Measuring Relationship Quality in an International Study: Exploratory and Confirmatory Factor Validity

SAGE

Jill M. Chonody[1], Jacqui Gabb[2], Mike Killian[3], and Priscilla Dunk-West[4]

Abstract

Objective: This study reports on the operationalization and testing of the newly developed Relationship Quality (RQ) scale, designed to assess an individual's perception of his or her RQ in their current partnership. **Methods:** Data were generated through extended sampling from an original U.K.-based research project, *Enduring Love? Couple relationships in the 21st century*. This mixed methods study was designed to investigate how couples experience, understand, and sustain their long-term relationships. This article utilizes the cross-sectional, community sample (*N* = 8,132) from this combined data set, drawn primarily from the United Kingdom, United States, and Australia. A two-part approach to scale development was employed. An initial 15-item pool was subjected to exploratory factor analysis leading into confirmatory factor analysis using structural equation modeling. **Results:** The final 9-item scale evidenced convergent construct validity and known-groups validity along with strong reliability. **Conclusion:** Implications for future research and professional practice are discussed.

Keywords

relationship quality, scale, measurement, long-term couple relationships, relationship satisfaction

Introduction

Even though divorce is commonplace and many couples choose to live together without marrying, romantic coupling is a patterned and predictable feature of adulthood. This coupling has significant implications beyond the relationship, including personal emotional well-being (e.g., Proulx, Helms, & Buehler, 2007) and physical health (e.g., Kiecolt-Glaser & Newton, 2001). Understanding how individuals create enduring coupledom is, therefore, important for both research and practice, and measuring relationship quality (RQ) is an essential aspect. As such, the study of relationship satisfaction has a long history in the substantive literature. Many of the existing scales in this area are problem focused and/or validated with a sample of couples engaged in therapy. These scales may serve a specific function, but we sought to create an alternative—a *strengths-based approach* to the measurement of RQ. In other words, our aim was to develop a scale that measures the *positive* aspects of a relationship, namely, RQ, using a large international diverse community sample.

In our definition of RQ, we do not presuppose that couples are "happy" or that their relationships are trouble-free; however, we start from the premise that these partnerships are "working" at an emotional and/or practical level (Gabb & Fink, 2015a) in ways that meet the needs and/or expectations of the couple. RQ thus defined draws on ideas of emotion work

and working relationships within systemic psychotherapy wherein emotions have been seen as relational, embodied, and culturally determined (Bertrando, 2008) and are understood as relational and performative (Fredman, 2004) rather than located within individuals. This connects with sociologically informed theorizing which suggests that couples relate to and interact with each other within dynamic and intersecting micro and macro networks of relations (Burkitt, 2014) through everyday relationship practices (Gabb & Fink, 2015b).

RQ is often used interchangeably with relationship/marital satisfaction and is perhaps the most studied element of intimate relationships (Graham, Diebels, & Barnow, 2011; Heyman, Sayers, & Bellack, 1994). Research indicates that RQ, satisfaction, and adjustment are all highly correlated, indicating that these are perhaps aspects of one latent construct (Fincham &

[1] Department of Social Work, Indiana University Northwest, Gary, IN, USA
[2] Faculty of Social Sciences, The Open University, Milton Keynes, UK
[3] School of Social Work, University of Texas at Arlington, Arlington, TX, USA
[4] Faculty of Social and Behavioural Studies, Flinders University, Adelaide, Australia

Corresponding Author:
Jill M. Chonody, Department of Social Work, Indiana University Northwest, 3400 Broadway Ave, Gary, IN 46408, USA.
Email: jchonody@iupui.edu

Bradbury, 1987). Clarity of terms is crucial for research given that it is difficult to ensure that a latent construct has truly been captured when it is conflated with other, albeit similar, constructs. Delineation of an operational definition coupled with rigorous psychometric testing can advance a new scale for this substantive domain. Most existing scales, however, have failed to define their latent construct (Fincham & Bradbury, 1987; Sabatelli, 1988; Vaughn & Baier, 1999) before proceeding with scale development procedures, often drawing on items from widely used scales.

Further conceptual delimitation is, therefore, necessary to improve precision in measurement (Fincham & Bradbury, 1987; Walker & Luszcz, 2009). In addition to operationalization, other conceptual and methodological weaknesses are found in the most commonly used scales; these are discussed in detail subsequently. After the review of relationship satisfaction/quality scales, the process of scale development is outlined with specific reference to RQ. Our new instrument was tested with a community sample using a two-part approach, which involved both exploratory (EFA) and confirmatory (CFA) factor analyses. Results are presented, indicating the strength and reliability of this instrument. The final scale is provided for further validation alongside implications for future research and professional practice.

