivariate normality in Brazil, $M_{skewness} = 3,539.1$, $p < .001$, $M_{kurtosis} = 49.4$, $p < .001$, and Portugal, $M_{skewness} = 2,456.0$, $p < .001$, and $M_{kurtosis} = 41.9$, $p < .001$. We considered using weighted least squares mean and variance adjusted (WLSMV) or unweighted least squares mean and variance adjusted (ULSMV) estimators, as they are widely regarded as the most appropriate methods for ordinal data (Li, 2016; Rhemtulla et al., 2012). Conversely, as a few items lacked responses in some response categories, WLSMV and ULSMV were not allowed. Thus, choosing MLR stems from a Monte Carlo simulation study suggesting 7-point scales are appropriately tested by MLR (Rhemtulla et al., 2012). Goodness-of-fit indices cutoffs were based on Schreiber et al. (2006), who recommended comparative fit index (CFI) and Tucker-Lewis index (TLI) above or equal to .95, and standardized root mean square residual (SRMR) and root mean square error of approximation (RMSEA) below .06. The following cutoffs were considered acceptable: CFI and TLI above or equal to .90, and RMSEA and SRMR below .080 (Brown, 2006).

Multigroup CFA tested the invariance of the AS's factor structure (configural model), factor loadings (metric model), and item intercepts (scalar model) across groups. Consistent with previous studies, invariance models across gender, age, and occupation were tested. Transgender and gender non-conforming people were not considered for the models across gender because only a few participants belonged to this category. Invariance across age compared youths (up to 30 years old) and non-youths. Invariance across occupation compared workers and students, with participants both workers and students being ignored. Additionally, invariance models across education and culture were tested to ensure that individuals from diverse educational and cultural backgrounds respond to the AS in a similar psychometric pattern. Invariance across culture compared participants from Brazil and Portugal. Invariance across education compared participants with and without a college degree. Invariance models across gender, age, education, and occupation were tested in

---

[2] Cutoff values are discussed in the section Data Analysis.

each country separately. Moreover, invariance models across Covid-related concern and impact were tested to ensure the pandemic did not influence participants' response patterns. Invariance across Covid-related concern and impact considered the entire sample due to the reduced number of responses in some categories. The two first categories of Covid-related concern and the two last categories of Covid-related impact were collapsed. To establish invariance across groups, in addition to good fit, we expected no big differences between compared models (configural versus metric, and metric versus scalar), i.e., ΔRMSEA $\leq$ .050 and ΔCFI $\geq$ -.010 (Cheung & Rensvold, 2002).

The AS was tested by graded response model (GRM; Samejima, 1969), an IRT approach for polytomous data. The three first response categories for authentic living items were collapsed because the two first ones were answered by a reduced number of participants, $n < 19$. According to Linacre (1999), distributions with long tails of relatively infrequently used categories may bias item calibration. Hence, only four difficulty thresholds were estimated for authentic living. Separate GRM per factor were preferred over a multidimensional approach because tridimensional item characteristic curves (ICC) and test information curves (TIC) are hardly interpretable. Moreover, while response categories were collapsed for the authentic living subscale, the number of response categories varied across items. Hence, generalized partial credit model (GPCM) is better suited to the nature of the data, as it allows for the calibration of items with different numbers of response categories. However, GPCM is restricted to unidimensional models in mirt package (used in this study). Nevertheless, for the purpose of comparison, Appendix 2 shows the item parameters considering a multidimensional GRM, with no big differences being observed. In this model, response categories for authentic living were not collapsed, and self-alienation and accepting external influence items were reverse-coded to ensure all items were directly correlated. Lastly, GRM were preferred over Rasch models because they compute two item parameters (discrimination and difficulty) rather than one (difficulty). In addition, the fit indices of GRM were contrasted to Rasch models, with GRM showing the best performance (Appendix 3).

Statistical assumptions and fit indices were tested. Unidimensionality was tested by Loevinger's H coefficient, with values above .30 being expected (Sijtsma & Molenaar, 2002). Local dependence was tested by Q_3 test, with values below $|1/(L - 1)|$ (L meaning the length of the scale) being expected (Yen, 1993). Monotonicity was assessed by scalability coefficient H, with values above .30 being expected (Mokken, 1971). Item fit was tested by RMSEA (Cook et al., 2009). Person fit was tested by Zh statistics, with values below -3.0 suggesting potential aberrant response patterns (Paek & Cole, 2019). The reliability of latent trait was tested by *Rho* coefficient, with $\varrho < .70$ being expected (Sijtsma & Molenaar, 1987).

GRM parameters were then interpreted. Discrimination (a) informs the degree to which the responses are able to dis-

tinguish individuals with different latent trait levels (ө). GRM provides an index of discrimination that can be interpreted as follows: $a > 1.69$, very high; $a > 1.34$, high; $a < 0.64$, moderate; otherwise, low (Baker & Kim, 2017). GRM estimates $K - 1$ difficulty thresholds, where K represents the number of response categories. Item difficulty indicates the ө at which an individual is equally likely to endorse two adjacent response categories. Hence, b_1 designates the ө at which an individual is equally likely to respond to the first or second categories; and so on. General item difficulty (b) suggests the ө at which one has the same chance to respond to the first and last categories. ICC and TIC were plotted to assess the adequacy of the rating scale and the range of ө that the test assesses most effectively.

