TPM®

THE RELATIONSHIP ASSESSMENT SCALE REVISED (RAS-R): ADAPTATION AND PSYCHOMETRIC EVALUATION BY CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY

KENNEDY L. QUISPILAYA-CAPCHA
UNIVERSIDAD PERUANA CAYETANO HEREDIA, LIMA, PERU

DANIEL E. YUPANQUI-LORENZO
UNIVERSIDAD PRIVADA DEL NORTE, LIMA, PERU

GABRIELA MONTES-HERRERA
UNIVERSIDAD PERUANA CAYETANO HEREDIA, LIMA, PERU

FLOR M. ROMERO-ALVAREZ
UNIVERSIDAD NACIONAL FEDERICO VILLARREAL, LIMA, PERU

TANIA ARAUCO-LOZADA
UNIVERSIDAD DE CIENCIAS Y HUMANIDADES, LIMA, PERU

The study introduces the revised Relationship Assessment Scale (RAS-R) in response to discrepancies identified within existing RAS models. The first phase involved grammatical revisions, bias removal, and standardization of response options for the Spanish version of the RAS. The subsequent phase evaluated the psychometric properties of both the RAS and the RAS-R, as well as the specification of common models from the literature, using data from 547 participants for the RAS and 285 for the RAS-R. In contrast to the RAS, which required adjustments due to correlated error or item elimination, the RAS-R demonstrated optimal fit without requiring model respecification. Further analysis using the Graded Response Model (GRM) indicated a better fit for the RAS-R, with consistent item discrimination across items, particularly for Item 2. Evaluation of test and item curves revealed a wide range of responses, highlighting the reliability of the instrument. In conclusion, the RAS-R was found to be a robust instrument that addresses the factorial structure challenges encountered in the RAS literature.

Keywords: Relationship satisfaction; Love; Psychometric properties; Item Response Theory; Adaptation.

Correspondence concerning this article should be addressed to Daniel E. Yupanqui-Lorenzo, Universidad Privada del Norte, Av. Tingo Maria 1122, Lima 15083, Peru. Email: daniel.yupanqui@upn.pe

As individuals mature and progress through life, they encounter various forms of attraction — physical, emotional, and sexual — toward others. These attractions form the foundation of romantic relationships, some of which can emerge during adolescence, a critical period for acquiring skills essential for fostering intimate connections (Rivera et al., 2011). Evaluating the contentment within a romantic relationship becomes imperative for its sustenance. In a cross-national study spanning 32 countries, findings revealed that

a substantial 76% of participants expressed satisfaction with feeling loved, 84% were satisfied with their current partners, and 63% reported satisfaction with their sexual lives as couples. This sense of satisfaction demonstrated significant correlations with higher educational levels, elevated income brackets, and marital status (Institut de Publique Sondage d'Opinion Secteur, 2023).

Examining this scenario specifically within Peru, more than 70% satisfaction was found with feeling loved, being in their relationship, and sex life (Institut de Publique Sondage d'Opinion Secteur, 2023). According to the Instituto Nacional de Estadística e Informática [National Institute of Statistics and Informatics] (2023), in 2022, 77,537 marriages were registered. At the same time, it was reported that there is a higher proportion of cohabiting couples (24.2%) than married couples (23.5%), with divorced couples making up a small percentage (0.7%). Likewise, in recent years, the rate of married people has decreased, especially in urban areas, from 28.6% to 25.7% (Instituto Nacional de Estadística e Informática, 2020). Conversely, the incidence of divorce displayed an upward trajectory from 2011 (N = 5,625) to 2019 (N = 16,485), with a sudden dip in 2020 (N = 5,682), possibly attributable to the pandemic (Statista, 2022). By 2021, 14,708 divorces were reported in Peru (Instituto Nacional de Estadística e Informática, 2023). Furthermore, in 2022, there was a 31.6% increase in divorce cases compared to the previous year, with a total of 5,184 divorces in the first half of the year (Superintendencia Nacional de los Registros Públicos, 2022).

Generally, relationships are expected to last over time, but more often, they have a shorter duration. Relationship stability is influenced by a range of factors, and one determinant is the level of satisfaction experienced within the relationship. Relationship satisfaction is a person's feelings, thoughts, and behaviors regarding their romantic relationship (Hendrick et al., 1998). It also pertains to the extent of intimacy, affection, and support that partners exchange, culminating in an overall subjective evaluation of the relationship's status (Collins et al., 2009; Graham et al., 2011). Moreover, this assessment is linked to an individual's sense of well-being and their positive view of themselves and their partner. Thus, those who find satisfaction in their relationship tend to experience a sense of well-being, tranquility, and equilibrium, as their emotional and relational needs are met (Sánchez et al., 2020). In that sense, it is important to assess relationship satisfaction, as it is based on criteria that evaluate the success of the relationship, its duration, and the demands of couples' needs (Özdemir & Demir, 2019). Research consistently shows that couples who report satisfaction in their relationship exhibit better psychological well-being compared to those who do not (Kanter & Proulx, 2021). On the flip side, dissatisfaction raises the risk of issues like family conflict, alcohol and substance dependence, problems in parent-child relationships, and dissatisfaction with work. These negative outcomes significantly impact individuals' overall quality of life and well-being (Sánchez et al., 2020).

Therefore, it is essential to have tools to measure satisfaction in interpersonal relationships. The Relationship Assessment Scale (RAS; Hendrick, 1988) has been validated in several international contexts and has been used in several studies focusing on the dynamics of couple relationships. As a brief measure, it allows for quick assessments, mitigating fatigue-related biases and item similarity. Its focus on the assessment of relationship satisfaction facilitates the understanding of relationship dynamics and serves as an initial starting point for therapeutic interventions. Therefore, a review of the literature on psychometric studies conducted to validate this instrument was undertaken to assess its validity based on previous findings. A number of studies were identified, ranging from 1991 to the most recent research in 2023 (see Table 1).