Review of Scales

According to Graham, Diebels, and Barnow (2011) and Funk and Rogge (2007), The Locke-Wallace Marital Adjustment Test (MAT), the Kansas Marital Satisfaction Scale (KMSS), the Quality of Marriage Index (QMI), the Relationship Assessment Scale (RAS), and Karney and Bradbury's (1997) semantic differential scale are the most commonly used relationship satisfaction scales. Graham et al. (2001) also included the Marital Opinion Questionnaire (Huston & Vangelisti, 1991) and the Couples Satisfaction Index (CSI) while Funk and Rogge identified the Dyadic Assessment Scale. These scales are reviewed here to illustrate areas for measurement improvement for this substantive domain. Karney and Bradbury's semantic differential along with the Marital Opinion Questionnaire are not reviewed because these scales ask participants to rate their relationship using a series of reflective adjectives (e.g., good/bad and pleasant/unpleasant). As such, semantic differentials are substantively different from scales based on item development representing a latent construct. Thus, our review is limited to those scales that utilize a Likert-type response to a series of statements aimed at evaluating the relationship.

Locke-Wallace MAT

One of the most often cited measures for marital satisfaction (Funk & Rogge, 2007) is the Locke-Wallace MAT (Locke & Wallace, 1959). This 15-item scale asks participants to rate 9 items for the level of agreement that occurs between the participant and his or her partner (e.g., philosophy of life), and a further 6 items are posed as questions (sample item: "Do you ever wish you had not married?"). All of the items for this scale were gleaned from previously published marital adjustment scales, and thus no specific operational definition was utilized. Instead, Locke and Wallace created this scale by choosing those items that "had the highest level of discrimination in the original studies ... and would cover the important areas of marital adjustment and prediction as judged by the authors" (p. 252). This approach to scale development limits the conceptualization to one that is purely statistical in nature.

The first 9 items are on the same 6-point Likert-type scale, but the final six questions each have their own response options ranging from two to four. This inconsistent scaling of the items may be problematic in terms of τ equivalent (see Graham et al., 2011, for further details) as well as weighting (Norton, 1983). Furthermore, initial validation of the scale was based on a sample of 236 participants who were "young, native-white, educated, Protestant, white-collar and professional, urban group" (Locke & Wallace, 1959, p. 254). Greater diversity in sampling strengthens scale development in terms of potential applicability to a wider range of participants.

The MAT is one of the early attempts at systematic measurement of relationship satisfaction, but despite its previous widespread use in the literature, interest in this scale is waning (Graham et al., 2011). In part, this may be due to largely outdated item content, which is not appropriate for contemporary participants. For example, Sabatelli (1988) notes that an item dealing with companionship requires a respondent to engage in all outside interests with his or her partner to achieve the highest adjustment score. Furthermore, Graham and colleagues' meta-analysis of reliability generalization found that the MAT was the weakest among the scales that they assessed (see above for the complete list) at .785. This reliability coefficient (based on 639 reliability coefficients) is significantly lower than the Cronbach's α reported in the original report (.90).

Dyadic Adjustment Scale (DAS)

The DAS (Spanier, 1976) is a 32-item scale designed to measure marital quality or RQ among long-term couples and is also widely used in the literature (Funk & Rogge, 2007). Like the MAT, item development for this scale was based on scales available at the time, and prior to data collection, items were reviewed before the final item pool were factor analyzed to create the scale. While Spanier does include an operational definition for dyadic adjustment, the items were still drawn from known scales.

Given Spanier's approach, 12 of the 15 items from the MAT are included on the DAS (Funk & Rogge, 2007). Some of these items are simply outdated. For example, 1 item inquires about the degree to which the couple agrees on "conventionality," defined parenthetically as "correct or proper behavior." For contemporary participants, and a diversely constituted sample, understandings of "correct" or "proper" behavior are unlikely to be universal. Furthermore, the reliance on the MAT fails to address the problem of inconsistent weighting of items and conceptual overlap between relationship concepts (Norton,

1983). Many relational elements (e.g., finances, recreations, and major decisions) are included on the DAS, but these issues are not necessarily applicable to all couples (i.e., noncohabitating couples). Such factors are more effectively assessed via other scales that attempt to pinpoint other relationship issues.

One of the strengths of the DAS is that this scale can be used to assess RQ with both married and cohabiting couples. However, cohabitating couples were not included in the original sample used to test the scale. Given that psychometric studies are sample dependent, this approach to scale development is worrisome, particularly in light of limited couple and participant diversity. The sample used to develop the DAS was White, married individuals from working- and middle-class backgrounds. The diversity in socioeconomic sampling strengthens the potential applicability of the scale while its racial specificity is delimiting.