Reliability was tested by alpha (a), omega (ω), Spearman-Brown coefficient (r_{kk})[3], and average variance extracted (AVE)[4]. The following cutoffs were used to interpret a, ω, and r_{kk}: below .50, inacceptable; below 0.60, poor; below, 0.70, questionable; below 0.80, moderate; below 0.90, good; otherwise, excellent (Gliem & Gliem, 2003). AVE values above .50 were expected (Fornell & Larcker, 1981). The percentage of participants with the minimum and maximum punctuations were computed to assess ceiling or floor effects, with percentages over 15% indicating ceiling/floor effect (Terwee et al., 2007).

Associations between the AS and presence of meaning were assessed via Pearson correlations. Factor scores were computed using the maximum a posteriori method. Pearson correlation was selected because the values of skewness and kurtosis were between -1.0 and 1.0. The following cutoffs were used: $r < .30$, weak; $r < .50$ are moderate; otherwise, strong (Dancey & Reidy, 2007). Coefficients were compared with the AVE, with values lower than the root square of AVE being expected (Fornell & Larcker, 1981). R-to-Z transformations tested whether correlations between presence of meaning and self-alienation were stronger than correlations between presence of meaning and the other two factors.

The sample size was adequate for all analyses. For CFA, a sample size calculator (Soper, 2023) suggested a minimum of 100 participants for testing the model structure. Considering $a = .05$, $\beta = .20$, and the smallest subsample ($n = 159$), CFA were powerful to detect significant parameters with an effect of .264. All CFA models were implemented with and without outliers. The outliers were retained as no big differ-

---

[3] To calculate the Spearman-Brown coefficient, the subscales were randomly divided into two equal halves. This process was repeated, considering all possible combinations of item subsets. The final coefficient is based on all correlation magnitudes individually extracted.

[4] Although Fornell and Larcker (1981) defined AVE as an indicator of convergent validity, authors such as Peterson et a. (2020), Jorgensen et al. (2021), and Valentini and Damásio (2016), consider AVE an indicator of reliability. Their main argument is that AVE represents the portion of item variance that is not affected by residual variance. Thus, AVE assesses measurement errors. According to the Standards for Educational and Psychological Testing (American Psychological Association et al., 2014), measurement errors are more closely associated with reliability rather than validity.

ences were observed after their removal. For GRM, the sample size exceeded 500 (Nunes & Primi, 2005).

All analyses were conducted using R software 4.1.3 (R Core Team, 2023). The following packages were used: *lavaan* (version 0.6-9; Rosseel, 2012) for CFA; *mirt* (version 1.34, Chalmers, 2012) for GRM; *semTools* (version 0.5-5; Jorgensen et al., 2021) and *multicon* (version 1.6; Sherman, 2015) for reliability.

Ethical Statements

The study was evaluated and approved by ethical commissions from Brazil and Portugal. All participants provided their consent.

Results

The model with three correlated factors achieved excellent fit indices in Brazil, $\chi^2 = 151.9(51)$, $p < .001$, CFI = .973,

TLI = .965, RMSEA [90% C. I.] = .043 [.036; .050], SRMR = .038, and Portugal, $\chi^2 = 118.1(51)$, $p < .001$, CFI = .968, TLI = .958, RMSEA [90% C. I.] = .046 [.036; .056], SRMR = .041. The single-factor solution showed a poor fit in Brazil, $\chi^2 = 1,284.9(54)$, $p < .001$, CFI = .675, TLI = .602, RMSEA [90% C. I.] = .145 [.139; .152], SRMR = .113, and Portugal, $\chi^2 = 926.4(54)$, $p < .001$, CFI = .580, TLI = .487, RMSEA [90% C. I.] = .161 [.153; .170], SRMR = .115. The bifactor model reached the best fit indices in Brazil, 68.9(36), $p < .001$, CFI = .991, TLI = .984, RMSEA [90% C. I.] = .029 [.020; .038], SRMR = .021, and Portugal, $\chi^2 = 58.5(36)$, $p < .001$, CFI = .989, TLI = .980, RMSEA [90% C. I.] = .032 [.017; .045], SRMR = .025. However, in both samples, the covariance matrix was not positive definite, and several factor loadings were non-significant and below .50. Therefore, the 3-factor solution was retained. As seen in Figure 1, all factor loadings were above .500. Lastly, as seen in Table 2, the 3-factor solution demonstrated scalar invariance in all conditions.

