To be or not to be satisfied in your romantic relationship: evaluating the reliability and validity of the Valentine scale

Jón Ingi Hlynsson, Ísak Örn Ívarsson, Gerhard Andersson & Per Carlbring

Published online: 10 Feb 2025.

Submit your article to this journal ⬈

Article views: 3970

View related articles ⬈

View Crossmark data ⬈

To be or not to be satisfied in your romantic relationship: evaluating the reliability and validity of the Valentine scale

Jón Ingi Hlynsson [a,b], Ísak Örn Ívarsson [a,c], Gerhard Andersson [d,e,f] and Per Carlbring [b]

aDepartment of Psychology, University of Iceland, Reykjavík, Iceland; bDepartment of Psychology, Stockholm University, Stockholm SE-106 91, Sweden; cLandspítali, The National University Hospital of Iceland, Reykjavík, Iceland; dDepartment of Behavioural Sciences and Learning, Department of Linköping University, Linköping, Sweden; eDepartment of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden; fDepartment of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden

ABSTRACT

An intimate partner relationship is one of the most significant life goals for humans. Romantic relationships can promote healthy behavior and buffer against the development of psychiatric disorders. However, reliable and valid measures of relationship satisfaction are lacking. The Valentine scale is a freely available brief measure of relationship satisfaction (https://osf.io/fb72s), intended to provide an easily interpretable index of relationship satisfaction. Across two studies, we evaluated the reliability, validity, and factor structure of the Valentine scale. Study One ($n = 851$) explored the factor structure of the Valentine scale, assessed its test–retest reliability, and criterion-related validity. Study Two ($n = 527$) confirmed the factor structure of the Valentine scale, explored its measurement invariance, and further evaluated criterion-related validity. The results supported a unidimensional structure of the Valentine scale. Furthermore, the Valentine scale exhibited good internal reliability (Cronbach's alpha = .75 and .81 in Study One and Two, respectively), high test–retest reliability ($ICC_3 = .80$ at a two-week follow-up in Study One), and appropriate criterion-related validity demonstrating positive correlations with other measures of relationship satisfaction and positive affect, as well as and negative correlations with measures of psychopathology. Together, these findings provide good support for the usage of the Valentine scale to quantify relationship satisfaction.

Introduction

A fulfilling intimate relationship is one of the most significant life goals (Nair, 2003). Humans are inherently social beings (Björnsson, 2016), and a deep, personal connection with a partner represents the pinnacle of social engagement. Fulfilling intimate relationships not only bring joy but also offer essential social support and can mitigate the impact of various stressors (Dugal et al., 2018; Epstein & Zheng, 2017).

For instance, married individuals report higher psychological well-being and greater perceived social support, while divorced individuals report significantly lower levels of psychological well-being, an effect mediated by lower levels of perceived social support (Soulsby & Bennett, 2015).

Satisfaction in one's relationship is thus associated with a myriad of positive benefits (Dugal et al., 2018). For instance, supportive relationships enhance self-efficacy, aiding in healthier behavior changes such as diet and exercise, through mutual modeling and support. Marital quality also affects shared behaviors like eating and sleeping (for a review, see Robles, 2014). In contrast, relationship discord (cf. low relationship adjustment) is predictive of psychiatric disorders such as major depression and anxiety disorders (Miller et al., 2013; Whisman, 2013). Relationship discord can also lead to other negative coping behaviors like substance use or unhealthy eating (Robles, 2014).

Numerous instruments have been developed to assess relationship satisfaction (The CORE Lab, 2024). One example is the Dyadic Adjustment Scale (DAS; Spanier, 1976), which is a self-report measure of relationship quality comprising four factors: Dyadic consensus, dyadic satisfaction, dyadic cohesion, and affectional expression. Although the DAS is a widely used relationship satisfaction measure (Bühler et al., 2021), its reliability and factor structure have been contested (Graham et al., 2006; Gudjonsdottir et al., 2021; Norton, 1983). A recent analysis found support for a bifactor exploratory structural equation model (ESEM) solution for the DAS (Vajda et al., 2019). Although Vajda et al. (2019) reported strict measurement invariance across genders for this bifactor-ESEM solution, the implication is that the DAS total score cannot be justifiably used which heavily complicates the scale's interpretation (Norton, 1983). Furthermore, the DAS has also been criticized for its variation of the number of items assessing each factor, varying number of response alternatives, and inappropriate item weighting (Gudjonsdottir et al., 2021; Norton, 1983). Taken together, despite its popularity, the DAS suffers from critical imperfections as an overall index of relationship satisfaction.

The Quality of Dyadic Relationships (QDR) was developed as a measure of relationship satisfaction, built on the DAS to combat its limitations (Ahlborg et al., 2005). The QDR was aimed to be reflective of a broader conceptualization of relationship quality and comprises five factors: cohesion, consensus, satisfaction, sensuality, and sexuality. Although QDR has undergone preliminary construct validation and its internal consistency reliability has been estimated (Gudjonsdottir et al., 2021), it lacks sufficient validity evidence (The CORE Lab, 2024; for an overview of validity evidence, see; DeVellis & Thorpe, 2021); its factor structure has not been confirmed and data on test–retest reliability is lacking. Since the QDR represents a further development of the DAS, the QDR likely suffers from many of the same flaws as the DAS.

Other scales of relationship satisfaction include the Maudsley Marital Questionnaire (MMQ), a 20-item self-report questionnaire comprising three factors (marital, sexual, and general life adjustment; Arrindell et al., 1983); the Romance Qualities Scale (RQS), a 23-item measure evaluating five fundamental dimensions of relationship quality (companionship, conflict, help, security, and closeness; Ponti et al., 2010); and the Kansas Marital Satisfaction Scale (KMSS), a 3-item general index of relationship satisfaction (Schumm et al., 1986). Evaluating every relationship satisfaction measurement is beyond the scope of this article (for a review of relationship satisfaction instruments, see The CORE Lab, 2024), but the aforementioned instruments represent a sample of some of the

best-validated measures of relationship satisfaction. The MMQ, RQS, and KMSS have all been validated (The CORE Lab, 2024) but are limited by either a multifactor structure or exclusive inclusion of non-specific items.

Unidimensionality and careful attention to item selection are vital components in the development of a psychometrically sound relationship satisfaction instrument. Firstly, relationship satisfaction is a singular construct that operationalizes how satisfied a person is in their relationship. Although relationship satisfaction can be influenced by a myriad of other factors, the construct *relationship satisfaction* remains a singular dimension (cf. Delatorre & Wagner, 2020). Secondly, when all items from five of the most frequently cited relationship satisfaction measures are analyzed together, a unidimensional structure emerges (Funk & Rogge, 2007). This in turn raises the question of how items from multifactor scales form a unidimensional structure when combined. Finally, when a scale assesses multiple dimensions unintentionally, it introduces error, reducing the reliability of the measure. For instance, scales like the DAS have been criticized for heterogeneous item content that introduces potential confounding variance into the measure (e.g. confounding error variance from the construct "communication"; Funk & Rogge, 2007). Taken together, the lack of unidimensionality and careful item-selection can indicate a limitation with a relationship satisfaction instrument.

Such limitations may also indicate that insufficient attention has been paid to the theoretical underpinnings of *relationship satisfaction* as a construct (i.e. data-driven models as opposed to theory-driven models; Fried, 2020). Specifically, relationship satisfaction is operationalized as a unidimensional construct that quantifies how satisfied a person is in their relationship (cf. Delatorre & Wagner, 2020). This operationalization allows for precise measurement along the satisfaction continuum over time, which is typically how couple's therapy research outcomes are operationalized. Furthermore, many established instruments, while psychometrically sound, are often too lengthy, costly, or impractical for frequent administration, such as on a weekly basis, which is often required in therapeutic interventions for couples.

The Valentine scale (freely available in three languages at https://osf.io/fb72s/) is a novel scale, designed to address specific limitations observed in existing measures of relationship satisfaction, particularly in the context of their application within therapeutic settings. Designed be a single-factor index of relationship satisfaction (Burman et al., 2018), the Valentine scale consists of seven items rated on a 4-point Likert scale, with total scores ranging from 0 to 21, and comprises a combination of specific items (e.g. about conflict resolution and trust in the relationship) and general items (e.g. overall relationship satisfaction).

The development of the Valentine scale was guided by cognitive-behavioral theory (CBT), which posits that relationship satisfaction is intricately linked to the interplay of thoughts, behaviors, and emotions within the dyad. The items on the scale were carefully constructed to reflect key domains of relationship functioning that are central to CBT-based interventions, including problem-solving abilities, emotional intimacy, and conflict resolution. These domains are considered critical for both the maintenance and enhancement of relationship satisfaction, and their inclusion in the scale was informed by the extensive work of Neil S. Jacobson and his contributions to behavioral couple therapy (Jacobson et al., 2000).

Preliminary analyses of the Valentine scale ($n = 691$) indicated high test–retest reliability ($r = .80$) and adequate concurrent validity with the DAS ($r = .74$; Burman et al.,

2018). However, these findings are in need of replication. Additionally, the Valentine scale's internal reliability and factor structure have not been evaluated. Taken together, although the Valentine scale shows promise as a comprehensive and easily interpretable measure of relationship satisfaction, its reliability and unidimensionality needs to be assessed before the interpretation of its total score can be justified for clinical practice.