Due to its evidence of good reliability and factor structure, its popularity among researchers is understandable. The reported Cronbach's α in the original study was .96. A large coefficient such as this suggests that there may be item redundancy in the scale, which inflates the correlations between items. Scale length may also be contributing to this coefficient, and at 32-items, some further reduction of items seems warranted.

KMSS

KMSS (Schumm, Nichols, Schectman, & Grigsby, 1983) was originally developed in the late 1970s as three single-item indicators, but subsequent data collection and analyses suggested that these items could be combined to create a general MSS. These items ask participants to rate "how satisfied are you with . . . your husband [wife] as a spouse?; your marriage?; and your relationship with your husband [wife]?" (Schumm et al., 1983, p. 569). In Graham et al.'s (2011) recent meta-analyses of reliability generalization, researchers found that the KMSS was the strongest among the scales assessed with a Cronbach's α that averaged .95. Schumm, Nichols, Schectman, and Grigsby (1983) report interitem correlations between .93 and .95, and these high correlations suggest item redundancy as does the nearly perfect Cronbach'sα. Nonetheless, the KMSS exhibits good face validity and has been found to be related to other satisfaction instruments (see Graham et al., 2011). Its brevity is also a significant advantage, allowing it to be readily included alongside other relationship instruments. It addresses the issue of conceptual overlap between relationship satisfaction and relationship issues that may influence satisfaction (e.g., division of labor); however, as it was originally written, the items are geared toward marital relationships and its applicability to nonmarried couples is thus limited.

QMI

The QMI (Norton, 1983) is a 6-item scale that measures "the goodness of the relationship gestalt." Items include "We have a good marriage" and "Our marriage is strong" (Norton, 1983,

p. 143). As mentioned above, the use of the terms "marriage/ marital" in the items is problematic for researchers who seek to be more inclusive in their sampling frames. However, Norton took a rigorous approach to the development of the QMI and sought to tackle problems found in the existing scales. Specifically, Norton aimed to eliminate the conceptual overlap between RQ and *components* of a relationship that can impact RQ. Additionally, Norton clearly delineated a definition for RQ and how it should be captured, as evaluative not descriptive. In other words, Norton proposed that relationships can be *described* by a set of qualities or the relationship can be *evaluated* (i.e., is this relationship good?).

While Norton's operational definition guided his decision-making process, the items for the QMI were part of another scale, the Partner Communication Scale. Twenty of the 261 items on this scale were found to "loosely satisfy the criteria of evaluative" (p. 144). Once the data were collected, the items were subjected to a two-stage process of reduction. First, a correlational analysis was performed, then a factor analysis. In Graham and colleagues' (2011) reliability analyses, the QMI was found to exhibit strong reliability (.94). But there is inconsistent scaling across the items in the QMI, with 5 items being assessed on a 7-point scale and 1 global item (overall happiness with marriage) evaluated on a 10-point scale. In sum, the primary drawback of this scale is that the terminology is limiting. It also needs further testing to determine its applicability with a diverse sample, as the original sample was drawn from the Midwest without sociodemographic information on education, race/ethnicity, religion, and sexual orientation being provided.

RAS

The RAS (Hendrick, 1988) is a 7-item scale designed to measure satisfaction in general. Sample items include "How good is your relationship compared to most?" and "How many problems are there in your relationship?". The RAS aimed to address item content that was delineated by marital status, something that was common in standardized scales at the time. Hendrick previously tested 5 of the 7 items with married couples in another study, and the full RAS was then later psychometrically tested with undergraduate students enrolled in a psychology class. Only responses from students who reported that they were "in love" were retained for further analyses (*n* = 125). Factor analysis of these data indicated a strong factor structure. A second study was then undertaken with 57 dating couples attending the university who were given course credit or a small stipend for participation. No information is given about the sociodemographic composition of the sample.

The RAS correlates with the DAS showing evidence of concurrent validity, and reliability was good (.86). However, some of the items appear problematic given the focus of this scale is RQ. For example, one item reads, "To what extent has your relationship met your original expectations?" This item does not appear to have face validity, given that an individual could justifiably indicate that this relationship does not meet his or her original expectations, yet the *quality* of the

relationship may be very high. Furthermore, items that invite comparisons to others, such as "how good is your relationship compared to most?" presuppose a normative underlying concept and a shared understanding of what constitutes a "good relationship."

CSI

Utilizing factor analysis and item response theory (IRT), Funk and Rogge (2007) developed three versions (32-, 16-, and 4-items) of the CSI. IRT allows researchers to determine which items are providing the most information, thus identifying items that are more precise in their measurement. The CSI is the only instrument to take this approach (Graham et al., 2011). Once again, previous scales on relationship satisfaction were used to create this "new" scale, and an operational definition for the latent construct was not provided. Instead, all of the items from these scales were included, aside from redundancies. An additional 71 further items were included, 25 items selected from other scales of relationship satisfaction and 46 new items, of which 35 items were "written from scratch" and the remainder 11 were modified items from the MAT and the DAS. No further information is provided on how the new items were written.