Table 2
Multigroup CFA

Model	χ^2(df)	Sc. χ^2(df)	CFI	ΔCFI	TLI	RMSEA [90% C. I.]	ΔRMSEA	SRMR
Culture – Brazil, *n* = 1,077; Portugal, *n* = 622								
Configural	270.6(102)**		.971		.963	.044 [.038, .050]		.036
Metric	284.2(111)**	13.1(9)	.970	-.001	.965	.043 [.037, .048]	-.001	.039
Scalar	314.7(120)**	31.9(9)**	.967	-.003	.963	.044 [.038, .049]	.001	.040
Age (Brazil) – youths, *n* = 580; non-youths, *n* = 497								
Configural	200.1(102)**		.973		.965	.042 [.034, .050]		.041
Metric	209.9(111)**	9.3(9)	.973	.000	.968	.041 [.033, .048]	-.001	.042
Scalar	235.0(120)**	27.0(9)*	.968	-.005	.965	.042 [.035, .049]	.001	.045
Age (Portugal) – youths, *n* = 449; non-youths, *n* = 173								
Configural	193.8(102)**		.958		.945	.054 [.043, .065]		.050
Metric	196.1(111)**	3.1(9)	.961	.003	.953	.050 [.039, .060]	-.004	.050
Scalar	221.3(120)*	26.8(9)*	.953	-.008	.949	.052 [.042, .062]	.002	.051
Gender (Brazil) – females, *n* = 701; males, *n* = 355								
Configural	250.6(102)**		.961		.950	.053 [.045, .060]		.043
Metric	359.5(111)**	8.1(9)	.961	.000	.954	.050 [.043, .058]	-.003	.044
Scalar	274.8(120)*	14.2(9)	.960	-.001	.956	.049 [.042, .057]	-.001	.045
Gender (Portugal) – females, *n* = 422; males, *n* = 185								
Configural	170.1(102)**		.968		.958	.047 [.035, .058]		.044
Metric	175.3(111)**	6.1(9)	.969	.001	.964	.044 [.032, .055]	-.003	.047
Scalar	191.3(120)**	16.4(9)	.966	-.003	.963	.044 [.033, .055]	.000	.049
Education (Brazil) – college degree, *n* = 643; no college degree, *n* = 434								
Configural	220.8(102)**		.969		.960	.047 [.039, .054]		.041
Metric	227.3(111)**	5.8(9)	.970	.001	.964	.044 [.037, .052]	-.003	.041
Scalar	242.0(120)**	14.1(9)	.968	-.002	.965	.043 [.036, .051]	-.001	.042
Education (Portugal) – college degree, *n* = 380; no college degree, *n* = 242								
Configural	216.9(102)**		.949		.934	.060 [.050, .071]		.049
Metric	223.5(111)**	7.8(9)	.950	.001	.940	.057 [.047, .067]	-.003	.052
Scalar	236.8(120)**	12.6(9)	.948	-.002	.943	.056 [.046, .066]	-.001	.054
Occupation (Brazil) – workers, *n* = 573; students, *n* = 373								
Configural	202.2(102)**		.970		.962	.046 [.037, .054]		.041
Metric	209.3(111)**	7.3(9)	.971	.001	.965	.043 [.035, .051]	-.003	.043
Scalar	228.3(120)**	19.3(9)*	.968	-.003	.965	.044 [.036, .052]	.001	.045
Occupation (Portugal) – workers, *n* = 643; students, *n* = 434								
Configural	159.6(102)**		.970		.962	.044 [.031, .056]		.045
Metric	163.9(111)**	5.0(9)	.973	.003	.968	.041 [.028, .052]	-.003	.047
Scalar	178.8(120)**	15.2(9)	.970	-.003	.967	.041 [.029, .052]	.000	.048
Covid-concern – not at all/slightly, *n* = 302; moderately, *n* = 465; very much, *n* = 517, extremely, *n* = 159								

Model	χ^2(df)	Sc. χ^2(df)	CFI	ΔCFI	TLI	RMSEA [90% C. I.]	ΔRMSEA	SRMR
Configural	353.7(204)**		.971		.962	.045 [.038, .052]		.041
Metric	385.5(311)**	31.2(27)	.970	-.001	.966	.043 [.036, .050]	-.002	.048
Scalar	416.2(258)**	28.5(27)	.969	-.001	.969	.041 [.034, .048]	-.002	.049
Covid-impact – not at all, *n* = 553; slightly, *n* = 401; moderately, *n* = 294; very much/extremely, *n* = 195								
Configural	366.6(204)**		.968		.959	.047 [.040, .054]		.042
Metric	397.2(231)**	29.7(27)	.968	.000	.963	.045 [.038, .051]	-.002	.049
Scalar	436.2(258)**	37.9(27)	.965	-.003	.965	.044 [.037, .050]	-.001	.050

Notes. *p > .05. **p > .001.