This study aims to assess the factor structure, reliability, and criterion-related validity of the Valentine scale across two studies. In particular, we aimed to determine whether the Valentine scale can be viewed as a unidimensional measure of relationship satisfaction using factor analysis, assessing its internal and temporal reliability, and criterion-related validity. In light of the abovementioned prior findings and the theoretical underpinning of the scale's construction (cf. Burman et al., 2018), we expected the Valentine scale to be a unidimensional measure of relationship satisfaction. Furthermore, we expected total scores on the Valentine scale to be positively correlated with quality of life (Doss et al., 2016; Dugal et al., 2018; Soulsby & Bennett, 2015) and other instruments that evaluate relationship satisfaction (Burman et al., 2018), but negatively correlated with measures of psychopathology (Doss et al., 2016; Finn et al., 2018; Malivoire et al., 2024).

Due to the explorative nature of this study, no formal hypotheses are put forth. However, our overarching research question was: Is the Valentine scale a unidimensional, reliable measure of relationship satisfaction?

Study one: internet-based couple's therapy sample

The primary objective of this study was to evaluate the factor structure of the Valentine scale. As such, we only present information from Study One that is relevant to the present analysis here; detailed information about the original study is detailed elsewhere (Samuelsson & Way, 2018). In Study One, we aimed to evaluate the factor structure of the Valentine scale. Additionally, we conducted a reliability analysis and gather preliminary validity evidence by examining concurrent-convergent validity with related scales (i.e. the DAS and BBQ), as well as concurrent-discriminant validity with measures of psychiatric disorders (i.e. the GAD-7 and PHQ-9). Specifically, we predicted a positive correlation between the Valentine scale and DAS and the Valentine scale and BBQ and a negative correlation with measures of psychopathology (i.e. lower scores on the GAD-7 and PHQ-9).

Method

Sample characteristics

Of the 901 participants seeking an online course to strengthen their relationship, 851 completed the pre-treatment screening and responded to the Valentine scale. Descriptive statistics for these participants are presented in Table 1.

Study design

Participants were primarily recruited through social media, although the study was also advertised in traditional media outlets like radio and TV. The study implemented specific

Table 1. Descriptive statistics for the internet-based couple's therapy sample.

Age: M (SD) [min-max] in Years	41.34 (10.03) [21–75]
Gender: N (%)	
Female	443 (53.7%)
Male	382 (46.3%)
Years of Education: M (SD) [min-max]	16.26 (2.96) [2–32]
Marital Status: M (SD) [min-max]	
Living together	353 (43.5%)
Married	458 (56.5%)
Number of Children: M (SD) [min-max]	1.71 (1.03) [0–6]
Employment Status: N (%)	
Studying	43 (5.9%)
Working	682 (92.8%)
Sick Leave	10 (1.4%)
Measures: M (SD) [min-max]	
Valentine Scale	12.86 (3.03) [3–21]
DAS	95.82 (14.12) [46–142]
GAD-7	3.76 (3.47) [0–20]
PHQ-9	5.60 (4.62) [0–27]
BBQ	52.16 (19.63) [3–96]

Numerical variables are reported as mean (SD) [range: minimum-maximum]. Categorical variables are reported as count (percentage). Data were missing for gender from 26 participants, for age from 25 participants, for years of education from 40 participants, for employment status for 116 participants, for GAD-7 for 15 participants, and for the BBQ for 18 participants. Acronyms: DAS = Dyadic Adjustment Scale; GAD-7 = Generalized Anxiety Disorder scale 7-item; PHQ-9 = Patient Health Questionnaire 9-item; BBQ = Brunnsviken Brief Quality of Life Inventory.

inclusion and exclusion criteria for participating couples. Inclusion criteria mandated that couples had been in a relationship for at least 1 year and cohabiting for a minimum of 6 months. Both partners needed to be fluent in Swedish, have access to a computer with internet connection, and experience significant relationship dissatisfaction. All participants were required to be over 18 years old. Exclusion criteria encompassed ongoing psychological treatment, psychoactive pharmacological treatment with a dose stable for less than 6 months, severe depression and/or suicidal thoughts, serious physical or sexual violence in the relationship, sexual dysfunction as the primary reason for seeking help, and other ongoing primary psychiatric problems deemed to require priority treatment. The present study primarily leverages data from the pre-treatment screening phase of the larger study, although baseline assessment phase measures for the Valentine scale are used for evaluating test–retest reliability; detailed information about the original study is provided elsewhere (Samuelsson & Way, 2018).

Measures

The Valentine scale

The Valentine scale is a 7-item self-report questionnaire, designed to serve as a single-factor measure of relationship satisfaction, effectively summarizing key dimensions of relationship dynamics (Burman et al., 2018). Each item is rated on a 4-point Likert scale from 3 to 0, except for items 4 and 5, which are scored from 0 to 3 (reverse scoring). This results in total scores ranging from 0 to 21, with higher scores indicative of higher relationship satisfaction. The Valentine

scale is sensitive to changes in relationship satisfaction following couple's therapy (Samuelsson & Way, 2018). Furthermore, preliminary investigations into the Valentine scale indicated high test–retest reliability ($r = .80$) and high convergent validity with the DAS ($r = 74$; Burman et al., 2018). In this sample, the internal consistency reliability for the Valentine scale at screening was good, Cronbach's $\alpha = .75$ [95% CI: .73, .77].

The dyadic adjustment scale (DAS)

The DAS is a 32-item self-report questionnaire that indexes overall relationship satisfaction (Spanier, 1976). The DAS produces a total score but can also be broken down into four subscales: dyadic consensus, dyadic satisfaction, dyadic cohesion, and affectional expression (Spanier, 1976). Although previous studies contest the usage of a DAS total score (Graham et al., 2006; Norton, 1983; Vajda et al., 2019), it remains commonly used (Bühler et al., 2021; The CORE Lab, 2024). In this sample, the internal consistency reliability for the DAS total scale at screening was good, Cronbach's $\alpha = .89$ [95% CI: .88, .90].

Generalized anxiety disorder scale-7 (GAD-7)

The GAD-7 is a 7-item self-report questionnaire that quantifies the severity of anxiety symptoms (Spitzer et al., 2006). Although it was designed to measure symptoms of generalized anxiety disorders, its items are sensitive to various anxiety disorders (Kroenke et al., 2010) and demonstrate good accuracy and discrimination ability in both clinical settings and the general population (Hlynsson & Carlbring, 2024; Martin-Key et al., 2022). A total score of 8 or higher is considered a diagnostic indicator for an anxiety disorder (Luo et al., 2019). In this sample, the internal consistency reliability for the GAD-7 was good, Cronbach's $\alpha = .86$ [95% CI: .84, .87].

Patient health questionnaire 9-item scale (PHQ-9)

The PHQ-9 is a 9-item self-report questionnaire that quantifies the severity of depressive symptoms (Kroenke et al., 2001). Designed to capture all nine symptoms of major depression, the PHQ-9 routinely demonstrates good accuracy and discrimination ability in both clinical settings and the general population (Hlynsson & Carlbring, 2024; Martin-Key et al., 2022). A total score of 10 or higher is considered a diagnostic indicator for depression (Kroenke et al., 2001, 2010). In this sample, the internal consistency reliability for the PHQ-9 was good, Cronbach's $\alpha = .85$ [95% CI: .84, .87].

Brunnsviken brief quality of life inventory (BBQ)

The BBQ is a 12-item self-report questionnaire that indexes overall subjective quality of life across six life domains: Leisure, View on Life, Creativity, Learning, Friends and Friendship, and View on Self (Lindner et al., 2016). Previous studies have found the BBQ to be a reliable and unidimensional indicator of quality of life across divergent groups (i.e. age, gender, and psychiatric conditions; Hlynsson et al., 2024). In this sample, the internal consistency reliability for the BBQ was good, Cronbach's $\alpha = .79$ [95% CI: .77, .81].

Data analysis

Since the factor structure of the Valentine scale has not been evaluated previously, we began by examining the intercorrelations between items using a polychoric correlation matrix. Derived from Spearman's rho rank order correlation coefficient, polychoric correlation matrices are preferred when the data is ordinal (Olsson, 1979). Thereafter, the suitability of the data for factor analysis was evaluated using the Kaiser-Meyer-Olkin's (KMO) factor adequacy and Bartlett's test of homogeneity of variances. A KMO value above .80 and a significant Bartlett's test of homogeneity of variances was taken as an indicator of the suitability of data for factor analysis (DeVellis & Thorpe, 2021; Kaiser, 1970).

Upon meeting those criteria, an exploratory factor analysis (EFA) was conducted to ascertain the measure's dimensionality and item fit using the *psych* package in R Studio (Revelle, 2021). For additional information, we examined the "elbow" on the scree plot (Cattell, 1966) and conducted a parallel analysis (Horn, 1965). The "elbow" of the scree plot indicates the point where eigenvalues plateau. Parallel analysis, using 1,000 iterations, offers an estimated number of factors and compares this to the chance of randomly finding the same number of factors (Horn, 1965). Information from the scree plot and parallel analysis were juxtaposed with the Valentine scale's theoretical underpinnings to determine the number of factors of the scale. Finally, each item in the Valentine scale was fit to an EFA model using a weighted least squares (WLS) method, appropriate for ordinal data and robust against violations of normality, and an orthogonal (varimax) rotation due to our expectation of unidimensionality (Fabrigar & Wegener, 2011). Factor loadings below 0.30 were considered poor, factor loadings between 0.30 and 0.40 were considered acceptable, and factor loadings higher than 0.40 were considered good (Fabrigar & Wegener, 2011; Howard, 2016).