Evidence of convergent construct validity was demonstrated for the CSI suggesting that it is measuring the construct of satisfaction as conceptualized by past scale developments. Relatedly, reliability was quite high at .98, again raising concerns regarding item redundancies, but subsequent reports indicate lower αs (.90–.92; Graham et al., 2011). A large ($N =$ 5,315) and diverse sample (including people of color, range of educational backgrounds, and dating couples) was obtained to test the items for the CSI, which is a significant strength; however, the lack of conceptualization of the latent construct prior to item construction may be problematic when pinpointing what the summary scores are representing. Moreover, given that Funk and Rogge used previously developed scales, some of the items mentioned in the previous sections as potentially problematic are also found in this scale (e.g., "To what extent has your relationship met your original expectations?"). Like the MAT and DAS, some items use different types of response options; nevertheless, all of the items are placed on a 5-point scale, which eliminates the weighting issue.

Summary

The most salient issue across nearly all of the scales reviewed is that the item content is not specific to the domain of relationship satisfaction or quality. Most of these scales conflate a number of relationship constructs in that item content contains aspects of relationships that may pertain to quality, but are not a *measure* of quality, an issue that has been raised in the literature for many years (e.g., Sabatelli, 1988). For example, communication influences the quality of the relationship, but is not a determinant of it (Norton, 1983). To address this issue, our scale limits the definition of RQ to those key elements that

represent overall quality. In other words, relationship issues, such as division of labor within the household, are excluded from operationalization given that these issues can be assessed by other means to determine their role, if any, in explaining RQ. This creates a conceptually clean scale with the sole focus on RQ in-and-of-itself.

Based on our review of commonly used RQ scales, four other important issues were identified as areas of potential improvement for development of a new scale. First, lack of an operational definition that specifically articulates the focus of the scale is a key limitation. A clear conceptual definition that guides item development contributes to a parsimonious scale. As suggested above, relationship satisfaction and RQ are likely to share essential components, but further research is required to establish the precise demarcation of terms, and testing to determine the exact nature of the associations between similar latent constructs.

Second, lack of diversity in the sample used for scale validation is problematic for contemporary studies. The inclusion of racial/ethnically diverse samples as well as couples who represent modern day intimacies, including residency (e.g., cohabitating and living-apart-together relationships) and sexuality (same-sex, bisexual, and opposite-sex couples). These are important features for a scale that is meant to be representative and/or reflect community-wide diversity. Same-sex couples and cohabitators are increasingly more salient for research foci, yet available scales may not be appropriately designed to be inclusive of such diverse relationships (Graham et al., 2011).

Third, the use of the word "marital" or "marriage" in the items is problematic because these items are thus not inclusive of other coupled relationships, including those who are in a domestic partnership, civil union, or de facto relationship. Changes to currently used scales may need further psychometric testing to determine their usefulness with other populations (see Graham et al., 2011, for information on reliability of scales with different couple types). Finally, many of these scales are quite long. Respondent burden and relevance of item content are both important features in research that endeavors to inform practice and advance the substantive knowledge base.

Current Study

To address the weaknesses identified in the above scales, we sought to develop a new scale to measure RQ. Based on our review of the literature, we operationalized the construct of RQ, and then proceeded to test our items, both with experts and through advanced statistical analysis. Our primary goal was to create a strengths-based scale that addressed limitations in the currently available scales that are related to RQ, including the recruitment of a diverse, international sample.

Method

Item Development

Based on the literature, RQ was operationalized as the degree to which a commitment exists, mutual enjoyment (including

companionship) is present, and a sense that this person is the "right" one. To that end, 26 items that were geared toward these constructs were written; 13 were oriented toward commitment and companionship and the other 13 covered RQ as it relates to one's relationship with his or her partner (e.g., "My partner is usually aware of my needs"). Items were designed to interrogate the ways in which partnerships are sustained through ordinary (Brownlie, 2014) everyday relationship work (Gabb & Fink, 2015a), drawing on U.K. sociological analysis that has advanced a "practices approach" to study families (Morgan, 1996, 2011), intimacy (Jamieson, 1998), and personal life (Smart, 2007). This ongoing relationship work sustains RQ and maintains coupledom.

The initial study (*Enduring Love? Couple relationships in the 21st century* [RES-062-23-3056] was funded by the Economic and Social Research Council and completed in the United Kingdom) was designed in dialogue with members of a Strategy Board, including policy makers, professional practitioners, and senior researchers. At the outset of the project, interviews were completed with key stakeholders in U.K. family and relationship support services and government departments. Drafts of the survey were subsequently circulated among the Strategy Board and the research community more widely. This enabled us to edit and add items and refine the survey instrument. This aimed to ensure that the items were attentive to the concerns of relationship support organizations and the needs of adult couples (Walker, Barrett, Wilson, & Chang, 2010) and that findings would provide potentially useful information on how individuals experience and perceive their coupledom.