Figure 1
CFA
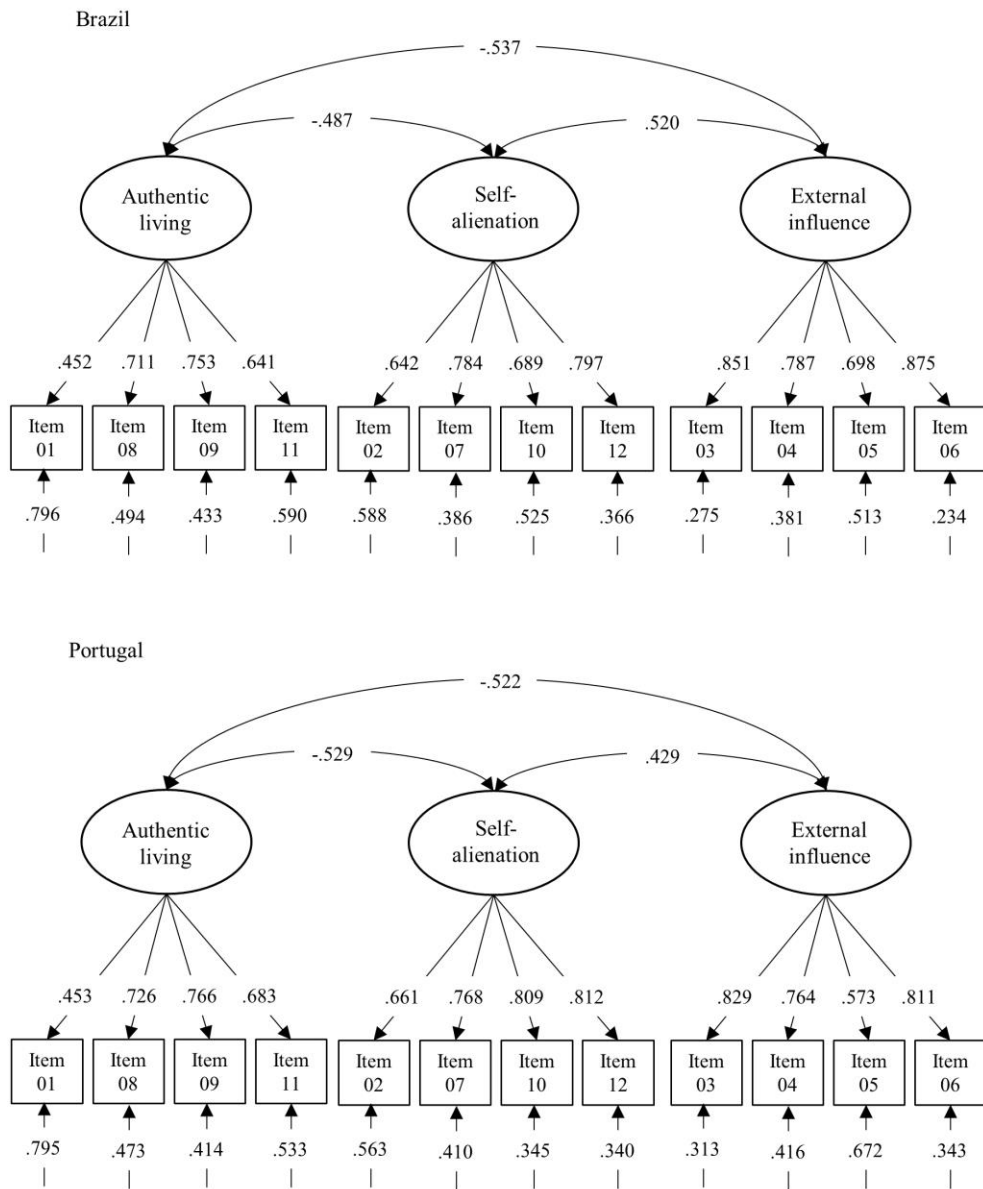
Notes. Top figure represents the Brazilian sample (*n* = 1,077), and bottom figure shows results for the Portuguese sample (*n* = 622).

As seen in Appendix 4, GRM assumptions were partially observed, whereas item fit and person fit were good. Loevinger's and scalability *H* coefficients were all above .30, which suggest unidimensionality and monotonicity. One

item pair of authentic living and self-alienation, and two item pairs of accepting external influence exceeded the expected cutoff, indicating local dependence and potential biases in the GRM calibration. *RMSEA* values were all below .060,

suggesting good item fit. As for person fit, less than 5% of respondents had Zh values below -3.0. All ϱ coefficients were above .70.

Table 3

GRM – item parameters

| Factors/Items (in English)	Logistic parameters							
	a	b	b_1	b_2	b_3	b_4	b_5	b_6
Authentic living								
1. I think it is better to be yourself, than to be popular	1.30	-1.72	---	---	-2.96	-2.30	-1.30	-0.17
8. I always stand by what I believe in	2.52	-0.88	---	---	-1.90	-1.26	-0.55	0.41
9. I am true to myself in most situations	2.41	-0.93	---	---	-2.01	-1.32	-0.59	0.37
11. I live in accordance with my values and beliefs	2.08	-1.15	---	---	-2.29	-1.62	-0.76	0.25
Self-alienation								
2. I don't know how I really feel inside	1.82	0.45	-1.20	-0.27	0.17	0.69	1.34	1.86
7. I feel as if I don't know myself very well	2.61	0.80	-0.62	0.21	0.62	0.99	1.44	2.03
10. I feel out of touch with the 'real me'	2.50	0.75	-0.65	0.16	0.54	0.96	1.37	1.95
12. I feel alienated from myself	3.62	1.14	0.01	0.58	0.94	1.33	1.74	2.10
Accepting external influence								
3. I am strongly influenced by the opinions of others	3.32	0.78	-0.79	-0.03	0.48	1.06	1.66	2.17
4. I usually do what other people tell me to do	2.60	0.95	-0.71	0.05	0.59	1.26	1.96	2.63
5. I always feel I need to do what others expect me to do	1.77	0.50	-1.10	-0.35	0.21	0.73	1.45	2.17
6. Other people influence me greatly	3.58	0.72	-0.91	-0.11	0.44	0.98	1.60	2.19

Notes. n = 1,699.

As seen in Table 3, all items exhibited very high discrimination, except for item 1. Authentic living items were less difficult compared to self-alienation and accepting external influence items. That is, higher rates of the latent trait level were needed to endorse the highest response categories for self-alienation and accepting external influence items. Conversely, in the case of authentic living items, the highest response categories were endorsed by participants with fewer rates in the latent trait level. As seen in Figure 2, authentic living subscale was most effective in assessing participants with ɵ between -3.0 and 2.0, whereas self-alienation and accepting external influence subscales demonstrated best performance among respondents with ɵ from -2.0 to 3.0. Furthermore, as seen in Figure 3, the 5-point rating scale with the three first categories merged did not fit item 1, from authentic living subscale. Participants did not discriminate the four first categories. In self-alienation subscale, the 7-point

scale did not fit the response patterns of items 2 and 10, with no differentiation between the second and third categories. In addition, participants did not distinguish the sixth and seventh categories of item 2. The 7-point scale fitted the response patterns of all accepting external influence items.