To evaluate the internal consistency reliability, both Cronbach's alpha and McDonald's omega were calculated (cf. Cortina et al., 2020). An alpha level of between .70 and .84 and an omega value between 65 and .80 were interpreted as tentatively acceptable reliability evidence, while alpha level of above .85 and an omega value over .80 were interpreted as tentatively strong reliability evidence (Kalkbrenner, 2023; Taber, 2018). To evaluate the temporal reliability, we calculated test–retest reliability estimates based on measurements obtained at pre-treatment screening and at the baseline measurement two-weeks later. We used the Intraclass Correlation Coefficient (ICC_3) two-way mixed effects, single measurement model for this calculation (Koo & Li, 2016). An ICC_3 value below 0.50 is interpreted as an indicator of poor reliability, values between 0.50 and 0.75 indicate moderate reliability, values between 0.75 and 0.90 indicate good reliability, and values greater than 0.90 indicate excellent reliability (Koo & Li, 2016; Nunnally, 1978).

Finally, criterion-related validity was evaluated by correlating total scores on the Valentine scale with total scores on another measure of relationship satisfaction (DAS), a measure of quality of life (BBQ), and measures of psychopathology; specifically, we predicted a positive correlation between the Valentine scale and DAS and the Valentine scale and BBQ, and a negative correlation with measures of psychopathology (i.e. lower scores on the GAD-7 and PHQ-9). Correlational effect sizes were primarily interpreted using Hemphill's (2003) interpretive framework (but see also Rönkkö & Cho, 2022).

Results

First, an assessment of the suitability of the data for factor analysis was conducted. The Kaiser-Meyer-Olkin's (KMO) factor adequacy was .81, and Bartlett's test of homogeneity of variances was $p < .001$, thus confirming the suitability of the data for factor analysis. Thereafter, a covariance matrix for all items on the Valentine scale at screening was computed (see Figure 1). In total, 50 participants did not respond to the Valentine scale at pre-treatment screening. Thus, the correlation matrix was computed using only complete cases ($n = 851$).

To evaluate the number of factors that are appropriate to derive from the Valentine scale, a parallel analysis was performed (see Figure 2). Despite a two-factor solution technically being preferred based on the parallel analysis, the scree plot's "elbow" suggested a single-factor solution. Moreover, only the one-factor solution satisfied the Kaiser criterium. Furthermore, since the Valentine scale was explicitly developed to match a singular relationship satisfaction construct (Burman et al., 2018), we proceeded with a single-factor solution.

Next, an exploratory principal axis factor analysis using the weighted least squares (WLS) factoring method was conducted. The factor solution was rotated with an orthogonal (varimax) factor rotation due to the abovementioned single-factor solution for the Valentine scale. Standardized factor loadings and internal consistency reliability estimates are displayed in Table 2.

As Table 2 displays, all factor loadings were at least acceptable in magnitude (cf. Howard, 2016). Specifically, Items 1 and 3 had a factor loading that was just below the threshold for good ($\geq .40$), while all other items had factor loadings that were above the threshold for good (i.e. $> .40$). Thus, these results provide preliminary support for a single-factor solution for the Valentine scale. However, it should be noted that the factor loading magnitude for Items 1–3 deviates substantially from that of Items 4–7.
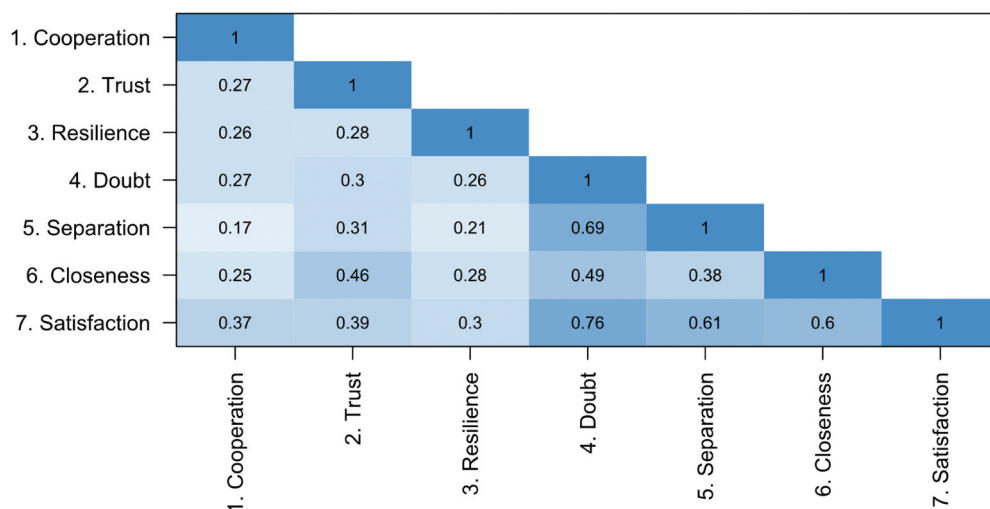

Figure 1. A Spearman polychoric correlation matrix for every item pair on the Valentine scale; darker colors indicate a larger correlation. Each item has been labelled with a descriptive name that aims to capture the essence of what it is intended to measure.
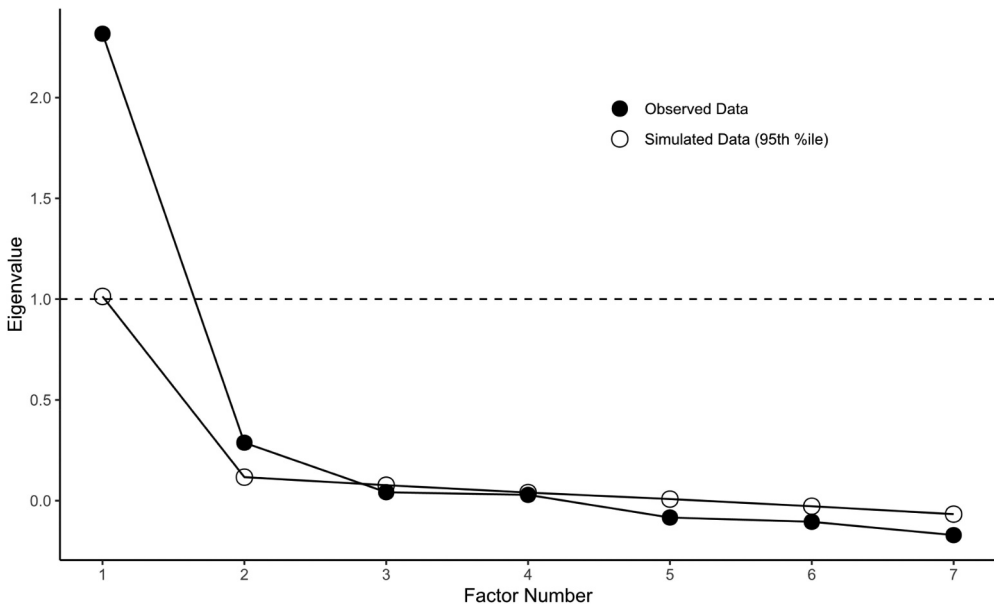
Figure 2. Scree plot and parallel analysis for the Valentine scale.

Table 2. Exploratory factor analysis of the Valentine scale at pre-treatment screening.

Item Number	Items	Standardized Factor Loading	h^2
Item 1	I can collaborate well and solve practical problems with my partner.	.39	.15
Item 2	I feel that I can confide in my partner.	.49	.24
Item 3	When my partner and I disagree or have a conflict, I quickly get over it.	.38	.15
Item 4	How often in the recent past have you thought that your relationship is not good?	.82	.68
Item 5	How often have you recently considered separating from your partner?	.69	.47
Item 6	How emotionally close do you feel to your partner?	.65	.43
Item 7	Overall, how satisfied are you with your relationship?	.89	.80
	Eigenvalue	2.91	
	Percentage of variance explained	41.6%	
	Reliability: Cronbach's alpha	.75 [95% CI: .73, .77]	
	Reliability: McDonald's omega*	.76 [95% CI: .68, .85]	

Principal factor analysis using weighted least squares (WLS) factoring and a varimax rotation. Standardized loadings (pattern matrix) based upon the polychoric correlation matrix. h^2 = communality. Items 1–5 are rated on a 4-point scale from "Always" to "Never"; Item 6 is rated on a 4-point scale from "Very close" to "Not at all close"; Item 7 is rated on a 4-point scale from "Very satisfied" to "Very dissatisfied".
*Simulated confidence interval from 100 iterations.

This may be related to the two-factor solution suggested by the parallel analysis reported above (i.e. Items 1–3 May signify a differentiable factor from Items 4–7).

Reliability analysis

The reliability analysis for the Valentine scale yielded acceptable evidence for its reliability. Specifically, in terms of internal consistency reliability, Cronbach's alpha was equal to .75 and McDonald's omega was equal to .76, which both fall under our definition of tentatively acceptable reliability (Kalkbrenner, 2023). Regarding temporal reliability, the

ICC_3 test–retest reliability coefficient was equal to .80 [95% CI: .78, .82], indicating that the Valentine scale possesses good test-retest reliability between pre-treatment screening and the baseline measurement two-weeks later (Koo & Li, 2016). Taken together, our reliability analyses indicate that the Valentine scale's items are acceptably homogenous and reliable over time.

Criterion-related validity

A correlation calculation assessing the bivariate correlation between total scores on the Valentine scale and DAS revealed a significant positive linear association ($r = .73$, $p < .001$ [95% CI: .72, .76]), where increased relationship satisfaction measured by the Valentine scale was associated with increased relationship satisfaction measured by the DAS. This correlational effect size is large in magnitude (Hemphill, 2003). Moreover, the factor correlation between the Valentine scale and DAS total score was $r = .87$ [95% CI: .84, .90], in turn establishing excellent evidence for concurrent-convergent validity for the Valentine scale (cf. Rönkkö & Cho, 2022). Table 3 displays other correlational coefficients relevant to criterion-related validity.