Web-based surveys have quickly moved from "novel idea to routine use" (Dillman, Smyth, & Christian, 2007, p. 447). Online surveys allow for a diverse and international group of individuals to be sampled and can capture opinions on RQ from a wide spectrum of people. This has the capacity to generate high-quality data (e.g., Chang & Krosnick, 2009; Gosling, Vazire, Srivastava, & John, 2004). Good practice guidelines for Internet-mediated research (IMR) are becoming well established (e.g., Hewson, 2003; Hewson & Laurent, 2012) and our survey was developed in line with these protocols. We also consulted with an online survey expert to make sure the instrument was technically and ethically robust, in accordance with the British Psychological Society guidelines, and that items were not double barreled or difficult to interpret.

In response to all of the above consultation, some items were reworked or replaced. These 26 reformulated items were used in our survey of couples to measure RQ; however, 1 item ("Raising children together makes our relationship stronger") was not used in any of the RQ analyses given that it does not apply to all survey participants. Items utilized a 5-point Likert-type scale (1 = *strongly disagree* and 5 = *strongly agree*).

Data Collection

Data were collected in two waves via anonymous online surveys on *Survey Monkey* as part of a larger survey on enduring

coupledom. In Wave 1, survey administration was targeted at a UK sample. Advantages of IMR methods include the capacity to recruit participants irrespective of their geographical location, and the ability to target specialist and/or "hard-to-reach" populations. Survey participants were recruited through features and news coverage of the research project posted on various online forums, newsletters, and community group noticeboards, especially those clustered around parenting and relationship support.

In Wave 2, the survey was replicated in the United States and Australia. However, recruitment in these two countries remained limited and as such there were smaller samples here than those obtained in the U.K. data collection phase. The primary method for recruitment in the United States and Australia was snowballing techniques that relied on sharing the survey link with interested participants, alongside posts (and reposts) made on Twitter and Facebook, as well through university networks where the researchers worked.

Once missing data were removed (i.e., those who opened the survey but did not complete any items) along with respondents not in a relationship (e.g., divorced), the final sample ($N = 8,132$) was obtained. While the study focused on long-term enduring relationships, what constitutes "long-term" was not specified because pilot research indicated that couples' perception of relationship duration is informed by age, childhood, personal relationship biographies, and an imagined future in this relationship (Gabb & Fink, 2015a). That is to say, perceptions of relationship duration are relative.

The survey items that were utilized in this study, in addition to the RQ scale, are described below along with the hypotheses related to their inclusion. Other survey questions were included in the questionnaire, but were not used in the current analysis; these are thus described elsewhere.

Convergent construct validity variable. A single-item indicator was used to determine overall *happiness with one's relationship,* and as a test of convergent construct validity. Participants were asked to rate this question: "How happy are you with your relationship overall?" employing a 5-point Likert-type scale (1 = *very unhappy* and 5 = *very happy*). We hypothesized a positive correlation between this single item and the RQ scale.

Known-groups validity variables. Gender included "male, female, and other" and was used as a test of known-groups validity. Substantive literature indicates that gender is not related to relationship satisfaction (Jackson, Miller, Oka, & Henry, 2014); thus, we anticipated no significant difference in RQ for this variable. *Parenthood* was assessed by a dichotomous question ("yes/no") and used as another test of known groups. A meta-analysis of the role of parenthood in relationship satisfaction (Twenge, Campbell, & Foster, 2003) indicated that parents are less satisfied than nonparents. Therefore, we hypothesized that parents would indicate less RQ than nonparents.

Sociodemographic variables. A number of other sociodemographic variables were also included and descriptively used

in this study. *Age* was measured categorically ("16–24, 25–34, 35–44, 45–54, 55–64, and 65+"). *Sexual orientation* included "heterosexual, gay/lesbian, bisexual, and other." *Religious affiliation* comprised all major religions as well as the opportunity to add one that was not listed. *Education* was measured categorically according to country-specific educational standards. Some of these categories were then combined to create a description of the overall sample. *Employment* was measured categorically and then later combined to create a description. *Relationship status* was measured as "married, living together, not living together, domestic partner, and dating." The *length of the relationship* was measured categorically ("under 1 year, 1–5 years, 6–10 years, 11–15 years, 16–20 years, and 20+ years"). *Past use of relationship support* (e.g., counseling with a therapist or pastor/religious leader, seeking consultation with a primary care physician/general practitioner) was a dichotomous question ("yes/no").