Between 1988 and 2016, the scale was mostly unifactorial, except for Rask et al.'s (2010) study, which identified a 2-factor structure. This was the first report to diverge from the initial proposed model. Until 2016, the RAS was predominantly considered unidimensional despite its multiple translations into different cultures. However, Sychev (2016) and later Maroufizadeh et al. (2018) demonstrated divergences in error correlations that

frequently occurred in subsequent studies. The literature consistently shows an association of errors between Items 1 and 2, as well as between Items 4 and 7. Both specifications were established in studies by González-Rivera (2020) and Ventura-León et al. (2023). However, in the latter study, correlating item errors improved model fit but led to a decrease in reliability (ω = .56). Therefore, Items 4 and 7 were eliminated in the final structure, resulting in a 5-item model (Ventura-León et al., 2023). It is worth noting that these modifications of the RAS are not specific to one language; they were present in the Swedish, Persian, Arabic, and Spanish translations.

TABLE 1
Literature review on the psychometric analysis of the RAS in different countries

Year	Authors	Country	Items	Models	Factorial modifications	Reliability
1991	Hassebrauck	Germany	7	One factor	None	.87[a]
1998	Hendrick et al.	USA	7	One factor	None	.85[a]
1999	Vaughn & Baier	USA	7	One factor	None	.91[a]
2005	Dinkel & Balck	Germany	7	One factor	None	.89[a]
2010	Rask et al.	Sweden	7	Two factors	Factor 1: ras1, ras2, ras5, ras7 Factor 2: ras4, ras6	.84[a] .89[a]
2011	Cassepp-Borges & Pasquali	Brasil	7	One factor	None	.85[a]
2014	Çelik	Turkey	7	One factor	None	.87[a]
2014	Hernandez	Brasil	7	One factor	None	.87
2014	Martos et al.	Ungarn	7	One factor	None	.93[a]
2015	De la Rubia	Mexico	7	One factor	None	.81[a]
2016	Chowdhury et al.	Bangladesh	7	One factor	None	.86[a]
2016	Dehshiri & Mousavi	Iran	7	One factor	None	.88[a]
2016	Sychev	Russia	7	One factor	ras4~~ras7	.86[a]
2018	Maroufizadeh et al.	Iran	7	One factor	ras4 ~~ ras7	.83[a]
2020	González-Rivera	Puerto Rico	7	One factor	ras1~~ras2 ras4~~ras7	.91
2021	Saramago et al.	France	7	One factor	None	.87
2022	Adamczyk et al.	Polish	7	One factor	None	.89[b]
			7	One factor	None	.55[b]
2023	Ventura-León et al.	Peru	7	One factor	ras1~~ras2 ras4~~ras7	56 [b]
			5	One factor	ras1~~ras2	.84[b]

Note. RAS = Relationship Assessment Scale. ~~ = error asociation; [a] = α coefficient; [b] = ω coefficient.

Therefore, we reviewed the items that present associations, specifically Item 1, “Do you consider that your partner meets your needs?,” and Item 2, “In general, to what extent are you satisfied with your relationship?” Both items evaluate satisfaction, but in different contexts: personal needs and overall relationship satisfaction. Although both items measure the same indicator, respondents may have been influenced by biases in item comprehension. However, Items 4 and 7, which ask about frequency of regret and number of problems in the relationship respectively, are phrased similarly and may not clearly distinguish what they are measuring. This issue will be further discussed in the study.

The Study

The study raised concerns about the structure of the RAS, particularly the wording of the items and their response options. Based on this information, the study has proposed a revised version of the Relationship Assessment Scale (RAS-R). The analysis included grammatical refinement, as well as an explanation of its psychometric attributes. This evaluation was developed in two distinct phases. Firstly, a comprehensive theoretical and content analysis of the instrument was conducted. Secondly, the aim was to assess the psychometric properties of both scales and analyze them using the Graded Response Model (GRM). These evaluations followed the paradigms of both Classical Test Theory (CTT) and Item Response Theory (IRT), thus strengthening the basis for the revision of the instrument.

METHODS

Participants

The study utilized two nonprobabilistic samples. The sample's inclusion and exclusion criteria were as follows: a) participants must have been in a couple relationship for at least six months, or currently be in a relationship of that length; b) only adults (> 18 years) in a monogamous relationship were included. Participants were excluded if c) they were involved in a long-distance relationship, and d) they were in couples' therapy at the time of application.

The first sample comprised 547 Peruvian participants, with 37.3% male and 62.7% female. Among them, 69.7% were single, 13.2% married, and 17.1% cohabiting. These individuals represented both heterosexual (93.6%) and homosexual (6.4%) relationships. The average age of the participants was 27 years (*SD* = 8.24 years), with an age range spanning from 18 to 58 years. Participants showed an average of two years in a relationship (*SD* = 0.47 years). This initial sample was subjected to RAS assessment, as well as model comparison against existing literature.

The second sample encompassed 285 Peruvian participants, with 34.7% identifying as male and 65.3% as female. Among them, 60.0% were single, 19.6% married, and 20.4% cohabiting. These participants represented both heterosexual (95.1%) and homosexual (4.9%) relationships. The participants in this sample had an average age of 29 years (*SD* = 10.1 years), ranging from 18 to 63 years. Their average relationship duration was two years (*SD* = 0.46 years). This second sample was specifically utilized to present the results of the RAS-R.

Instrument

Relationship Assessment Scale (RAS; Hendrick et al., 1998) adapted by De la Rubia (2015). Comprising seven items, respondents provide answers on a 5-point Likert-type scale, where Items 4 and 7 necessitate reverse scoring due to their inverse wording. The response options have different alternatives for each item (e.g., *very dissatisfied* to *very satisfied*, *very poor* to *very good*, *never* to *constantly*). The scale was initially a single-factor scale measuring overall relationship satisfaction. Higher scores denote greater satisfaction levels. Internal consistency reliability was assessed through the alpha coefficient, revealing satisfactory results ($\alpha = .89$).

Procedure

The study was conducted within the principles outlined in the Declaration of Helsinki (World Medical Association, 1964), and was approved by the authors' university. Subsequent to obtaining ethical clearance, an online survey was prepared to administer the measurement instruments. The online form contained the study's objectives and informed consent details. Participation hinged upon consenting to the study terms, allowing participants to proceed to sociodemographic queries and the measurement tools. The data collection occurred from February to June 2023. The participants were accessed through a dissemination strategy using the social networking platform Facebook and WhatsApp groups managed by the study authors. Additionally, individuals were approached to achieve the desired sample size.