The correlational effect between relationship satisfaction measured by the Valentine scale and quality of life as measured by the BBQ was large (Hemphill, 2003), showing that relationship satisfaction is positively associated with quality of life. Calculating the factor correlation between the Valentine scale and BBQ yielded an isomorphic result to the bivariate correlation reported in Table 3; $r = .37$ [95% CI: .30, .45]; in turn corroborating the abovementioned concurrent-convergent validity of the Valentine scale (cf. Rönkkö & Cho, 2022). Moreover, relationship satisfaction was negatively associated with measures of psychiatric disorders; the correlational effect size between relationship satisfaction and anxiety symptom severity was moderate, and the correlational effect size between relationship satisfaction and depressive symptom severity was large (Hemphill, 2003). The factor correlation between the Valentine scale and GAD-7 was $r = -.24$ [95% CI: -.32, -.16] and between the Valentine scale and PHQ-9 was $r = -.39$ [95% CI: -.46, -.32], which further corroborates the concurrent-discriminant validity of the Valentine scale.

Taken together, the Valentine scale was significantly associated with another measure of relationship satisfaction (i.e. the DAS) and showed concurrent-convergent validity with established measures (cf. Rönkkö & Cho, 2022). Moreover, relationship satisfaction was significantly associated with quality of life which further establishes its convergent validity (Dugal et al., 2018; Soulsby & Bennett, 2015). Finally, relationship satisfaction as measured by the Valentine scale was negatively associated with measures of negative

Table 3. Correlations between the Valentine scale and relevant measures.

Measurement Compared with the Valentine Scale	Pearson's r	95% CI	
		Lower	Upper
DAS	.73	.72	.76
BBQ	.33	.27	.39
GAD-7	−.24	−.32	−.17
PHQ-9	−.37	−.37	−.25

All correlations were significant at the $p < .001$ level. Acronyms: DAS = Dyadic Adjustment Scale; BBQ = Brunnsviken Brief Quality of Life Inventory; GAD-7 = Generalized Anxiety Disorder Scale-7; PHQ-9 = Patient Health Questionnaire 9-Item Scale.

affectivity (cf. psychiatric disorder symptom severity), which substantiates its concurrent-discriminant validity (Miller et al., 2013; Robles, 2014; Whisman, 2013). These results provide preliminary evidence for the criterion-related validity of the Valentine scale.

Discussion

In Study One, our aim was to evaluate the factor structure of the Valentine scale, investigate its internal and temporal reliability, and collect information on convergent and discriminant validity. The results indicate that a single-factor solution is appropriate for the Valentine scale, a finding supported by both the scale's theoretical framework (Burman et al., 2018) and the scree plot. However, it should be noted that the parallel analysis suggested that a two-factor solution might be a better fit for the data. This may explain the deviation in the factor loading magnitude for Items 1–3 compared to that of Items 4–7. That said, the first factor had an eigenvalue of 2.32 and the second factor and an eigenvalue of 0.29, in turn suggesting that the first factor explains the majority of the variance in the Valentine scale while the second factor explains less than one item in the scale (DeVellis & Thorpe, 2021). Furthermore, a two-factor solution does not align with the theoretical underpinnings of the Valentine scale's conception. Blind data-driven decisions are undesirable when conducting factor analyses (Fried, 2020), and the difference in factor loading magnitude may also relate to conceptual redundancy (cf. Cortina et al., 2020), indicating that the Valentine scale consists of both general and specific items that measure relationship satisfaction. This is further supported by the polychoric correlation matrix, as the scale items did not display extraordinary levels of correlation with one-another (i.e. correlations ranged between $r = .17$ and $.76$).

The Valentine scale emerged as a reliable instrument of relationship satisfaction, with tentatively acceptable internal consistency reliability and good temporal reliability. Our reliability analyses indicated that the Valentine scale consists of homogenous items, reliable over repeated administrations. Finally, concurrent-convergent validity was established for the Valentine scale via its strong positive association with a previously established measure of relationship satisfaction (i.e. the DAS; cf. Rönkkö & Cho, 2022). Auxiliary criterion-related validity evidence for the Valentine scale was also established through its positive association with a measure of quality of life and negative associations with measures of psychiatric disorders (i.e. anxiety and depression). Taken together, this study provides preliminary evidence supporting the reliability and unidimensionality of the Valentine scale, corroborating its single-factor construct.

This study has limitations. Firstly, 50 out of 901 participants did not respond to the Valentine scale at screening. This issue likely arises from individuals starting the screening questions but losing interest before completing them. However, given the already substantial sample size, it is unlikely that this issue will have a significant impact on the current psychometric evaluation analysis. Second, our use of the total score of the DAS to validate the Valentine scale may be suboptimal considering the contested use of the DAS (Graham et al., 2006; Norton, 1983; Vajda et al., 2019). Nonetheless, the DAS total score remains commonly used (Bühler et al., 2021; The CORE Lab, 2024) and thus provides us with an adequate idea about the convergent validity of the Valentine scale. Finally, this study uses data from a couple's therapy sample which may not be representative of the

general population. However, given the intended goal of the Valentine scale (i.e. index relationship satisfaction, particularly within therapeutic settings), this limitation is likely trivial. This does, however, limit the generalizability of the findings to a general population. The strengths of this study include its large sample size, use of various measures to establish criterion-related validity, and design that allowed for the evaluation of temporal reliability.

Future studies must confirm the factor structure of the Valentine scale and replicate the criterion-related validity findings. Importantly, the measurement invariance of the Valentine scale must be established, particularly across gender and depressive symptoms due to the differential ratings of relationship satisfaction that are moderated by these variables. For instance, a wife's marital satisfaction is predictive of their husband's depressive symptomology, but not vice versa (Meyer et al., 2019; Miller et al., 2013).

Study two: ADHD couple's therapy sample

In Study Two, our goal was to collect further validity data for the Valentine scale's intended use. As such, we only present information from Study Two that is relevant for the present psychometric evaluation here. Detailed information about the original study is presented elsewhere (Aarikka & Chlaidze, 2024). Study Two aimed to validate the factor structure of the Valentine scale reported in Study One. Moreover, we aimed to evaluate the measurement invariance of the Valentine scale across gender and depressive symptom severity, given the differential ratings of relationship satisfaction between males and females and depressed/non-depressed individuals in a relationship dyad (Meyer et al., 2019; Miller et al., 2013). Additionally, we aimed to corroborate the validity findings found in Study One by further examining concurrent-convergent validity with related scales (i.e. the QDR and BBQ), as well as concurrent-discriminant validity with measures of psychiatric disorders (i.e. the GAD-7 and PHQ-9).

Method

Sample characteristics

Of the 678 potentially interested participants, 527 completed the pre-treatment screening and responded to the Valentine scale; descriptive statistics for these participants are presented in Table 4.

Study design

Participants were primarily recruited through social media, with additional advertising via flyers displayed at outpatient facilities across Sweden. The study had specific inclusion and exclusion criteria for participating couples. Inclusion criteria required both partners to be willing participants, at least 20 years old, fluent in Swedish, cohabiting for a minimum of 6 months within a relationship of at least 1 year's duration. Additionally, at least one partner needed an ADHD diagnosis, and the couple had to reside within an hour's travel from central Stockholm. If medication was involved, dosages needed to have been stable for at least 3-months prior to the onset of

Table 4. Descriptive statistics for the ADHD couple's therapy sample.

Age: M (SD) [min-max] in Years	40.13 (7.71) [18.00–63.00]
Gender: N (%)	
Female	296 (56%)
Male	230 (44%)
Other	1 (0.2%)
Highest Level of Education Completed: N (%)	
Primary School	23 (4.4%)
High School	171 (32%)
Higher Education ≤3 years	153 (29%)
Higher Education >3 years	180 (34%)
Occupation: N (%)	
Student	43 (8.2%)
Seeking Work	18 (3.4%)
Retired	1 (0.2%)
Parent Leave	13 (2.5%)
Sick Leave	11 (2.1%)
Working	441 (84%)
Measures: M (SD) [min-max]	
Valentine Scale	13.81 (3.43) [3.00–21.00]
QDR	20.71 (3.46) [9.51–29.33]
GAD-7	7.16 (5.14) [0.00–21.00]
PHQ-9	7.02 (5.36) [0.00–26.00]
BBQ	53.49 (20.02) [1.00–96.00]

Numerical variables are reported as mean (SD) [range: minimum-maximum]. Categorical variables are reported as count (percentage). Acronyms: QDR = Quality of Dyadic Relationships; GAD-7 = Generalized Anxiety Disorder Scale 7-item; PHQ-9 = Patient Health Questionnaire 9-item; BBQ = Brunnsviken Brief Quality of Life Inventory.

participation. Exclusion criteria encompassed relationship abuse, severe depression (PHQ-9 score > 19), suicidality, ongoing psychological treatments, or the presence of more severe psychiatric conditions requiring prioritized treatment, such as psychosis, schizophrenia, or bipolar spectrum disorder. The present study exclusively leverages data from the pre-treatment screening phase of the larger study; detailed information about other components of the original study is provided elsewhere (Aarikka & Chlaidze, 2024).

Measures

The Valentine scale
The Valentine scale has already been described in detail in Study One. In Study Two, the internal consistency reliability for the Valentine scale at screening was good, Cronbach's α = .81 [95% CI: .78, .83] and McDonald's ω = .81 [95% CI: .80, .94].