Table 4 displays the results of ceiling/floor effect and reliability. The mean of authentic living was much higher compared to self-alienation and accepting external influence. Over 15% of participants from the two samples reached the maximum punctuation in authentic living, meaning ceiling effect. Self-alienation and accepting external influence demonstrated good reliability. In authentic living subscale, a, ω, and r_{kk} reached moderate values, and AVE was below .50. With the exclusion of item 1, the reliability of authentic living would have yielded better results in Brazil, a = .75, ω = .75, r_{kk} = .73, and AVE = .50, and Portugal a = 0.77, ω = .77, r_{kk} = .75, and AVE = .53.
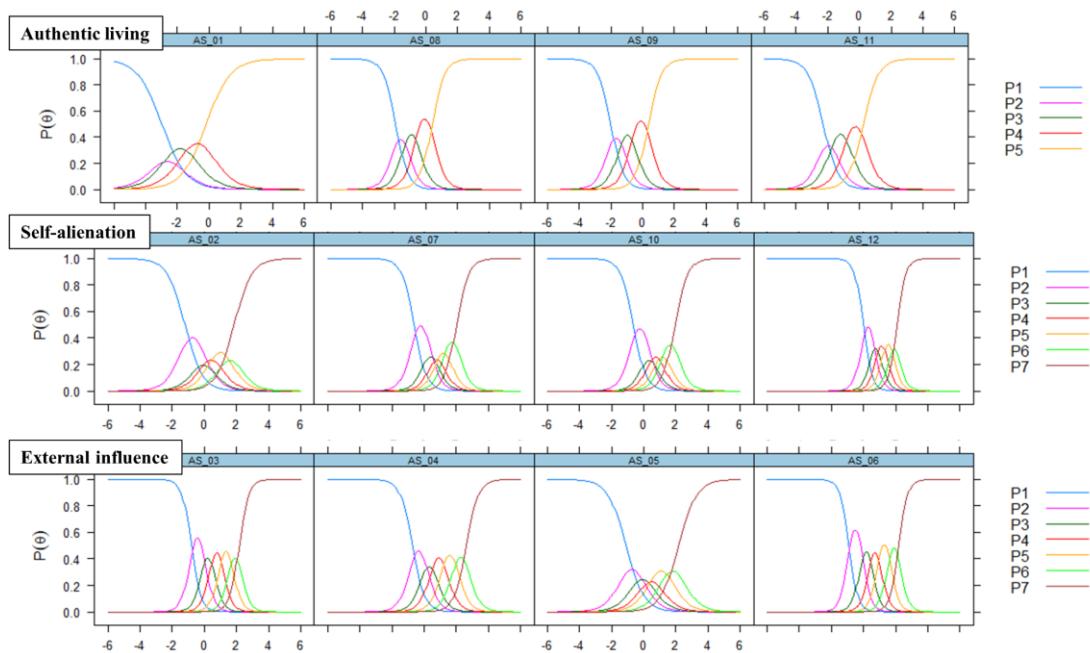
Figure 2
Test Information Curves

Authentic living

Self-alienation

External Influence
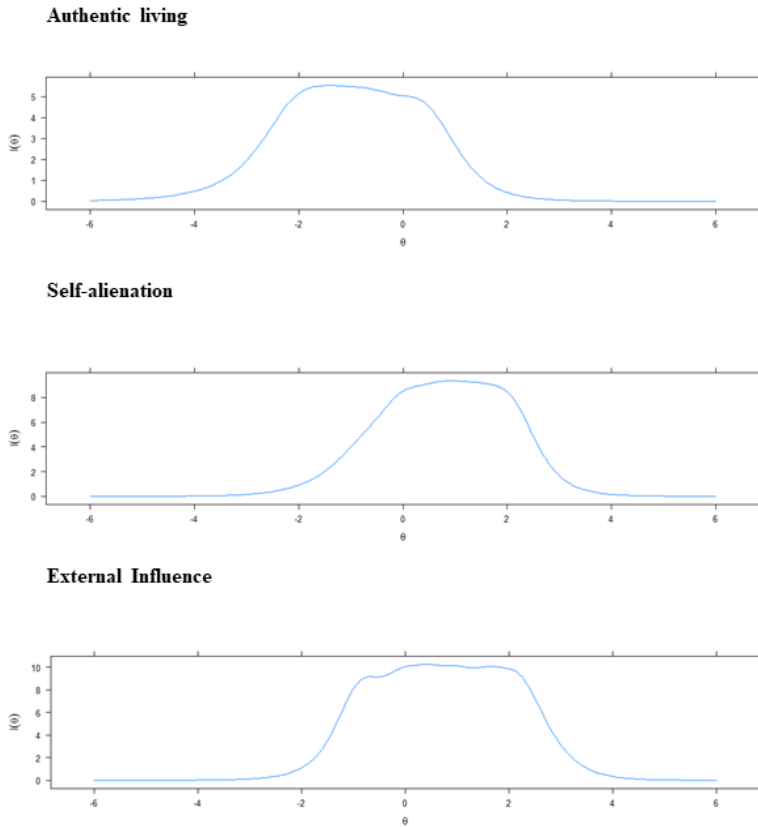
Note. n = 1,699.

Figure 3
Item Characteristic Curves

Notes. n = 1,699.