Data Analysis

R Studio Version 4.2.2 and the packages psych (Revelle, 2023), lavaan (Rosseel, 2023), semTools (Jorgensen et al., 2022), mirt (Chalmers, 2012), ltm (Rizopoulos, 2006), Rcpp (Eddelbuettel & François, 2011), and knitr (Xie, 2023) were used. Descriptive statistics such as mean, standard deviation, skewness, and kurtosis were used to evaluate the items. Univariate normality was obtained according to the criteria of skewness (±2) and kurtosis (±7) coefficients (Finney & DiStefano, 2013). We investigated the existence of floor and ceiling effects, with extreme responses comprising less than 15% of the total (McHorney & Tarlov, 1995). Likewise, the matrix of polychoric correlations was obtained due to the ordinal nature of the items.

For the confirmatory factor analysis (CFA), the WLSMV model was used because of the categorical nature of the responses, and also because it is the estimator used when there is the presence of floor and ceiling effect (Koziol & Bovaird, 2018). Models were evaluated through the following fit indices: chi-square (χ^2), degrees of freedom (*df*), comparative fit index (CFI > .95), Tucker-Lewis index (TLI > .95), root-mean-square error of approximation (RMSEA < .08), and standardized root-mean-square residual (SRMR < .08) (Hu & Bentler, 1999). Factor loadings must exceed .50 to represent the measured latent variable (Dominguez-Lara, 2018). At the same time, modification indexes (MI) were used to evaluate the models and suggested changes for a better model fit (Thoemmes et al., 2018). Finally, reliability by internal consistency was assessed through the omega coefficient where values ≥ .80 are considered adequate (Raykov & Hancock, 2005).

Also, an Item Response Theory (IRT) analysis was applied using the 2-parameter version of the Graded Response Model (GRM) (Hambleton et al., 2010). To establish the feasibility of the IRT, we confirmed the unidimensional model through CFA. We then analyzed the correlation between item errors to evaluate local independence (Erhart et al., 2010). Once the assumptions were confirmed, we used the GRM for analysis. Model fit was assessed using the M2 index for ordinal items (Cai & Monroe, 2013), and RMSEA ≤ .08, SRMR ≤ .05, CFI and TLI ≥ .95 (Cai et al., 2021; Maydeu-Olivares & Joe, 2014). Finally, the discrimination (a) and difficulty (b) parameters were evaluated, which show the subjects' ability to answer the items and their increased difficulty as a function of the responses.

RESULTS

Phase 1

Upon dissecting the psychometric attributes of the RAS, a notable disparity emerged between the present study's findings and the extant literature. Of particular note was the divergence from the outcomes

reported by Ventura-León et al. (2023) in a Peruvian sample. This discovery held substantial import due to the congruence of our Phase 1 results, where an unmodified unidimensional model showcased satisfactory fit indices. However, to circumvent notable disparities in the psychometric attributes of the RAS within the Spanish context, a meticulous evaluation of both the items' content and the coherence of responses was undertaken. This comprehensive examination culminated in a refined version that mitigates the interpretation biases inherent in the original instrument (content analysis and consistency of responses to RAS items are available upon request from the corresponding author).

After analyzing the items, it came to light that some items harbored the potential for subjective and expansive interpretations. Inconsistencies in terms such as "good" or "normal" were also identified, inadvertently leading to responses predicated on personal perceptions of an ideal relationship. Moreover, while retaining the same number of response options as recommended by literature, we unearthed incongruities where certain options failed to enable precise item responses. Notably, Item 4's absence of a neutral or intermediate choice posed measurement challenges. Similarly, utilizing *frequently*, *very frequently*, and *constantly* as response alternatives for the same item introduced discordance, lacking uniform interval distances and thereby presenting difficulties in response selection.

To preserve the theoretical and psychometric unidimensionality of the RAS, we kept the seven items and made the following modifications. First, the items were reworded avoiding the biases initially observed. Thus, the reverse items were converted into direct items. The response options were unified into a Likert scale from *strongly disagree* to *strongly agree*. It is expected that the answers to the items will be equivalent.

Phase 2

Table 2 presents the descriptive statistics for the RAS and RAS-R items. It is evident that all items adhere to the univariate normality assumption. Notably, assessing the mean trend led to computation of the floor and ceiling effects; remarkably, a pronounced ceiling effect was observable across all items (>15%) of RAS and RAS-R. Further, the polychoric correlations unveiled robust associations, except for Items 4 and 7, which exhibited comparatively weaker connections, likely attributed to their reverse-wording nature. Considerable changes were observed in the ceiling effect and the interitem correlation of the RAS-R. The polychoric correlation indicates associations above .50, demonstrating a higher proportion of variance between items than in the RAS. The change to direct wording favored the stability of the associations. Likewise, the univariate normality assumption was used.

In addition, the study presented an interest in evaluating the unidimensional model of Hendrick et al. (1998), which has been confirmed by several studies, but not by studies in Latin America. Table 3 shows that the fit indices of Hendrick's model were acceptable (χ^2/df = 7.94; SRMR = .04; TLI = .94; CFI = .96; RMSEA = .11). Nevertheless, some indices, such as the χ^2/df, RMSEA, and their confidence intervals, showed that the model does not fit correctly. The model modification indices were reviewed, and the errors were suggested to be correlated as models in the literature. Table 3 shows the different models tested; a model with the greatest differences were the RAS^c, which had adequate fit indices (χ^2/df = 3.76; SRMR = .02; TLI = .98; CFI = .97; RMSEA = .79), but the reliability decreased considerably to .72 (ω), which is at the border of the acceptable limit (> .70). On the other hand, the model of Ventura-León et al. (2023) (RAS^e) was tested with five items and error correlations (ras1~~ras2), which showed an adequate fit (χ^2/df = 5.08; SRMR = .01; TLI = .99; CFI = .98; RMSEA = .09; ω = .91). However, the elimination of items represents a significant change to the original instrument. In this sense, the χ^2/df index for the RAS-R was lower than for the other models, its TLI and CFI indices were adequate for the model, but its RMSEA was .08, which may be due to different factors

affecting the error. The factorial loadings proved to be stable above .78, which differs from the original RAS model, which has at least one item with a loading of .51. In turn, the highest loading was for Item 2, which we will see later has a higher discrimination than the other items. Furthermore, the reliability of the instrument increased slightly compared to the RAS. When reviewing the modification indexes, there were no sufficiently substantiated suggestions for changes in the unidimensional model.