Quality of dyadic relationships (QDR)
The QDR is a 36-item self-report questionnaire that measures overall relationship satisfaction (Ahlborg et al., 2005). The QDR was developed as a response to the problems inherent with the DAS by revising the original DAS to reflect a broader scope of relationship quality (see Ahlborg et al., 2005). The QDR produces a total score but can also be broken down into five subscales: Cohesion, consensus, satisfaction, sensuality, and sexuality. The QDR has undergone preliminary construct validation and been subjected to a reliability analysis (Gudjonsdottir et al., 2021). However, it is possible that the DQR suffers from similar problems as the DAS (cf. The CORE Lab, 2024). In this

sample, the internal consistency reliability for the QDR total score was excellent, Cronbach's α = .93 [95% CI: .92, .94].

Generalized anxiety disorder scale-7 (GAD-7)
The GAD-7 (Spitzer et al., 2006) was described in detail in Study One. In Study Two, the internal consistency reliability for the GAD-7 was excellent, Cronbach's α = .88 [95% CI: .86, .89].

Patient health questionnaire 9-item scale (PHQ-9)
The PHQ-9 (Kroenke et al., 2001) was described in detail in Study One. In Study Two, the internal consistency reliability for the PHQ-9 was excellent, Cronbach's α = .85 [95% CI: .83, .87].

Brunnsviken brief quality of life inventory (BBQ)
The BBQ (Lindner et al., 2016) was described in detail in Study One. In Study Two, the internal consistency reliability for the BBQ was excellent, Cronbach's α = .79 [95% CI: .76, .82].

Data analysis

The suitability of the data for factor analysis was evaluated using the Kaiser-Meyer-Olkin's (KMO) factor adequacy and Bartlett's test of homogeneity of variances. A KMO value above .80 and a significant Bartlett's test of homogeneity of variances were taken as an indicator of the suitability of data for factor analysis (DeVellis & Thorpe, 2021; Kaiser, 1970). Intercorrelations between items were evaluated using a polychoric correlation matrix, the preferred method when the data is ordinal (Olsson, 1979). To evaluate the stability of the factor structure of the Valentine scale obtained in Study One, a confirmatory factor analysis (CFA) model was fit to the data using the *lavaan* package; version 0.6–16 (Rosseel, 2012). The factor model was generated using a diagonally weighted least squares (DWLS) estimator, as it is robust against violations of normality making it suitable for ordinal data (Li, 2016). CFA models were plotted using the *semPlots* package (Epskamp, 2022). Factor loadings were standardized on the latent variable to allow factor loadings to be expressed as standard deviations of the latent factor (Brown, 2012, 2015).

Model fit indices were assessed using the χ^2 test statistic together with auxiliary model fit indices as the χ^2 test statistic's tendency to overemphasize minor and insignificant departures from a theoretically perfect model fit (Putnick & Bornstein, 2016). Thus, the Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR) were also used to evaluate model fit. Hu and Bentler (1999) suggested that a cutoff value close to .95 on the CFI, a cutoff value close to .06 for RMSEA, and a cutoff value close to .08 for SRMR results in lower Type II error rates (with acceptable Type I error rates). Accordingly, a suboptimal model fit was defined as a CFI value above .90, SRMR values below .10, and RMSEA values between .08 and .10; adequate model fit was defined as a CFI value between .92 and .95, SRMR values between .08 and .10, and a RMSEA value below .08; and good model fit was defined as a CFI value higher

than .92, SRMR values below .08, and RMSEA value below .05. Additionally, model fit was deemed excellent if it exhibited all characteristics of a good model fit but also yielded a non-significant χ^2 test statistic. A separate analysis wherein the dynamic cutoff framework (McNeish, 2023) was used to evaluate fit indices was also conducted (see Supplementary Material on OSF, https://osf.io/fb72s). Finally, the modification indices were used to evaluate potential changes to the model that could enhance the χ^2 fit. The modification index represents the χ^2 value, with 1 degree of freedom, by which model fit would improve if a particular path was added or constraint freed. A modification index value of 3.84 or higher indicates a significant improvement in the χ^2 statistic, corresponding to a significant change at a one-sided p-value of .05 (Sörbom, 1989).

Measurement invariance (MI) analyses were performed to evaluate the stability of the Valentine scale across different groups. Model parameters were successively constrained to assess configural, metric, scalar, and residual invariance factor structures across groups in cases when it was appropriate (Gregorich, 2006; Kline, 2023; van de Schoot et al., 2012). A poor model fit to any of these models indicates that the constrained parameter differentially operates across groups. Measurement invariance across groups in each model was determined when $\Delta\chi^2$ $p > .05$, ΔCFI < .01, ΔRMSEA < .015, and ΔSRMR < .015 (Bikos, 2022; Putnick & Bornstein, 2016). Finally, the data were reanalyzed by implementing a subsampling approach to examine MI across clinical-subclinical levels of depression, as these groups were unbalanced in size. This was implemented to substantiate the results and involved randomly selecting subsets from the larger group to match the size of the smaller group and conducting invariance testing across 100 replications (Yoon & Lai, 2018).

Criterion-related validity was evaluated by correlating total scores on the Valentine scale with total scores on another measure of relationship satisfaction (QDR), a measure of quality of life (BBQ; cf. convergent and concurrent validity), and measures of psychopathology; specifically, we predicted a positive correlation between the Valentine scale and DAS and the Valentine scale and BBQ, and a negative correlation with measures of psychopathology (i.e. lower scores on the GAD-7 and PHQ-9). Correlational effect sizes were primarily interpreted using Hemphill's (2003) interpretive framework (but see also Rönkkö & Cho, 2022).

Results

Confirmatory factor analysis

First, an assessment of the suitability of the data in Study Two for factor analysis was conducted. The Kaiser-Meyer-Olkin's (KMO) factor adequacy was .85 and Bartlett's test of homogeneity of variances was $p < .001$, thereby confirming the suitability of the data for factor analysis. Thereafter, a polychoric correlation matrix was produced by computing the Spearman's rho rank order correlation coefficient to provide a visual representation of the intercorrelations between the items on the Valentine scale (see Figure 3).

As Figure 3 shows, the intercorrelations between Item 3 and other items were noticeably weaker than between other items, suggesting that Item 3 May not align as
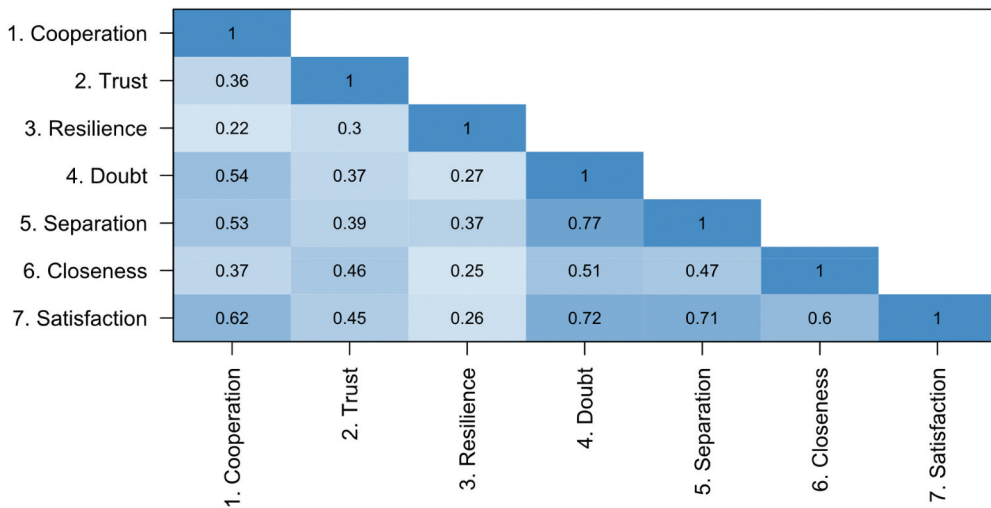
Figure 3. A Spearman polychoric correlation matrix for every item pair on the Valentine scale; darker colors indicate a larger correlation. Each item has been labelled with a descriptive name that aims to capture the essence of what it is intended to measure.

closely with the underlying construct of relationship satisfaction as do other items on the scale.

To evaluate whether the single-factor solution obtained in Study One fit the data, a confirmatory factor analysis was conducted. All items on the Valentine scale were fit to a single factor (see Figure 4).

A single-factor solution for the Valentine scale resulted in a good fit for the data: $\chi2$ (14) = 68.53, p = 0.000, CFI = 0.938, SRMR = 0.049, RMSEA = 0.120 [90% CI: 0.091, 0.150]. Thus, our single-factor model for the Valentine scale exhibits our predetermined criteria for an excellent model fit in all cases except for RMSEA, which was problematic. An evaluation of the modification indices revealed two potential changes that could be made to the model to improve model fit: adding a covariance between Item 2 and Item 6 (MI = 10.82) and between Item 4 and Item 5 (MI = 16.29). See Table 5 for the standardized CFA factor solution.

As Table 5 shows, Item 3 has the lowest factor loading, which further suggests that it may not align as closely with the underlying construct of relationship satisfaction as do other items on the scale.