Data Analysis

Our data analysis plan for testing the newly developed RQ scale commenced with a review of item performance, including skew and kurtosis and a correlational analysis. Next, an (EFA with SPSS 22.0 was performed to determine the factor structure of the scale, and any poorly performing items were eliminated. The EFA was conducted using principle component analysis as the extraction method, and eigenvalues greater than 1 were used to identify the factors. To improve the interpretation of the factor loadings, an orthogonal rotation was used (Varimax). A CFA using structural equation modeling with MPlus 7.3 provided the final factor structure of the scale. Modification indices generated during the CFA were considered if the modification would create a change in the model χ^2 value greater than 3.84 ($p < .05$), which is a statistically significant improvement in the model. Though researchers should use these post hoc modifications to the model with care, these changes can be done where supported by theory (Jackson, Gillaspy, & Purc-Stephenson, 2009; Kline, 2011). The final model was then used for tests of convergent construct validity and known-groups validity.

Results

Demographics

The first wave of the survey was administered through a mixed methods study based in the United Kingdom ($n = 7,654$), with the subsequent wave in the United States ($n = 917$) and Australia ($n = 465$) producing additional responses. Individuals responding to the online survey across both waves included participants representing 60 different countries, including Japan, Botswana, China, Peru, India, and the Dominican Republic as well as a number of other European countries. The two waves resulted in a final sample of 8,132 individuals who fully completed the survey and reported being in a long-term relationship.

Demographic characteristics reported by respondents indicate a diverse and multinational sample of individuals (Table 1). Approximately 12% of the sample identified as a sexual minority, nearly 50% reported being either Atheist or Agnostic, and 25% of participants were not married, but rather were living together/in a civil union. However, the sample was also highly educated (75.7% with a university degree) and the most frequent response for length of relationship was over 20 years (mode with 30.6% of responses). The race and ethnicity characteristics of respondents are provided in Table 2 and demonstrate the complex nature of this international sample of individuals. From the total sample across all countries, around 25% of respondents reported their race/ethnicity as Black, Asian, or biracial/mixed ethnicity.

Evaluating Item Performance

Measures of central tendency were checked prior to undertaking the analysis of the factor structure for the RQ scale. Skew and kurtosis were not greater than 2.5 on any item, and variance in responses was acceptable. As a result of this evaluation, no items were removed.

Next, bivariate correlations between all of the items were performed to determine the degree to which these items were related to one another. No items were found to exceed a correlation of .90 (range of $r = .081$ to $r = .644$, all $p < .001$); however, 10 items were found to have no correlations >.30, indicating that the item had a weak relationship with the other items. These items were removed, and the remaining 15 items were utilized for the EFA.

EFA

The overall sample ($N = 8,132$) was split randomly and equally into two subsamples (Table 1). The two subsamples significantly differed only by gender ($\chi^2 = 4.13$, $df = 1$, $p = 0.42$), with a greater proportion of women in the EFA subsample ($n = 3,203$, 81.7% vs. $n = 3,161$, 79.9% in the CFA subsample). This difference between the two subsamples was deemed negligible, and there were no other significant between-group differences by demographic variables ($p > .05$). Split-half validation was then conducted on the two separate samples of 4,066 respondents using first EFA and then CFA. The subsample for the EFA contained 3,675 complete responses across the initial set of items (90.4%). Bartlett's test of sphericity ($\chi^2 = 20,904.61$, $df = 105$, $p < .001$) and KMO's (Kaiser-Myer-Olkin) measure of sample adequacy were excellent (KMO = .946), and the amount of explained variance was good (50.5%). Items were removed from the model based on their factor loadings and amount of variance in the item explained by the factor model. The initial factor solution indicated two factors with 15 items; however, several items had significant loadings (greater than .40) on both factors. After several iterations and removal of poor performing items as indicated by cross loadings or a weak factor loading (less than .40), a final factor model was achieved. This model contained 9 items and had a

Table 1. Sociodemographic Description of Sample.