TABLE 2
Descriptive data, floor and ceiling effect, polychoric matrix of the RAS and RAS-R

Measure	Items	*M*	*SD*	g_1	g_2	Effects		Polychoric matrix					
						Floor (%)	Ceiling (%)	1	2	3	4	5	6
RAS	1	3.72	1.00	–0.89	0.62	4.4	19.7	–					
(*N* = 547)	2	3.94	0.95	–0.95	0.95	2.9	30.2	.81	–				
	3	4.03	0.88	–0.86	0.81	1.5	32.7	.72	.76	–			
	4*	4.34	0.87	–1.67	3.19	1.8	51.6	.41	.42	.46	–		
	5	3.68	0.98	–0.64	0.08	2.7	19.2	.76	.77	.80	.46	–	
	6	4.16	0.99	–1.45	2.16	4.2	43.7	.68	.73	.73	.40	.69	–
	7*	3.54	1.02	–0.03	–0.65	2.2	23.4	.35	.39	.49	.40	.38	.35
RAS-R	1	3.95	1.00	–0.94	0.54	2.8	33.3	–					
(*N* = 285)	2	4.13	0.98	–1.17	1.05	2.5	43.2	.77	–				
	3	3.71	1.01	–0.47	–0.24	2.8	25.3	.55	.63	–			
	4	4.12	1.05	–1.17	0.74	3.2	47.0	.66	.77	.53	–		
	5	3.90	0.94	–0.73	0.14	1.4	28.4	.66	.83	.54	.67	–	
	6	4.36	0.84	–1.50	2.51	1.4	54.0	.62	.74	.55	.74	.66	–
	7	4.17	0.84	–1.00	1.09	1.1	40.0	.61	.73	.41	.64	.72	.66

Note. g_1 = skewness; g_2 = kurtosis. RAS = Relationship Assessment Scale. RAS-R = Relationship Assessment Scale-Reversed. *Reverse-worded items.

TABLE 3
Comparison of different models

Models	$\chi^2(df)$	χ^2/df	SRMR	TLI	CFI	RMSEA	CI 90%	Factor loadings							
								1	2	3	4	5	6	7	ω
RAS	111.2 (14)	7.94	.04	.94	.96	.11	[.09, .13]	.87	.89	.88	.51	.88	.80	.48	.89
RAS[a]	80.2 (13)	6.17	.04	.98	.96	.10	[.08, .12]	.83	.86	.89	.52	.89	.81	.48	.71
RAS[b]	78.6 (13)	6.05	.03	.97	.95	.10	[.08, .13]	.87	.89	.88	.50	.88	.80	.46	.74
RAS[c]	45.2 (12)	3.76	.02	.98	.97	.08	[.06, .10]	.83	.86	.89	.50	.89	.81	.47	.72
RAS[d]	50.4 (5)	10.08	.02	.98	.96	.13	[.10, .17]	.87	.89	.87	–	.88	.80	–	.91
RAS[e]	20.3 (4)	5.08	.01	.99	.98	.09	[.05, .13]	.84	.87	.88	–	.89	–	.81	.90
RAS-R	42.8 (14)	3.06	.03	.97	.96	.08	[.06, .12]	.79	.95	.65	.83	.85	.82	.78	.91

Note. SRMR = standardized root-mean-square residual; TLI =Tucker-Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval. RAS = Relationship Assessment Scale. RAS-R = Relationship Assessment Scale-Reversed; [a] = ras1~~ras2; [b] = ras4~~ras7; [c] = ras1~~ras2 and ras4~~ras7; [d] = 5-item model; [e] = Ventura-León et al. (2023) model with five items and ras1~~ras2.

At the same time the TCT analysis was performed, the instruments were evaluated by IRT. Table 4 shows the results according to the GRM analysis for polytomous items. Primarily, the RAS obtained adequate fit (M2 = 61.5; *df* = 14; RMSEA = .08; SRMR = .06; TLI = .98; CFI = .98), which is an adequate fit of the model, but the RAS-R obtained more adequate fit indices (M2 = 26.5; *df* = 14; RMSEA = .06; SRMR = .04; TLI = .99; CFI = .99). In terms of item discrimination (*a*), Item 2 was the most discriminant in both measures, while for the RAS Item 4 was the least discriminant and for the RAS-R Item 3 was the least discriminant. Contrary to the RAS, which has an item with low discrimination, the RAS-R improved a stability of discrimination among the items, with the exception of Item 4, according to its own indicator, is a direct question on satisfaction. Conversely, the increase in difficulty (*b*) was consistently monotonic for both measures.

TABLE 4
Graded response model

Model	Items	*a*	*b*1	*b*2	*b*3	*b*4	M2	*df*	SRMR	RMSEA	TLI	CFI
RAS	1	3.01	–2.03	–1.47	–0.54	0.82	61.5	14	.06	.08	.98	.99
	2	3.70	–2.18	–1.79	–0.74	0.58						
	3	3.47	–2.53	–1.92	–0.82	0.51						
	4	1.17	–3.93	–3.03	–2.25	–0.11						
	5	3.57	–2.19	–1.37	–0.41	0.96						
	6	2.44	–2.17	–1.96	–1.13	0.20						
	7	0.96	–4.45	–2.39	–0.28	1.45						
RAS-R	1	2.95	–2.42	–1.24	–0.54	0.68	26.5	14	.04	.06	.10	.99
	2	4.46	–1.58	–1.33	–0.57	0.39						
	3	1.52	–2.69	–1.59	–0.09	1.12						
	4	2.76	–1.99	–1.43	–0.66	0.28						
	5	2.86	–2.07	–1.27	–0.44	0.81						
	6	2.61	–2.33	–1.90	–0.99	0.14						
	7	2.35	–2.53	–1.94	-0.80	0.51						

Note. *a* = discrimination; *b* = difficulty; SRMR = standardized root-mean-square residual; RMSEA = root-mean-square error of approximation; TLI = Tucker-Lewis index; CFI = comparative fit index. RAS = Relationship Assessment Scale. RAS-R = Relationship Assessment Scale-Reversed.

Figure 1 shows the differences between the RAS and RAS-R in terms of instrument functionality. In most cases, more stable curves are observed in the RAS-R compared to the RAS, which can be seen in the test information function (TIF), conditional standard error (CSE), test expected score (TES), option response function (ORF), and item information function (IIF). In this sense, greater heterogeneity in the item response function can be observed, with most items showing better accuracy in measuring latent traits. However, these comparisons cannot be taken as decisive criteria, as changing the item wording produces a completely different measure with the same theoretical properties.