Measurement invariance (MI)

Measurement invariance of gender

The sample was stratified by gender (male: n = 230; female: n = 296). As only one participant identified as non-binary, we did not include the person in the gender analyses. Thereafter, all items on the Valentine scale were fit to one factor, grouped by gender. The configural model, which constrained only the relative configuration of item data to be equal across genders, had an excellent fit to the data: $\chi^2(28)$ =47.69, p = .012, CFI = .996, SRMR = .052, RMSEA = .052. The metric invariance model constrained the
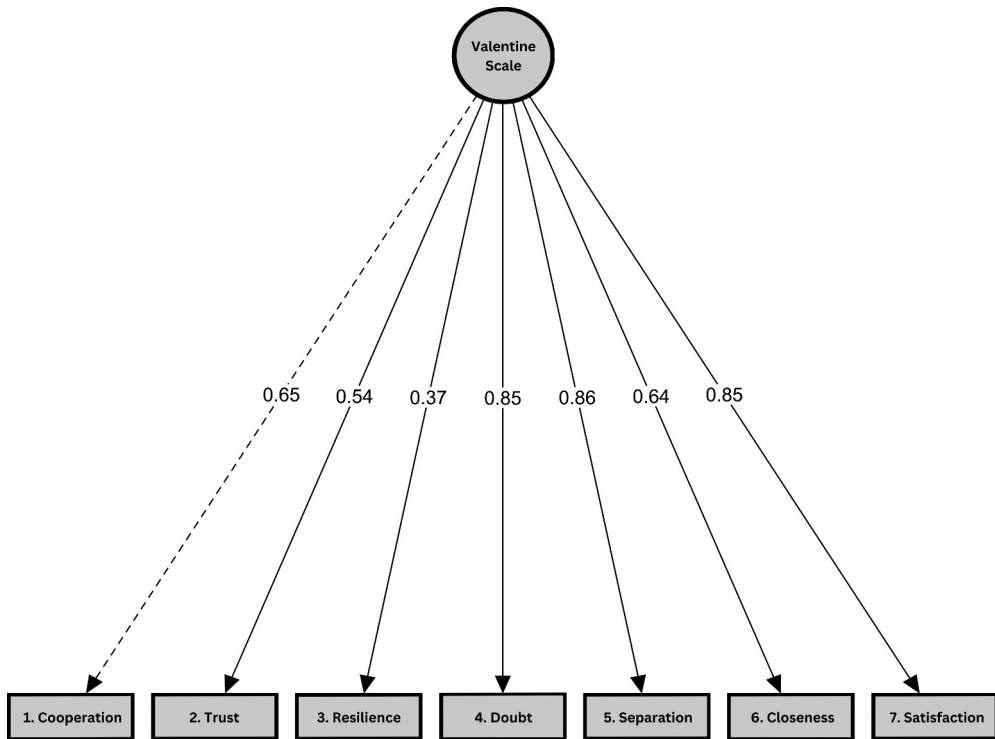
Figure 4. Structural equation model path diagram with standardized factor loadings for the Valentine scale.

Table 5. Confirmatory factor analysis for the Valentine scale.

Item Number	Items	Standardized Factor Loadings	Explained variance (R²)
Item 1	I can collaborate well and solve practical problems with my partner.	.65	.42
Item 2	I feel that I can confide in my partner.	.54	.28
Item 3	When my partner and I disagree or have a conflict, I quickly get over it.	.37	.14
Item 4	How often in the recent past have you thought that your relationship is not good?	.86	.73
Item 5	How often have you recently considered separating from your partner?	.86	.74
Item 6	How emotionally close do you feel to your partner?	.64	.42
Item 7	Overall, how satisfied are you with your relationship?	.86	.73

configuration of item data and factor loadings to be constant across genders. Fit indices were comparable to the configural model and resulted in an excellent fit: $\chi^2(34) = 60.54$, $p = .003$, CFI = .995, SRMR = .058, RMSEA = .055. Moreover, metric invariance was supported by non-significant difference tests that evaluated configural and metric model similarity: $\Delta\chi^2(6) = 12.8$, $p = .046$; ΔCFI = -.001; ΔSRMR = .006; ΔRMSEA = .003.

In the scalar invariance model, configuration of items, loadings, and indicator means/intercepts were constrained to be equal across genders. Fit indices were comparable to the metric model resulting in an excellent fit: $\chi^2(47) = 64.32$, $p = .047$, CFI = .997, SRMR =

.054, RMSEA = .037. However, scalar invariance was unsupported by the large ΔRMSEA value in the evaluation between metric and scalar model similarity: $\Delta\chi^2(13) = 3.78$, $p = .99$; ΔCFI =.002; ΔSRMR = -.004; ΔRMSEA = -.017. Taken together thus, only metric MI was established for the Valentine scale across genders but strong indication for scalar invariance was found (cf. only ΔRMSEA was problematic).

Measurement invariance of clinical-subclinical levels of depression

Participants were split into groups according to their scores on the PHQ-9, representing above ($n = 147$) and below ($n = 380$) threshold for clinical depression (i.e., PHQ-9 ≥ 10). Thereafter, Valentine scale item data were fit to one factor, grouped by clinical–subclinical levels of depression. The configural model, which was constrained only by the relative configuration of item data to be equal across clinical–subclinical levels of depression, had an excellent fit to the data: $\chi^2(28) = 51.36$, $p = .005$, CFI = .996, SRMR = .052, RMSEA = .056. A resampling approach that corrected for unbalanced groups yielded the same fit indices (CFI = .996, SRMR = .052, RMSEA = .056), in turn corroborating configural invariance across clinical–subclinical levels of depression. The metric invariance model constrained the configuration of item data and factor loadings to be constant across clinical–subclinical levels of depression. Fit indices were comparable to the configural model and resulted in an excellent fit: $\chi^2(34) = 78.43$, $p = .000$, CFI =.992, SRMR = .066, RMSEA = .071.

A resampling approach that corrected for unbalanced groups yielded the same fit indices (CFI = .992, SRMR = .066, RMSEA = .071), in turn corroborating metric invariance across clinical–subclinical levels of depression. However, metric invariance was not supported by difference tests that evaluated configural and metric model similarity: $\Delta\chi^2(6) = 27.1$, $p = .000$; ΔCFI = -.004; ΔSRMR = .014; ΔRMSEA = .014. Thus, given the statistically significant $\Delta\chi^2$ test, we cannot claim metric invariance across clinical–subclinical levels of depression and therefore do not test stricter models. Taken together, we only found support for the configural model, which constrained only the relative configuration of items in the model to be the same across clinical–subclinical levels of depression.

Criterion-related validity

A correlation test assessing the bivariate correlation between total scores on the Valentine scale and QDR revealed a significant positive linear association ($r = .71$, $p < .001$ [95% CI: .67, .75]), where increased relationship satisfaction measured by the Valentine scale was associated with increased relationship satisfaction measured by the QDR. This correlational effect size is large in magnitude (Hemphill, 2003). Moreover, the factor correlation between the Valentine scale and DAS total score was $r = .91$, [95% CI: .88, .94], in turn establishing excellent evidence for concurrent-convergent validity for the Valentine scale (cf. Rönkkö & Cho, 2022). Table 6 displays other correlation coefficients relevant to criterion-related validity.

The correlational effect size between relationship satisfaction measured by the Valentine scale and quality of life as measured by the BBQ is large (Hemphill, 2003), wherein relationship satisfaction is positively associated with quality of life. Evaluating the factor correlations between the Valentine scale and BBQ yielded an isomorphic result to the bivariate correlation reported in Table 6; $r = .36$ [95% CI: .27, .45]; in turn

Table 6. Correlations between the Valentine scale and relevant measures.

Measurement Compared with the Valentine Scale	Pearson's r	95% CI	
		Lower	Upper
QDR	.71	.67	.75
BBQ	.33	.25	.40
GAD-7	−.25	−.33	−.17
PHQ-9	−.24	−.32	−.16

All correlations were significant at the p < .001 level. Acronyms: QDR = Quality of Dyadic Relationships; BBQ = Brunnsviken Brief Quality of Life Inventory; GAD-7 = Generalized Anxiety Disorder Scale-7; PHQ-9 = Patient Health Questionnaire 9-Item Scale.

corroborating the abovementioned concurrent-convergent validity of the Valentine scale (cf. Rönkkö & Cho, 2022). Moreover, relationship satisfaction is negatively associated with measures of psychiatric disorders; the correlational effect sizes between relationship satisfaction and anxiety symptom severity and relationship satisfaction and depressive symptom severity were large (Hemphill, 2003). The factor correlation between the Valentine scale and GAD-7 was $r = -.28$ [95% CI: −.37, −.19] and between the Valentine scale and PHQ-9 was $r = -.27$ [95% CI: −.36, −.18], which further corroborates the concurrent-discriminant validity of the Valentine scale.

Discussion

Study Two primarily aimed to confirm the single-factor structure for the Valentine scale obtained in Study One, while also conducting a preliminary analysis of measurement invariance across gender and depressive symptom severity, given the differential ratings of relationship satisfaction between males and females and depressed/non-depressed individuals in a relationship dyad (Meyer et al., 2019; Miller et al., 2013). We found strong support for a unidimensional factor structure for the Valentine scale, thereby confirming its single-factor structure. However, two significant modification indices were identified, suggesting that adding a covariance between Item 2 and Item 6 and between Item 4 and Item 5 would improve model fit if they were added to the model. Finally, only Item 3 resulted in a weak factor loading, contrasting with the results from Study One, further suggesting that the scale is conceptually redundant (as opposed to grammatically redundant; cf. Cortina et al., 2020) and providing further support for the notion that the Valentine scale consists of both general and specific items that measure relationship satisfaction.