Variable	Total Sample, N = 8,132	EFA Half, n = 4,066	CFA Half, n = 4,066
Gender			
Male	1,516 (19.2%)	719 (18.3%)	797 (20.1%)
Female	6,364 (80.8%)	3,203 (81.7%)	3,161 (79.9%)
Age			
16–24	631 (8.0%)	310 (7.8%)	321 (8.1%)
25–34	2,177 (27.5%)	1,099 (27.8%)	1,078 (27.2%)
35–44	2,023 (25.5%)	1,014 (25.7%)	1,009 (25.4%)
45–54	1,565 (19.8%)	733 (19.6%)	792 (20.0%)
55–64	1,116 (14.1%)	546 (13.8%)	570 (14.4%)
65+	409 (5.2%)	210 (5.3%)	199 (5.0%)
Sexual orientation			
Heterosexual	6,839 (88.0%)	3,405 (88.0%)	3,434 (87.9%)
Gay/lesbian	499 (6.4%)	247 (6.4%)	252 (6.5%)
Bisexual	437 (5.6%)	219 (5.6%)	218 (5.6%)
Country			
United Kingdom	5,683 (69.9%)	2,837 (69.8%)	2,846 (70.0%)
United States	1,652 (20.3%)	820 (20.2%)	832 (20.5%)
Australia	491 (6.0%)	255 (6.3%)	236 (3.7%)
Other country	306 (3.8%)	154 (3.8%)	152 (3.7%)
Education level			
No high school diploma	102 (1.5%)	46 (1.4%)	56 (1.7%)
High school diploma/equivalency	309 (4.6%)	163 (4.9%)	146 (4.3%)
Vocational training/some college	1,227 (18.2%)	598 (17.8%)	629 (18.6%)
Professional/bachelor's degree	2,855 (42.3%)	1,434 (42.7%)	1,421 (41.9%)
Master's/PhD	2,257 (33.4%)	1,119 (33.3%)	1,138 (33.6%)
Employment			
Part-time work	1,796 (26.4%)	894 (26.3%)	902 (26.4%)
Full-time work	3,143 (46.2%)	1,540 (45.3%)	1,603 (47.0%)
Retired	503 (7.4%)	256 (7.5%)	247 (7.2%)
Homemaker/carer	519 (7.6%)	256 (7.5%)	263 (7.7%)
Volunteer	85 (1.2%)	51 (1.5%)	34 (1.0%)
Full-/part-time student	454 (6.7%)	237 (7.0%)	217 (6.4%)
Not employed or working	180 (2.6%)	102 (3.0%)	78 (2.3%)
Disabled	129 (1.9%)	60 (1.8%)	69 (2.0%)
Religious affiliation			
Christian (Protestant, Catholic)	2,976 (46.7%)	1,479 (46.8%)	1,497 (46.5%)
Jewish	111 (1.7%)	51 (1.6%)	60 (0.5%)
Muslim	53 (0.8%)	28 (0.9%)	25 (0.8%)
Buddhist	81 (1.3%)	47 (1.5%)	34 (1.1%)
None	3,118 (48.9%)	1,534 (48.6%)	1,584 (49.3%)
Other (Sikh, Hindu)	34 (0.5%)	18 (0.6%)	16 (0.5%)
Parent (yes)	2,966 (44.4%)	1,477 (44.3%)	1,489 (44.4%)
Relationship status			
Married	4,981 (62.7%)	2,500 (63.1%)	2,481 (62.3%)
Couple-not living together	832 (10.5%)	406 (10.3%)	426 (10.7%)
Living together	1,744 (22.0%)	859 (21.7%)	885 (22.2%)
Civil partnership	250 (3.1%)	129 (3.3%)	121 (3.0%)
Dating	133 (1.7%)	65 (1.6%)	68 (1.7%)
Number of years in relationship			
Under 1 year	336 (4.2%)	169 (4.2%)	167 (4.2%)
1–5	1,813 (22.6%)	915 (22.8%)	898 (22.4%)
6–10	1,506 (18.8%)	746 (18.6%)	760 (18.9%)
11–15	1,133 (14.1%)	567 (14.2%)	566 (14.1%)
16–20	779 (9.7%)	384 (9.6%)	395 (9.8%)
20+	2,451 (30.6%)	1,224 (30.6%)	1,227 (30.6%)
Relationship support (no)	4,775 (65.7%)	2,372 (65.2%)	2,403 (66.2%)
Happy with relationship[b]	4.29 (0.87)	4.28 (0.86)	4.30 (0.87)
Relationship quality[c]	37.70 (5.97)	37.63 (5.94)	37.79 (6.01)

Note: CFA = confirmatory factor analysis; EFA = exploratory factor analysis.
[a]Sample sizes are different on each variable due to missing data.
[b]Theoretical range = 1–5.
[c]Theoretical range = 9–45 (based on final scale).

Table 2. Ethnicity by Country.