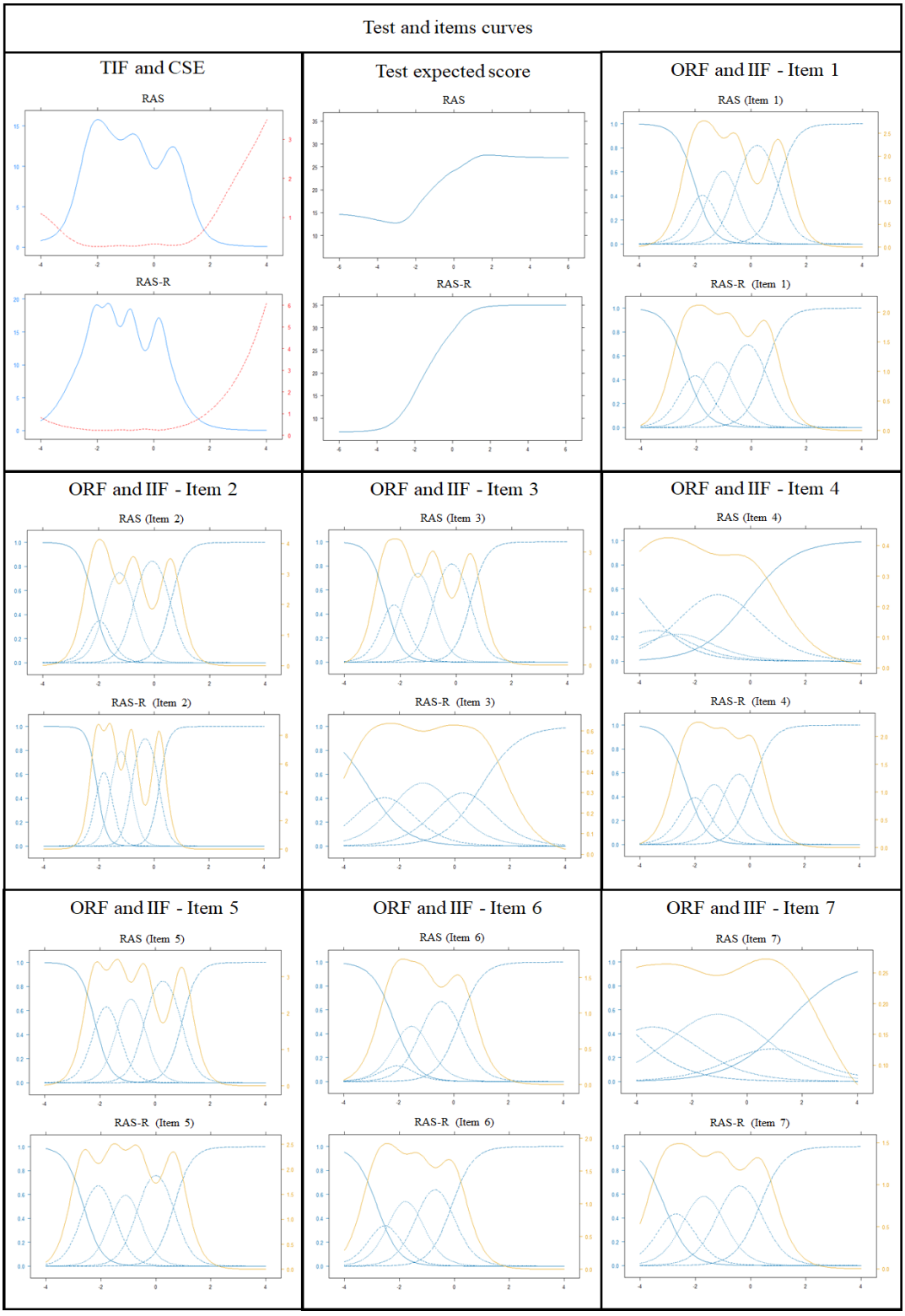

FIGURE 1
Test and items information curves of the RAS and RAS-R

Note. RAS = Relationship Assessment Scale. RAS-R = Relationship Assessment Scale-Reversed.

DISCUSSION

Consolidating an instrument that evaluates relationship satisfaction is of significant importance for the scientific and psychological field. Such an instrument enables the measurement of satisfaction within the affective and emotional dynamics of a couple, a facet intricately tied to their psychological and physical well-being (Novak et al., 2023). An appropriate measuring tool can identify satisfaction indicators. This allows for designing interventions to promote stable relationships. For the purpose of the study, the RAS was first applied, and the findings showed that the unidimensional model presents acceptable psychometric properties as established in some antecedents (e.g., Dinkel & Balck, 2005; Hassebrauck, 1991; Hendrick et al., 1998; Martos et al., 2014; Vaughn & Baier, 1999). Nevertheless, in the Spanish RAS version, a considerable diversity of models was found because the psychometric properties did not fit adequately (e.g., González-Rivera, 2020; Ventura-León et al., 2023).

Our analysis thoroughly examined the content and phrasing utilized in the Spanish RAS version models. To avoid bias in the measurement, a specific item analysis was conducted. This helped in reformulating the wording of the items as there was subjectivity in the interpretation of the content and responses. This process was necessary to ensure accuracy. Also, because the Likert-type items did not maintain a mean or neutral option and did not allow an understanding of the differences between the distances of the responses (Abad et al., 2011), it was decided to unify the response options according to the items. Therefore, a revised model (RAS-R) was proposed.

This process permits the incorporation of all items into the model, including those with nondiscriminatory RAS inverse items (Oropeza et al., 2010; Ventura-León et al., 2023). We proceeded to change the inverse wording due to its controversial potential to confuse participants when interpreting the items (Swain et al., 2008; van Sonderen et al., 2013), as well as its difficulty in controlling for acquiescence bias in scale scores (Savalei & Falk, 2014). Moreover, despite the items being measured in an identical way, it is possible that other factors may be generated due to the correlations of errors that have been documented in previous studies (DiStefano & Motl, 2006; Lindwall et al., 2012).