Table 4
Descriptive statistics, reliability, and correlations

	M (SD)	Min. punct.	Max. punct.	Reliability Coefficients					Correlations		
				α	ω	n_{kk}	AVE	$\sqrt{AVE}$	Ali	Ext	MLQ
Brazil, *n* = 1,077											
Aut	5.9 (0.96)	0%	18.6%	.73	.74	.73	.42	.65	-.579**	-.622**	.420**
Ali	2.9 (1.48)	11.6%	1.3%	.81	.81	.82	.52	.72		.586**	-.524**
Ext	3.0 (1.46)	11.5%	0.6%	.87	.87	.87	.64	.80			-.337**
Portugal, *n* = 622											
Aut	6.0 (0.82)	0.2%	18.2%	.75	.76	.76	.46	.68	-.614**	-.606**	.397**
Ali	2.8 (1.41)	11.3%	0.5%	.84	.84	.85	.58	.76		.489**	-.546**
Ext	3.0 (1.24)	6.1%	0.3%	.83	.83	.83	.55	.74			-.248**

Notes. **p* < .05, ***p* < .001, Aut = authentic living, Ali = self-alienation, Ext = accepting external influence, Min. punct. = minimum punctuation, Max. punct. = maximum punctuation, MLQ = presence of meaning subscale.

Table 4 displays the correlation results. As expected, self-alienation and accepting external influence were positively correlated with each other, as well as inversely correlated to authentic living. The correlation magnitudes were moderate and below the root square of the AVE. As expected, MLQ was positively correlated to authentic living and inversely correlated to self-alienation and accepting external influence. The magnitudes of correlations were moderate, except for accepting external influence in Portugal, which showed a weak correlation. R-to-Z transformations suggested that correlations to self-alienation were stronger compared to authentic living ($Z = 4.357$, $p < .001$, in Brazil; $Z = 4.947$, $p < .001$, in Portugal) and accepting external influence ($Z = 7.697$, $p < .001$, in Brazil; $Z = 8.319$, $p < .001$, in Portugal).

Discussion

This study introduces distinct types of validity and reliability evidence of the AS in a sample from Brazil and Portugal. Based on CFA, the 3-factor solution assessing self-alienation, authentic living, and accepting external influence was selected. The previous adaptation in Portugal (Balbino et al., 2018) also concluded for a 3-factor solution. To our knowledge, this is the first study testing the internal structure of the AS in Brazil. Unlike the original internal structure, a second-order factor was not extracted because CFA does not compute the fit indices of models with zero degrees of freedom (Kline, 2011). Two alternative models were tested: the unidimensional model and the bi-factor model. The unidimensional model was rejected due to its poor fit to the data. In turn, despite the bi-factor model showing the best fit indices, the covariance matrix in both samples was not positive definite, and several factor loadings were non-significant and below .50. Although the bi-factor model was retained in Serbia (Grijak, 2017), the factor loadings were not reported, limiting the drawing of conclusions.

Multigroup CFA concluded for scalar invariance across culture, gender, age, education, occupation, and Covid-related concern and impact. The results are in line with previous studies that tested the invariance of the AS across gender (Nartova-Bochaver et al., 2021, Wood et al., 2008; Xia et al., 2022; Zoysa et al., 2021). In China (Xia et al., 2022), metric invariance across occupation was observed yet the scalar

model was not tested. Metric invariance across age was observed in China (Xia et al., 2022) and Russia (Nartova-Bochaver et al., 2021), though in the latter the scalar model reached good fit only after freeing three item intercepts. Thus, there might be aspects in Russian culture that influence intercepts across age, while item intercepts are equivalent across age groups in Brazil and Portugal. To our knowledge, this is the first study testing invariance across culture, education, and Covid-related concern and impact.

So far, it seems this is the first study discussing the AS item parameters following an IRT approach. Discrimination was very high for nearly all items, meaning the AS is able to distinguish people with different trait levels. Items' difficulty and TIC suggested the AS is most effective in assessing participants with medium to high rates of self-alienation and accepting external influence, as well as participants with medium to low rates of authentic living. Therefore, the AS demonstrates low reliability among participants with high authenticity traits. Lastly, ICC suggested the inadequacy of the rating-scale. This might be attributed to the lack of anchoring for the intermediate points on the rating-scale. Indeed, this was the reason for the elimination of 19 participants who potentially had difficulty understanding how to use the rating-scale.

The results suggest a ceiling effect for the authentic living subscale. In both samples, the mean of direct scores was very high and the percentage of participants with the maximum punctuation was above 15%. Moreover, a reduced number of participants selected the two first response categories for authentic living items, prompting the collapse of the three first response categories prior to GRM calibration. Hence, the ceiling effect might have exacerbated the bad performance in the assessment of people with high authentic living traits.

This study reported four types of reliability coefficients. Self-alienation and accepting external influence demonstrated good reliability, while authentic living achieved moderate reliability and an AVE below .50. The existing literature mostly reported alpha, yielding values similar to those obtained in this study. In Russia (Nartova-Bochaver et al., 2021), omega was assessed and achieved similar results to the present study. Lastly, Spearman-Brown coefficient was assessed only in Iran, considering the whole set of items. To

our knowledge, this is the first study assessing the AVE of AS factors, enabling the identification of discriminant validity by comparing the correlation coefficients to the square root of AVE.

The study identified associations between authenticity and presence of meaning. The results are in line with the literature on the relationship between authenticity and well-being (e.g., Di Fabio, 2014; Grijak, 2017), including studies with the MLQ (Akin & Taş, 2015). This study advances the literature by comparing the correlation coefficients between factors. R-to-Z transformation revealed that correlations to self-alienation were stronger than correlations to authentic living and accepting external influence. This corroborates our hypothesis that self-alienation would have higher correlation coefficients. This is the authenticity dimension mostly related to knowledge about the self and, consequently, meaning in life.