We found strong support for measurement invariance across gender at the scalar level of analysis (i.e., scalar measurement invariance). These findings substantiate the validity of the single-factor structure of the Valentine scale and contest previous findings that have suggested gender differences in relationship satisfaction ratings (Meyer et al., 2019; Miller et al., 2013). However, we only found support for configural measurement invariance across clinical—subclinical levels of depression, indicating that the same factors are manifested in each group, albeit in potentially different ways. These differences may be due to unequal pattern coefficients, unequal intercepts, or unequal error variances across clinical—subclinical levels of depression for the Valentine scale (Bikos, 2022). Thus, total scores on the Valentine scale must be interpreted with caution when it is used in studies with a high prevalence of depressive

symptoms. Although only weak support for measurement invariance across clinical—subclinical levels of depression could be established, this aligns with previous findings that suggest depression symptom severity moderate ratings of relationship satisfaction among individuals in a relationship dyad (Meyer et al., 2019; Miller et al., 2013).

We also aimed to corroborate the validity evidence from Study One by further examining concurrent-convergent validity with related scales (i.e. the QDR and BBQ), as well as concurrent-discriminant validity with measures of psychiatric disorders (i.e. the GAD-7 and PHQ-9). The results showed that the Valentine scale was significantly associated with another measure of relationship satisfaction (i.e. the QDR) and has an appropriately high factor correlation to establish concurrent-convergent validity (cf. Rönkkö & Cho, 2022). Moreover, relationship satisfaction was significantly associated with quality of life which further establishes its convergent validity (Dugal et al., 2018; Soulsby & Bennett, 2015). Finally, relationship satisfaction as measured by the Valentine scale was negatively associated with measures of negative affectivity (cf. psychiatric disorder symptom severity), which substantiates its concurrent-discriminant validity (Miller et al., 2013; Robles, 2014; Whisman, 2013). These results replicate the results obtained in Study One, in turn further substantiating the criterion-related validity of the Valentine scale.

This second study has limitations. Similar to Study One, our use of the total score of the QDR to validate the Valentine scale may be suboptimal considering that it is based on the DAS (Ahlborg et al., 2005), which has been heavily contested (Graham et al., 2006; Norton, 1983; Vajda et al., 2019). Nonetheless, the QDR total score is often used (Ahlborg et al., 2005; The CORE Lab, 2024; Gudjonsdottir et al., 2021) and may thus provide us with information about the convergent validity of the Valentine scale. Additionally, this study used data from a couple's therapy sample which may not be representative for the general population. However, given the intended goal of the Valentine scale (i.e. index relationship satisfaction, particularly within therapeutic settings), this limitation is likely trivial. The strengths of this study include its large sample size, no missing data at pre-treatment screening, and use of various measures to establish criterion-related validity.

General discussion

The Valentine scale is a freely available, unidimensional measure of relationship satisfaction. Across two studies, we found support for a single-factor structure for the Valentine scale, support for its high internal and temporal reliability, and appropriate validity evidence for convergent and discriminant validity evidence (DeVellis & Thorpe, 2021; Rönkkö & Cho, 2022). Specifically, the Valentine scale is highly correlated with other measures of relationship satisfaction which establishes its convergent validity. The operational characteristics of the Valentine scale (cf. measurement invariance) are isomorphic across genders. Furthermore, we found weak support for equivalency of the Valentine scale across clinical—subclinical levels of depression, indicating that the same factors are manifested in each group, albeit in potentially different ways. Although this aligns with previous findings that suggest depression symptom severity moderate ratings of relationship satisfaction among individuals in a relationship dyad (Meyer et al., 2019; Miller et al., 2013), total scores on the Valentine scale must be interpreted with caution when it is used in studies with a high prevalence of depressive symptoms. Finally, the Valentine scale is positively associated with a measure of quality of life and negatively correlated with measures of psychiatric disorders. Furthermore, criterion-related validity

evidence was consistently found in both Study One and Study Two, in turn corroborating the findings.

To our knowledge, the Valentine scale is the first freely available, unidimensional measure of relationship satisfaction that has been subjected to as stringent criteria for validation as detailed in this paper (for a review, see The CORE Lab, 2024). However, further psychometric evaluation is warranted to comprehensively understand the Valentine's scale performance as the studies presented in this paper have limitations. Chiefly, both Study One and Two consist of treatment-seeking participants in a relationship dyad which limits the generalizability of the findings to the general population. Other limitations include the use of contested relationship satisfaction instruments to validate the Valentine scale (Graham et al., 2006; Norton, 1983; Vajda et al., 2019). Nevertheless, the DAS and QDR remain commonly used (Bühler et al., 2021; The CORE Lab, 2024) and thus can inform the convergent validity of the Valentine scale.

Future studies should continue to validate the Valentine scale. For instance, the modification indices revealed in Study Two warrant further investigation. Specifically, the modification indices suggested that adding a covariance between Item 2 and Item 6 and between Item 4 and Item 5 would improve model fit. Employing a Mokken scale analyses would inform scalability of the Valentine scale, as well as provide information about item discrimination, difficulty, and potential guessing effects (Sijtsma & Molenaar, 2002). Such analyses would also help to elucidate the evident problems with Item 3 found in both Study One and Two.

In summary, our studies indicate that the Valentine scale is a reliable unidimensional measure of relationship satisfaction. Our unidimensional factor structure greatly improves the interpretability of the relationship satisfaction construct. Furthermore, our findings indicate that the Valentine scale consists of both general and specific items of relationship satisfaction that are conceptually redundant, but not grammatically redundant (Cortina et al., 2020). This feature is fairly unique when compared to other scales of relationship satisfaction (The CORE Lab, 2024). Future studies should aim to evaluate its operational characteristics when used in non-clinical populations as well as when it is used to monitor in-treatment fluctuations in outcomes during couple's therapy treatment (as opposed to only pre-treatment screening and post-assessment measurements). Doing so could allow for a greater understanding of the mechanisms driving change and potential threats to treatment gainsfor instance, in studies involving individuals in a relationship dyad.

ORCID

Jón Ingi Hlynsson https://orcid.org/0009-0000-5703-1068
Ísak Örn Ívarsson https://orcid.org/0009-0002-3012-4402
Gerhard Andersson https://orcid.org/0000-0003-4753-6745
Per Carlbring https://orcid.org/0000-0002-2172-8813

Ethical statement

The studies reported in this paper were approved by the by the Swedish Ethical Review Authority. Study One was approved in 2017 (Dnr: 2017/1726-31/5) and Study Two was approved in 2023 (Dnr: 2023-04245-01).

Open science statement

The data that support the findings of this study are available at https://osf.io/fb72s

References

Aarikka, V., & Chlaidze, M. (2024). *An improved romantic relationship: A randomized controlled trial of the course 'you and I and Adhd'* [Unpublished master's thesis]. Stockholms universitet, Psykologiska institutionen.

Ahlborg, T., Persson, L.-O., & Hallberg, L. R.-M. (2005). Assessing the quality of the dyadic relationship in first-time parents: Development of a new instrument. *Journal of Family Nursing*, *11*(1), 19–37. https://doi.org/10.1177/1074840704273462

Arrindell, W. A., Boelens, W., & Lambert, H. (1983). On the psychometric properties of the Maudsley Marital Questionnaire (MMQ): Evaluation of self-ratings in distressed and 'normal' volunteer couples based on the Dutch version. *Personality & Individual Differences*, *4*(3), 293–306. https://doi.org/10.1016/0191-8869(83)90151-4

Bikos, L. H. (2022). Chapter 12: Invariance testing. In *ReCentering psych Stats: Psychometrics*. https://lhbikos.github.io/ReC_Psychometrics/Invariance.html

Björnsson, A. S. (2016). Maðurinn er félagslegt dýr. In A. S. Björnsson, G. Eydal, & K. Kristjánsdóttir (Eds.), *Af sál: Ritgerðasafn handa álfheiði steinþórsdóttur sálfræðingi* (pp. 273–313). Háskólaútgáfan. https://iris.rais.is/is/publications/ma%C3%B0urinn-er-f%C3%A9lagslegt-d%C3%BDr

Brown, T. A. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed., pp. 261–276). The Guilford Press.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Bühler, J. L., Krauss, S., & Orth, U. (2021). Development of relationship satisfaction across the life span: A systematic review and meta-analysis. *Psychological Bulletin*, *147*(10), 1012–1053. https://doi.org/10.1037/bul0000342

Burman, M., Norlander, A.-K., Carlbring, P., & Andersson, G. (2018). *Närmare varandra: Nio veckor till en starkare parrelation*. Natur & Kultur.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

The CORE Lab. (2024). A novel, network-based approach to assessing romantic-relationship quality. *Perspectives on Psychological Science*, 17456916231215248. https://doi.org/10.1177/17456916231215248

Cortina, J. M., Sheng, Z., Keener, S. K., Keeler, K. R., Grubb, L. K., Schmitt, N., Tonidandel, S., Summerville, K. M., Heggestad, E. D., & Banks, G. C. (2020). From alpha to omega and beyond! A look at the past, present, and (possible) future of psychometric soundness in the Journal of Applied Psychology. *Journal of Applied Psychology*, 105(12), 1351–1381. https://doi.org/10.1037/apl0000815

Delatorre, M. Z., & Wagner, A. (2020). Marital quality assessment: Reviewing the concept, instruments, and methods. *Marriage & Family Review*, 56(3), 193–216. https://doi.org/10.1080/01494929.2020.1712300

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.