Nevertheless, keeping the items ensures that the instrument can measure all the original RAS indicators. One item of significant change was Item 1 (“Do you consider that your partner meets your needs?”), in which the term “needs” can be interpreted in multiple ways that affect the objectivity of the test (Brown, 2015). The response options containing statements with “normal” also incite subjectivity (Báez, 2012) about the number of problems within a couple.

The second phase of the study allowed the RAS-R psychometric properties to be evaluated. When observing the polychoric correlations between the RAS and RAS-R, we observed more robust stability in the RAS-R associations. Therefore, when testing the fit of the unidimensional model, we obtained higher indices than the RAS. Likewise, its factorial structure was preserved as initially proposed by Hendrick et al. (1998). In addition, the factor loadings were within the expected range ($\lambda > .70$). Specifically, the improvement was in the RMSEA index, implying that the proposed model is close to a perfect one (Xia & Yang, 2019). In addition, the incremental fit indices, CFI and TLI, had slight improvements compared to the RAS (Lai & Green, 2016). As a comparative estimator, the chi-square ratio was lower than the other models, which is more appropriate (Marsh & Hocevar, 1985). At the same time, the reliability increased in omega compared to the unidimensional RAS and the Latin American priors. However, it is important to approach these differences with caution as the version used in the antecedents only incorporated the RAS and not the RAS-R. Additionally, many Latin American antecedents changed the model, which impacted the instrument’s reliability.

The RAS and RAS-R were tested through the GRM (IRT) as a last procedure. It was observed that Item 2 of RAS-R ("I am satisfied with my relationship") was the item with the highest discrimination between those with more satisfaction and those who do not feel satisfied. It was to be expected that Item 4 represented the highest discrimination because it is a direct indicator of relationship satisfaction. The other items maintained adequate discrimination of satisfaction. Although Item 3 of RAS-R ("My relationship is better compared to other couples") had the lowest discrimination, it was of equal theoretical relevance to the instrument. Compared with the RAS, the RAS-R has lower item discrimination, but the discrimination between items is equal and compensates for the much lower discrimination of Item 7 of the RAS. Differences and similarities can be observed in the information values of the two scales, with the RAS-R showing greater stability of responses due to the standardization of previously established responses. On the other hand, if these results are compared with the only antecedent using IRT, it can be concluded that Item 2 tends to be the most discriminating, while Item 6 is the least discriminating (Ventura-León et al., 2023). On the other hand, the monotonic increase of the items allows us to conclude that the difficulty is in accordance with the discrimination of the measured indicator.

Limitations

The study carries limitations that warrant consideration. Initially, the sampling strategy employed was nonprobabilistic convenience sampling, consequently constraining the generalizability of the findings. It's prudent to exercise caution when extending the study's implications beyond its specific context. Secondly, the study did not delve into RAS-R invariance, underscoring the need for subsequent research endeavors to address this gap and furnish new insights into invariance patterns. Additionally, detecting ceiling and floor effects posed another constraint during data analysis. Also, the RAS and RAS-R were not administered to the same sample, which would have been advantageous for developing item comparison analyses and for determining whether the items measure as accurately as the original version. Finally, the lack of assessment of the RAS-R with other criterion measures did not allow us to provide evidence of validity based on association with other variables. Further studies could address these limitations to test both measures and their functionality.

Conclusions

In conclusion, the RAS-R was introduced, based on the adaptations observed in the factorial structure across the literature. The psychometric revision of the RAS was initiated, and the unidimensional model was tested and found to be adequate. Nevertheless, these results may be altered in other scenarios because the literature shows that the RAS may vary in structure and reliability. Therefore, a content review of the RAS was conducted and the result was a modified scale in wording and directionality, called RAS-R. This version showed adequate psychometric properties and reliability, even higher than the RAS. Thus, it proved a more robust instrument regarding the associations between items and their internal consistency. Likewise, it was observed that the IRT analysis showed that the items discriminate adequately and the difficulty index increases as the person possesses or does not possess the trait. Finally, the RAS-R is a suitable instrument for measuring satisfaction in a couple's relationship and can be used in different contexts, both in research and to identify couple satisfaction in preventive or promotional programs.

REFERENCES

Abad, F. J., Olea, J., Ponsoda, V., & García, C. (2011). *Medición en ciencias sociales y de la salud* [*Measurement in health and social sciences*]. Editorial Síntesis, S. A.

Adamczyk, K., Kleka, P., & Frydrychowicz, M. (2022). Psychometric functioning, measurement invariance, and external associations of the Relationship Assessment Scale in a sample of Polish adults. *Scientific Reports*, *12*(1), 22157. https://doi.org/10.1038/s41598-022-26653-6

Báez, J. (2012). Normalidad, anormalidad y crisis [Normality, abnormality and crisis]. *Tesis Psicológica*, *7*(2), 134-145. https://revistas.libertadores.edu.co/index.php/TesisPsicologica/article/view/270

Brown, T. (2015). *Confirmatory factor analysis for applied research*. The Guilford Press.

Cai, L., Chung, S. W., & Lee, T. (2021). Incremental model fit assessment in the case of categorical data: Tucker-Lewis index for Item Response Theory modeling. *Prevention Science*, *24*, 455-466. https://doi.org/10.1007/s11121-021-01253-4

Cai, L., & Monroe, S. (2013). IRT model fit evaluation from theory to practice: Progress and some unanswered questions. *Measurement: Interdisciplinary Research & Perspective*, *11*(3), 102-106. https://doi.org/10.1080/15366367.2013.835172

Cassepp-Borges, V., & Pasquali, L. (2011). Características psicométricas da Relationship Assessment Scale [Relationship Assessment Scale's psychometric characteristics]. *Psico-USF*, *16*(3), 255-264. https://doi.org/10.1590/S1413-82712011000300002

Çelik, E. (2014). Adaptation of Relationship Assessment Scale to Turkish culture: Study of validity and reliability. *International Journal of Psychology and Educational Studies*, *1*(1), 1-7. https://doi.org/10.17220/ijpes.2014.01.001

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

Chowdhury, T., Islam, A., & Rahman, S. T. (2016). Adaptation of the Relationship Assessment Scale for use in Bangladesh. *Bangladesh Psychological Studies*, *26*, 25-32. https://www.researchgate.net/profile/Azharul-Islam/publication/313376592_Adaptation_of_the_Relationship_Assessment_Scale_for_Use_in_Bangladesh/links/589895784585158bf6f6bbee/Adaptation-of-the-Relationship-Assessment-Scale-for-Use-in-Bangladesh.pdf