Finally, multiple findings indicated that item 1 has a dubious performance. In the CFA, item 1 yielded a factor loading below .50. In GRM, item 1 was the only one that did not exhibit very high discrimination. In addition, ICC indicated that participants did not distinguish between the four first response categories, potentially exacerbating ceiling effect. Lastly, the elimination of item 1 would have improved the reliability of authentic living in both samples. The bad performance of item 1 has also been observed in Turkey (İlhan & Özdemir, 2013), Canada (Grégoire et al., 2014), and Sri Lanka (Zoysa et al., 2021). In Russia (Nartova-Bochaver et al., 2021), item 1 migrated to accepting external influence subscale. Item 1 assesses the respondent's inclination to prioritize alignment with their values over popularity. The bad performance might be attributed to variations in individuals' valuing of popularity. In other words, individuals may value not being popular yet still accept and introject others' expectations.

Limitations and Future Directions

This study has some limitations. First, the data collection was online with a convenience sample. Future studies employing representative samples in more controlled environments may assess the AS under less biased conditions. Second, participants' responses were maldistributed, especially in authentic living subscale. The maldistribution of the data might have affected the quality of data analysis. For instance, the lack of responses in certain categories precluded the use of ordinal estimators. Third, violations of local dependency might have biased the GRM calibration. Lastly, given that the AS has only three factors, the fit indices for the original

model with a second-order could not be computed. Therefore, we cannot ensure that the model without a second-order factor is actually better than the original model with a second-order factor.

Based on the results of this study, future versions of the AS may be constructed. First, alternative rating-scales with fewer response categories and with all points properly anchored may be used. Second, items may be rephrased or created to properly distinguish individuals with high authenticity traits. The new items should be more difficult, in the case of authentic living, and less difficult, in the case of self-alienation and accepting external influence. Third, because of its bad performance, item 1 should be removed or rephrased.

Conclusion

This study introduces distinct types of validity and reliability evidence of the AS in Brazil and Portugal, based on CFA, invariance models, GRM, four types of reliability coefficients, and relations to presence of meaning. The findings indicate that the internal structure of the AS has three factors assessing self-alienation, authentic living, and accepting external influence, with the three subscales showing moderate to good reliability. The internal structure, factor loadings, and item intercepts are invariant across different groups, and associations with presence of meaning showed additional validity evidence. All items properly assess authenticity, except for item 1, which showed a few unsatisfactory results. Despite the good evidence, the rating-scale is inappropriate for some items, especially for authentic living subscale, which is affected by ceiling effect. Moreover, the three subscales are not able to distinguish individuals with high authenticity traits. Regardless of limitations, the study suggests that the AS may be effectively employed to assess authenticity, especially if refined versions are created to overcome the limitations acknowledged in this study.

References

Akin, A., & Taş, İ. (2015). Yaşam anlami ölçeği: Geçerlik ve güvenirlik çalişmasi [Meaning in life questionnaire: A study of validity and reliability]. *Turkish Studies, 10*(3), 27-36. https://doi.org/10.7827/TurkishStudies.7860

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Baker, F. B., & Kim, S. H. (2017). The basics of item response theory using R. *Springer International Publishing.* https://doi.org/10.1007/978-3-319-54205-8

Balbino, I. F., Galinha, I. C., Morais, C. C., & Calado, S. S. (2018). Contributo para a validação da versão portuguesa da Escala de Autenticidade [Contribution to the validation of the Portuguese version of the Authenticity Scale]. *Psicologia, Saúde & Doenças*, *19*(3), 564–577. https://doi.org/10.15309/18psd190308

Barrett-Lennard, G. T. (1998). *Carl Rogers' helping system: Journey and substance.* Sage.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research.* The Guilford Press.

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1-29. https://doi.org/10.18637/jss.v048.i06

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Cook, K. F., Kallen, M. A., & Amtmann, D. (2009). Having a fit: Impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Quality of Life Research, 18*(4), 447-460. https://doi.org/10.1007/s11136-009-9464-4

Damásio, B. F., & Koller, S. H. (2015). Meaning in Life Questionnaire: Adaptation process and psychometric properties of the Brazilian version. *Revista Latinoamericana de Psicología*, *47*(3), 185-195. https://doi.org/10.1016/j.rlp.2015.06.004

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology.* Pearson Education.

Di Fabio, A. (2014). Authenticity Scale: Un primo contributo alla validazione della versione italiana [Authenticity Scale: A first contribution to validation of the Italian version]. *Counseling: Giornale Italiano di Ricerca e Applicazioni, 7*(2), 231-238.

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

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50. https://doi.org/10.2307/3151312

Gable, S. L., & Haidt, J. (2005). What (and why) is positive psychology? *Review of General Psychology, 9*, 103-110. https://doi.org/10.1037/1089-2680.9.2.103

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *Proceedings of the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, USA.*

Grégoire, S., Baron, L., Ménard, J., & Lachance, L. (2014). The Authenticity Scale: Psychometric properties of a French translation and exploration of its relationships with personality and well-being. *Canadian Journal of Behavioural Science, 46*(3), 346-355. https://doi.org/10.1037/a0030962

Grijak, Đ. (2017). Psychometric evaluation of the authenticity scale on the sample of students in Serbia. *Psihologija, 50*(1), 85-99. https://doi.org/10.2298/PSI160504001G