Doss, B. D., Cicila, L. N., Georgia, E. J., Roddy, M. K., Nowlan, K. M., Benson, L. A., & Christensen, A. (2016). A randomized controlled trial of the web-based OurRelationship program: Effects on relationship and individual functioning. *Journal of Consulting & Clinical Psychology*, 84(4), 285–296. https://doi.org/10.1037/ccp0000063

Dugal, C., Bakhos, G., Bélanger, C., Godbout, N., Dugal, C., Bakhos, G., Bélanger, C., & Godbout, N. (2018). Cognitive-behavioral psychotherapy for couples: An insight into the treatment of couple hardships and struggles. In Ö. Şenormancı & G. Şenormancı (Eds.), *Cognitive behavioral therapy and clinical applications* (pp. 119–148). IntechOpen. https://doi.org/10.5772/intechopen.72104

Epskamp, S. (2022). *semPlot: Path diagrams and visual analysis of various SEM packages' output* [Manual]. https://CRAN.R-project.org/package=semPlot

Epstein, N. B., & Zheng, L. (2017). Cognitive-behavioral couple therapy. *Current Opinion in Psychology*, 13, 142–147. https://doi.org/10.1016/j.copsyc.2016.09.004

Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.

Finn, E., Morrison, T. G., & McGuire, B. E. (2018). Correlates of sexual functioning and relationship satisfaction among men and women experiencing chronic pain. *Pain Medicine*, 19(5), 942–954. https://doi.org/10.1093/pm/pnx056

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, 31(4), 271–288. https://doi.org/10.1080/1047840X.2020.1853461

Funk, J. L., & Rogge, R. D. (2007). Testing the ruler with item response theory: Increasing precision of measurement for relationship satisfaction with the couples satisfaction index. *Journal of Family Psychology: JFP: Journal of the Division of Family Psychology of the American Psychological Association (Division 43)*, 21(4), 572–583. https://doi.org/10.1037/0893-3200.21.4.572

Graham, J. M., Liu, Y. J., & Jeziorski, J. L. (2006). The dyadic adjustment scale: A reliability generalization meta-analysis. *Journal of Marriage & Family*, 68(3), 701–717. https://doi.org/10.1111/j.1741-3737.2006.00284.x

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, 44(11 Suppl 3), S78–S94. https://doi.org/10.1097/01.mlr.0000245454.12228.8f

Gudjonsdottir, J., Eklund, M., Ingberg, E., & Ström, J. O. (2021). Psychometric evaluation of the Swedish quality of dyadic relationships scale – Homogeneity and construct validity. *Scandinavian Journal of Caring Sciences*, 35(2), 468–475. https://doi.org/10.1111/scs.12858

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *The American Psychologist*, 58(1), 78–79. https://doi.org/10.1037/0003-066x.58.1.78

Hlynsson, J. I., & Carlbring, P. (2024). Diagnostic accuracy and clinical utility of the PHQ-2 and GAD-2: A comparison with long-format measures for depression and anxiety. *Frontiers in Psychology*, 15, 15. https://doi.org/10.3389/fpsyg.2024.1259997

Hlynsson, J. I., Lindner, P., Barri, B., & Carlbring, P. (2024). Exploring cutoff points and measurement invariance of the Brunnsviken brief quality of life inventory. *Frontiers in Psychology*, 14, 1305682. https://doi.org/10.3389/fpsyg.2023.1305682

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. https://doi.org/10.1007/BF02289447

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human–Computer Interaction*, *32*(1), 51–62. https://doi.org/10.1080/10447318.2015.1087664

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

Jacobson, N. S., Christensen, A., Prince, S. E., Cordova, J., & Eldridge, K. (2000). Integrative behavioral couple therapy: An acceptance-based, promising new treatment for couple discord. *Journal of Consulting & Clinical Psychology*, *68*(2), 351–355. https://doi.org/10.1037/0022-006X.68.2.351

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401–415. https://doi.org/10.1007/BF02291817

Kalkbrenner, M. T. (2023). Alpha, omega, and H internal consistency reliability estimates: Reviewing these options and when to use them. *Counseling Outcome Research and Evaluation*, *14*(1), 77–88. https://doi.org/10.1080/21501378.2021.1940118

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Löwe, B. (2010). The patient health questionnaire somatic, anxiety, and depressive symptom scales: A systematic review. *General Hospital Psychiatry*, *32*(4), 345–359. https://doi.org/10.1016/j.genhosppsych.2010.03.006

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

Lindner, P., Frykheden, O., Forsström, D., Andersson, E., Ljótsson, B., Hedman, E., Andersson, G., & Carlbring, P. (2016). The Brunnsviken brief quality of life scale (BBQ): Development and psychometric evaluation. *Cognitive Behaviour Therapy*, *45*(3), 182–195. https://doi.org/10.1080/16506073.2016.1143526

Luo, Z., Li, Y., Hou, Y., Zhang, H., Liu, X., Qian, X., Jiang, J., Wang, Y., Liu, X., Dong, X., Qiao, D., Wang, F., & Wang, C. (2019). Adaptation of the two-item generalized anxiety disorder scale (GAD-2) to Chinese rural population: A validation study and meta-analysis. *General Hospital Psychiatry*, *60*, 50–56. https://doi.org/10.1016/j.genhosppsych.2019.07.008

Malivoire, B. L., Rowa, K., Milosevic, I., & McCabe, R. E. (2024). Partner symptom accommodation in generalized anxiety disorder: A preliminary examination of correlates with symptoms and cognitive behavioural therapy outcome. *Behavioural and Cognitive Psychotherapy*, *52*(5), 478–494. https://doi.org/10.1017/S1352465824000213

Martin-Key, N. A., Spadaro, B., Funnell, E., Barker, E. J., Schei, T. S., Tomasik, J., & Bahn, S. (2022). The current state and validity of digital assessment tools for psychiatry: Systematic review. *JMIR Mental Health*, *9*(3), e32824. https://doi.org/10.2196/32824

McNeish, D. (2023). Dynamic fit index cutoffs for categorical factor analysis with Likert-type, ordinal, or binary responses. *American Psychologist*, *78*(9), 1061–1075. https://doi.org/10.1037/amp0001213

Meyer, D., Kemper-Damm, B., Parola, F., & Salas, J. (2019). Depressive symptoms as a predictor of men's relationship satisfaction. *Family Journal*, *27*(1), 37–43. https://doi.org/10.1177/1066480718809058

Miller, R. B., Mason, T. M., Canlas, J. M., Wang, D., Nelson, D. A., & Hart, C. H. (2013). Marital satisfaction and depressive symptoms in China. *Journal of Family Psychology*, *27*(4), 677–682. https://doi.org/10.1037/a0033333

Nair, K. P. S. (2003). Life goals: The concept and its relevance to rehabilitation. *Clinical Rehabilitation*, *17*(2), 192–202. https://doi.org/10.1191/0269215503cr599oa

Norton, R. (1983). Measuring marital quality: A critical look at the dependent variable. *Journal of Marriage and the Family*, *45*(1), 141–151. https://doi.org/10.2307/351302

Nunnally, J. C. (1978). An overview of psychological measurement. In B. B. Wolman (Ed.), *Clinical diagnosis of mental disorders: A handbook* (pp. 97–146). Springer US. https://doi.org/10.1007/978-1-4684-2490-4_4

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443–460. https://doi.org/10.1007/BF02296207

Ponti, L., Guarnieri, S., Smorti, A., & Tani, F. (2010). A measure for the study of friendship and romantic relationship quality from adolescence to early-adulthood. *The Open Psychology Journal*, *3*(1), 76–87. https://doi.org/10.2174/1874350101003010076

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Revelle, W. (2021). *Psych: Procedures for psychological, psychometric, and personality research* [Manual]. https://CRAN.R-project.org/package=psych

Robles, T. F. (2014). Marital quality and health: Implications for marriage in the 21st century. *Current Directions in Psychological Science*, *23*(6), 427–432. https://doi.org/10.1177/0963721414549043

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

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

Samuelsson, M., & Way, E. (2018). *The valentine study: Online couple therapy with a multifactorial comparison of therapist support and varied treatment structure* [Unpublished master's thesis]. Stockholms universitet, Psykologiska institutionen.

Schumm, W. R., Paff-Bergen, L. A., Hatch, R. C., Obiorah, F. C., Copeland, J. M., Meens, L. D., & Bugaighis, M. A. (1986). Concurrent and discriminant validity of the Kansas Marital Satisfaction Scale. *Journal of Marriage and the Family*, *48*(2), 381–387. https://doi.org/10.2307/352405

Sijtsma, K., & Molenaar, I. (2002). *Introduction to nonparametric item response theory*. SAGE Publications, Inc. https://doi.org/10.4135/9781412984676

Sörbom, D. (1989). Model modification. *Psychometrika*, *54*(3), 371–384. https://doi.org/10.1007/BF02294623

Soulsby, L. K., & Bennett, K. M. (2015). Marriage and psychological wellbeing: The role of social support. *Psychology*, *6*(11), Article 11. 1349–1359. https://doi.org/10.4236/psych.2015.611132

Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family*, *38*(1), 15–28. https://doi.org/10.2307/350547

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Vajda, D., Thege, B. K., & Rózsa, S. (2019). Factor structure of the dyadic adjustment scale. *European Journal of Psychological Assessment*, *35*(3), 326–334. https://doi.org/10.1027/1015-5759/a000405

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *The European Journal of Developmental Psychology*, *9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

Whisman, M. A. (2013). Relationship discord and the prevalence, incidence, and treatment of psychopathology. *Journal of Social & Personal Relationships*, *30*(2), 163–170. https://doi.org/10.1177/0265407512455269

Yoon, M., & Lai, M. H. C. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling*, *25*(2), 201–213. https://doi.org/10.1080/10705511.2017.1387859