Collins, W. A., Welsh, D. P., & Furman, W. (2009). Adolescent romantic relationships. *Annual Review of Psychology*, *60*(1), 631-652. https://doi.org/10.1146/annurev.psych.60.110707.163459

De la Rubia, J. M. (2015). Validación del modelo unidimensional de la Escala de Valoración de la Relación en personas casadas y en unión libre de Monterrey, México [Validation of the single-factor modelof the Relationship Assessment Scale among married and cohabiting persons from Monterrey, Mexico]. *Revista Internacional de Psicología*, *14*(02), 1-71. https://doi.org/10.33670/18181023.v14i02.134

Dehshiri, G. R., & Mousavi, S. F. (2016). Psychometric properties of the Persian version of the Relationship Assessment Scale among couples. *Family Research*, *12*(1), 141-154. https://jfr.sbu.ac.ir/index.php/ayenehmarefat/article/view/article_97223_c1e66f8a8382e31a1917891425c4ca01.pdf

Dinkel, A., & Balck, F. (2005). An evaluation of the German Relationship Assessment Scale. *Swiss Journal of Psychology*, *64*(4), 259-263. https://doi.org/10.1024/1421-0185.64.4.259

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(3), 440-464. https://doi.org/10.1207/s15328007sem1303_6

Dominguez-Lara, S. A. (2018). Propuesta de puntos de corte para cargas factoriales: una perspectiva de fiabilidad de constructo [Proposal for cut-offs for factor loadings: A construct reliability perspective]. *Enfermería Clínica*, *28*(6), 401-402. https://doi.org/10.1016/j.enfcli.2018.06.002

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1-18. https://doi.org/10.18637/jss.v040.i08

Erhart, M., Hagquist, C., Auquier, P., Rajmil, L., Power, M., & Ravens-Sieberer, U. (2010). A comparison of Rasch item-fit and Cronbach's alpha item reduction analysis for the development of a Quality of Life scale for children and adolescents. *Child: Care, Health and Development*, *36*(4), 473-484. https://doi.org/10.1111/j.1365-2214.2009.00998.x

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 439-492). IAP Information Age Publishing.

González-Rivera, J. A. (2020). Fiabilidad y validez de la Escala de Evaluación de Relaciones en mujeres puertorriqueñas [Reliability and validity of the Relationship Assessment Scale in Puerto Rican women]. *Informes Psicológicos*, *20*(1), 11-18. https://doi.org/10.18566/infpsic.v20n1a01

Graham, J. M., Diebels, K. J., & Barnow, Z. B. (2011). The reliability of relationship satisfaction: A reliability generalization meta-analysis. *Journal of Family Psychology*, *25*(1), 39-48. https://doi.org/10.1037/a0022441

Hambleton, R. K., van der Linden, W. J., & Wells, C. S. (2010). IRT models for the analysis of polytomously scored data: Brief and selected history of modelbuilding advances. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response models* (pp. 21-42). Routledge.

Hassebrauck, M. (1991). ZIP — Ein Instrumentarium zur Erfassung der Zufriedenheit in Paarbeziehungen [ZIP: A scale for assessment of satisfaction in close relationships]. *Zeitschrift Für Sozialpsychologie*, *22*(4), 256-259. https://www.researchgate.net/publication/261180161_ZIP_-Ein_Instrumentarium_zur_Erfassung_der_Zufriedenheit_in_Paarbeziehungen

Hendrick, S. S. (1988). A generic measure of relationship satisfaction. *Journal of Marriage and the Family*, *50*(1), 93-98. https://doi.org/10.2307/352430

Hendrick, S. S., Dicke, A., & Hendrick, C. (1998). The Relationship Assessment Scale. *Journal of Social and Personal Relationships*, *15*(1), 137-142. https://doi.org/10.1177/0265407598151009

Hernandez, J. A. E. (2014). Evidências de validade da Escala de Avaliação do Relacionamento [Validity evidence of the Relationship Assessment Scale]. *Estudos de Psicologia (Campinas)*, *31*(3), 327-336. https://doi.org/10.1590/0103-166X2014000300001

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Institut de Publique Sondage d'Opinion Secteur. (2023). *Love life satisfaction around the world*. Retrieved from https://www.ipsos.com/sites/default/files/ct/news/documents/2023-02/Global Advisor - Love Life Satisfaction - Graphic Report.pdf

Instituto Nacional de Estadística e Informática. (2020). *Perú: cambios en el estado civil o conyugal* [*Peru: Changes in civil or marital status*]. Retrieved from https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1742/Libro.pdf

Instituto Nacional de Estadística e Informática. (2023). *Más de 77 mil 500 matrimonios fueron inscritos a nivel nacional* [*More than 77,500 marriages were registered nationwide*]. Retrieved from https://m.inei.gob.pe/media/MenuRecursivo/noticias/nota-de-prensa-no-017-2023-inei-002.pdf

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling. R package Version 0.5-6*. https://cran.r-project.org/package=semTools

Kanter, J. B., & Proulx, C. M. (2021). Understanding the early years of socioeconomically disadvantaged couples' marriages. *Family Process*, *60*(3), 806-822. https://doi.org/10.1111/famp.12602

Koziol, N. A., & Bovaird, J. A. (2018). The impact of model parameterization and estimation methods on tests of measurement invariance with ordered polytomous data. *Educational and Psychological Measurement*, *78*(2), 272-296. https://doi.org/10.1177/0013164416683754

Lai, K., & Green, S. B. (2016). The problem with having two watches: Assessment of fit when RMSEA and CFI disagree. *Multivariate Behavioral Research*, *51*(2-3), 220-239. https://doi.org/10.1080/00273171.2015.1134306

Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment*, *94*(2), 196-204. https://doi.org/10.1080/00223891.2011.645936

Maroufizadeh, S., Omani-Samani, R., Almasi-Hashiani, A., Navid, B., Sobati, B., & Amini, P. (2018). The Relationship Assessment Scale (RAS) in infertile patients: A reliability and validity study. *Middle East Fertility Society Journal*, *23*(4), 471-475. https://doi.org/10.1016/j.mefs.2018.04.001