İlhan, T., & Özdemir, Y. (2013). Otantiklik Ölçeğinin Türkçe'ye uyarlanması: Geçerlik ve güvenirlik çalışması [Adaptation of Authenticity Scale to turkish: A validity and reliability study]. *Turkish Psychological Counseling and Guidance Journal, 4*(40), 142-153.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). *semTools: Useful tools for structural equation modeling.* https://CRAN.R-project.org/package=semTools

Kernis, M. H., & Goldman, B. M. (2005). From thought and experience to behavior and interpersonal relationships: A multicomponent conceptualization of authenticity. In A. Tesser, J. V. Wood, & D. Stapel (Eds.), *On building, defending, and regulating the self: A psychological perspective* (pp. 31-52). Psychology Press.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). The Guilford Press.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavioral Research Methods, 48*(3), 936-949. https://doi.org/10.3758/s13428-015-0619-7

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 3*(2), 103-122.

Mokken, R. J. (1971). *Theory and procedure of scale analysis: With applications in political research.* The Hague Mouton.

Nartova-Bochaver, S., Reznichenko, S., & Maltby, J. (2021). The Authenticity Scale: validation in Russian culture. *Frontiers in Psychology, 11*, 609617. https://doi.org/10.3389/fpsyg.2020.609617

Nunes, C. H. S. D. S., & Primi, R. (2005). Impact of the sample size in the item and subject's parameters estimates under item response theory. *Avaliação Psicológica, 4*(2), 141-153.

Paek, I., & Cole, K. (2019). *Using R for item response theory model applications.* Routledge.

Peterson, R. A., Kim, Y., & Choi, B. (2020). A meta-analysis of construct reliability indices and measurement model fit metrics. *Methodology, 16*(3), 208-223. https://doi.org/10.5964/meth.2797

Portugal, M. V. (2017). *Versão portuguesa do Questionário do Sentido da Vida: Primeiros estudos psicométricos* [Portuguese version of the Meaning in Life Questionnaire]. Non-published master thesis, Universidade de Lisboa. https://repositorio.ul.pt/bitstream/10451/33211/1/ulfpie052851_tm.pdf

R Core Team. (2023). R: A language and environment for statistical computing (version 4.1.3) [Computer software]. *R Foundation for Statistical Computing.* https://www.R-project.org/

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354-373. https://doi.org/10.1037/a0029315

Rogers, C. R. (1961). *On becoming a person: A therapist's view of psychotherapy.* Constable.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, *34*(Suppl 1), 1-97. https://doi.org/10.1007/BF03372160

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323-338. https://doi.org/10.3200/JOER.99.6.323-338

Shamsi, A., Ghamarani, A., Samadi, M., & Ahmadzadeh, M. (2012). The study of the validity and reliability of the Authentic Personality Scale. *Psychological Methods and Models, 2*(8), 89-100.

Sherman, R. A. (2015). *Multicon: An R Package for the Analysis of Multivariate Contructs, R package Version 1.6.* https://cran.r-project.org/web/packages/multicon/multicon.pdf

Sijtsma, K., & Molenaar, I. W. (1987). Reliability of test scores in nonparametric item response theory. *Psychometrika, 52*(1), 79–97. https://doi.org/10.1007/BF02293957

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory.* SAGE Publications.

Soper, D.S. (2023). *A-priori sample size calculator for structural equation models* [software]. Available from https://www.danielsoper.com/statcalc

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: assessing the presence of and search for meaning in life. *Journal of Counseling Psychology, 53*(1), 80-93. https://doi.org/10.1037/0022-0167.53.1.80

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., ... & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34-42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Vainio, M. M., & Daukantaitė, D. (2016). Grit and different aspects of well-being: Direct and indirect relationships via sense of coherence and authenticity. *Journal of Happiness Studies, 17*(5), 2119-2147. https://doi.org/10.1007/s10902-015-9688-7

Valentini, F., & Damásio, B. F. (2016). Variância Média Extraída e Confiabilidade Composta: Indicadores de precisão [Average Variance

Extracted and Composite Reliability: Reliability coefficients]. *Psicologia: Teoria e Pesquisa, 32*(2), 1-7. https://doi.org/10.1590/0102-3772e322225

Winnicott, D. W. (1965). *The maturational processes and the facilitating environment*. International Universities Press.

Wood, A. M., Linley, P. A., Maltby, J., Baliousis, M., & Joseph, S. (2008). The authentic personality: A theoretical and empirical conceptualization and the development of the Authenticity Scale. *Journal of Counseling Psychology, 55*(3), 385-399. https://doi.org/10.1037/0022-0167.55.3.385

Xia, M., Lv, H., & Xu, X. (2022). Validating the Chinese version authenticity scale: Psychometrics in college and community samples. *Current Psychology*, 41, 7301–7313. https://doi.org/10.1007/s12144-020-01326-7

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*, 187-213. https://doi.org/10.1111/j.1745-3984.1993.tb00423.x

Zlyvkov, V. L., Lukomska, S. O., Kotukh, O. V., Dykhovichnyi, O. O., & Kruglova, N. V. (2019). Authenticity of the english language teacher's: the validation of authenticity questionnaire using item response theory. *Science progress in European countries: new concepts and modern solutions*, 335.

Zoysa, P., Kumar, S., Amarasuriya, S. D., & Mendis, N. S. (2021). Being yourself: An assessment of authenticity in undergraduates of a University in Sri Lanka. *Asia Pacific Journal of Counselling and Psychotherapy, 12*(2), 138-153. https://doi.org/10.1080/21507686.2021.1924810