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin*, *97*(3), 562-582. https://doi.org/10.1037/0033-2909.97.3.562

Martos, T., Sallay, V., Szabó, T., Lakatos, C., & Tóth-Vajna, R. (2014). Psychometric characteristics of the Hungarian version of the Relationship Assessment Scale (RAS-H). *Mentálhigiéné És Pszichoszomatika*, *15*(3), 245-258. https://doi.org/10.1556/Mental.15.2014.3.6

Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, *49*(4), 305-328. https://doi.org/10.1080/00273171.2014.911075

McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research*, *4*(4), 293-307. https://doi.org/10.1007/BF01593882

Novak, J. R., Wilson, S. J., Ermer, A. E., & Harper, J. M. (2023). Aging together: Dyadic profiles of older couples' marital quality, psychological well-being, and physical health. *Journal of Social and Personal Relationships*, *40*(6), 1897-1919. https://doi.org/10.1177/02654075221135451

Oropeza, R. E., Armenta, C., García, A., Padilla, N., & Díaz-Loving, R. (2010). Validación de la Escala de Evaluación de Relaciones en la población Mexicana [Validation and adaptation of a Marital Satisfaction Scale in Mexican population]. *Psicología Iberoamericana*, *18*(2), 56-65. https://doi.org/10.48102/pi.v18i2.253

Özdemir, B., & Demir, A. (2019). Romantic relationship satisfaction, age, course load, satisfaction with income and parental status as predictors of instructors' burnout: Evidence from a correlational study. *Current Psychology*, *38*(5), 1083-1098. https://doi.org/10.1007/s12144-017-9724-3

Rask, M., Malm, D., Kristofferzon, M.-L., Roxberg, A., Svedberg, P., Arenhall, E., Baigi, A., Brunt, D., Fridlund, B., Ivarsson, B., Nilsson, U., Sjöström-Strand, A., Wieslander, I., Benzein, E., & SAMMI-Study Group. (2010). Validity and reliability of a Swedish version of the Relationship Assessment Scale (RAS): A pilot study. *Canadian Journal of Cardiovascular Nursing*, *20*(1), 16-21. http://www.ncbi.nlm.nih.gov/pubmed/20301858

Raykov, T., & Hancock, G. R. (2005). Examining change in maximal reliability for multiple-component measuring instruments. *British Journal of Mathematical and Statistical Psychology*, *58*(1), 65-82. https://doi.org/10.1348/000711005X38753

Revelle, W. (2023). *psych: Procedures for psychological, psychometric, and personality research.* https://cran.r-project.org/web/packages/psych/index.html

Rivera, D., Cruz, C., & Muñoz, C. (2011). Satisfacción en las relaciones de pareja en la adultez emergente: el rol del apego, la intimidad y la depresión [Satisfaction in intimate relationships in emerging adulthood: The role of attachment, intimacy and depression]. *Terapia Psicológica*, *29*(1), 77-83. https://doi.org/10.4067/S0718-48082011000100008

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and Item Response Theory analyses. *Journal of Statistical Software*, *17*(5), 1-25. https://doi.org/10.18637/jss.v017.i05

Rosseel, Y. (2023). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. https://www.jstatsoft.org/v48/i02/

Sánchez, C., Carreño, J., & Mota, C. (2020). Indicadores psicológicos y de pareja relacionados con la insatisfacción marital: estrategias de intervención [Psychological and couple indicators related to marital dissatisfaction: Intervention strategies]. *Psicología y Salud*, *31*(1), 81-92. https://doi.org/10.25009/pys.v31i1.2678

Saramago, M., Lemétayer, F., & Gana, K. (2021). Adaptation et validation de la version française de l'échelle d'évaluation de la relation [Adaptation and validation of the French version of Relationship Assessment Scale]. *Psychologie Française*, *66*(4), 333-343. https://doi.org/10.1016/j.psfr.2020.09.004

Savalei, V., & Falk, C. F. (2014). Recovering substantive factor loadings in the presence of acquiescence bias: A comparison of three approaches. *Multivariate Behavioral Research*, *49*(5), 407-424. https://doi.org/10.1080/00273171.2014.931800

Statista. (2022). *Número de divorcios que se registraron en Perú de 2014 a 2023* [*Number of divorces registered in Peru from 2014 to 2023*]. https://es.statista.com/estadisticas/1241930/numero-de-divorcios-registrados-en-peru/

Superintendencia Nacional de los Registros Públicos. (2022). *Más de 8 mil parejas inscribieron su divorcio en la Sunarp durante el 2021* [*More than eight thousand couples registered their divorce with Sunarp during 2021*]. https://www.gob.pe/institucion/sunarp/noticias/586266-mas-de-8-mil-parejas-inscribieron-su-divorcio-en-la-sunarp-durante-el-2021

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, *45*(1), 116-131. https://doi.org/10.1509/jmkr.45.1.116

Sychev, O. A. (2016). Русскоязычная версия Шкалы оценки отношений [Russian version of the Assessment Relationship Scale]. *Психология. Журнал Высшей Школы Экономики*, *13*(2), 386-409. https://doi.org/10.17323/1813-8918-2016-2-386-409

Thoemmes, F., Rosseel, Y., & Textor, J. (2018). Local fit evaluation of structural equation models using graphical criteria. *Psychological Methods*, *23*(1), 27-41. https://doi.org/10.1037/met0000147

van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS ONE*, *8*(7), Article e68967. https://doi.org/10.1371/journal.pone.0068967

Vaughn, M., & Baier, M. E. M. (1999). Reliability and validity of the Relationship Assessment Scale. *The American Journal of Family Therapy*, *27*(2), 137-147. https://doi.org/10.1080/019261899262023

Ventura-León, J., Lino-Cruz, C., & Caycho-Rodríguez, T. (2023). Relationship satisfaction in young couples: Evidence for validity of short scale combining CFA and IRT. *Journal of Sex & Marital Therapy*, *49*(4), 420-431. https://doi.org/10.1080/0092623X.2022.2129119

World Medical Association. (1964). *Declaration of Helsinki — Ethical principles for medical research involving human subjects*. https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409-428. https://doi.org/10.3758/s13428-018-1055-2

Xie, Y. (2023). *knitr: A general-purpose package for dynamic report generation in R* (R package Version 1.45). https://yihui.org/knitr/