		Country of Respondent			
Ethnicity	Total N	UK	USA	AUS	Other
White British, American, Australian	5,004 (74.3%)	3,874 (81.5%)	670 (49.2%)	393 (97.3%)	67 (30.9%)
Other White	1,286 (19.1%)	601 (12.6%)	561 (41.2%)	5 (1.2%)	119 (54.8%)
Caribbean	29 (0.4%)	23 (0.5%)	4 (0.3%)	5 (1.2%)	2 (0.9%)
African/African American	69 (1.0%)	41 (0.9%)	27 (2.0%)	0 (0.0%)	1 (0.5%)
Other African decent	11 (0.2%)	5 (0.1%)	6 (0.4%)	0 (0.0%)	0 (0.0%)
Indian, Asian subcontinent	63 (0.9%)	53 (1.1%)	5 (0.4%)	0 (0.0%)	5 (2.3%)
Asian	64 (1.0%)	36 (0.8%)	17 (1.2%)	1 (0.2%)	10 (4.6%)
Hispanic/Latino	18 (0.3%)	0 (0.0%)	17 (1.2%)	0 (0.0%)	1 (0.5%)
Native/aboriginal	5 (0.1%)	0 (0.0%)	2 (0.1%)	3 (0.7%)	0 (0.0%)
Mixed ethnicity, other	186 (2.8%)	119 (2.5%)	53 (3.9%)	2 (0.5%)	12 (5.5%)

Table 3. Exploratory Factor Analysis: Factor Loadings ($n = 4,066$).

Relationship Quality Item	Factor Loading	Commonality Score
I am content in our relationship	.838	.703
This is the relationship I always dreamed of	.794	.630
We have grown apart over time[a]	.748	.559
I am totally committed to making this relationship work	.745	.554
We enjoy each other's company	.733	.537
My partner is usually aware of my needs	.706	.499
I think of my partner as my soul mate	.703	.495
My partner makes me laugh	.686	.471
We have shared values	.655	.430

[a]Reverse scored.

Cronbach's α reliability coefficient of .888. Bartlett's test of sphericity ($\chi^2 = 14,780.16$, $df = 36$, $p < .001$) and KMO's measure of sample adequacy were again excellent (KMO = .928), and the amount of explained variance improved (54.2%) from the initial model. Table 3 provides the final factor loadings and communalities for the scale.

CFA

With the other half of the sample ($n = 4,066$), CFA was conducted. This subsample contained 3,858 complete responses (94.8%) across the 9 items identified in the EFA. The initial and final CFA models are listed in Table 4.

To assess the fit of the obtained RQ data to the EFA measurement model, multiple fit indices were obtained. The model χ^2 per degrees of freedom (χ^2/df), Tucker–Lewis index (TLI), comparative fit index (CFI), standardized root mean square residual (SRMR), and the root mean square error of approximation (RMSEA) were all used. Lower scores for the model χ^2 statistic (Kline, 2011) and the model χ^2/df indicate better fit between the data and model. Bollen (1989) suggests a χ^2/df value between 2.0 and 3.0 indicates adequate model fit. The CFI and TLI compare the model to the fit of a baseline model

with values ≥ 0.95 indicative of acceptable model fit (Hu & Bentler, 1999; Kline, 2011). SRMR is a measure of the residuals between the input covariance and measurement model matrices. SRMR values less than .08 or .10 indicate good model fit (Brown, 2006). Lastly, RMSEA adjusts for model parsimony and estimates the difference between model covariances and the observed covariances. Values between .08 and .10 are indicative of adequate fit. RMSEA values for each model were tested for significant differences from .05 along with the 90% confidence interval of the estimates (Kline, 2011).

The CFA tested the factor structure for the RQ scale found in the EFA. Initial results indicated adequate fit (see Table 4), but further improvements to the model were suggested through the indices produced ($\chi^2 > 3.84$, $p < .05$). Correlations between individual item error terms were added to the model, given that they offered the greatest decrease in the model χ^2 value. Two modifications were made to the model prior to the final model. In order, the error terms for items were allowed to correlate which reduced the χ^2 by 170.35 ($p < .001$) and then by 74.85 ($p < .001$). The final model demonstrated excellent fit across all fit indices (see Table 4). Figure 1 provides the factor loadings and error terms for the final RQ scale, and Table 5 lists the items.

Reliability

A total RQ score was calculated by summing responses to the items identified from the EFA and CFA. The resulting measure demonstrated high internal consistency reliability with a Cronbach's α of .891 when analyzed over the total sample.

Convergent Construct and Known-Group Criterion-Related Validities

Respondents were asked to assess their happiness regarding their current relationship. The RQ scale was highly, positively correlated with these self-reports of happiness ($r = .787$, $p < .001$) and indicated evidence of convergent construct validity. Relatedly, respondents children had significantly different RQ scores than those without children, $t = 9.56$, $df = 5,609$,

Table 4. Confirmatory Factor Analysis: Model Fit Indices (*n* = 4,066).

Model	χ^2	df	χ^2/df	RMSEA	90% CI	*p* Value	CFI	TLI	SRMR
Initial	594.12***	27	22.0	.074	[.069, .079]	*p* < .001	.965	.953	.027
Final	292.73***	25	11.7	.053	[.047, .058]	*p* = .199	.983	.976	.020

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; TLI = Tucker–Lewis index; χ^2 = chi-square; *df* = degrees of freedom.
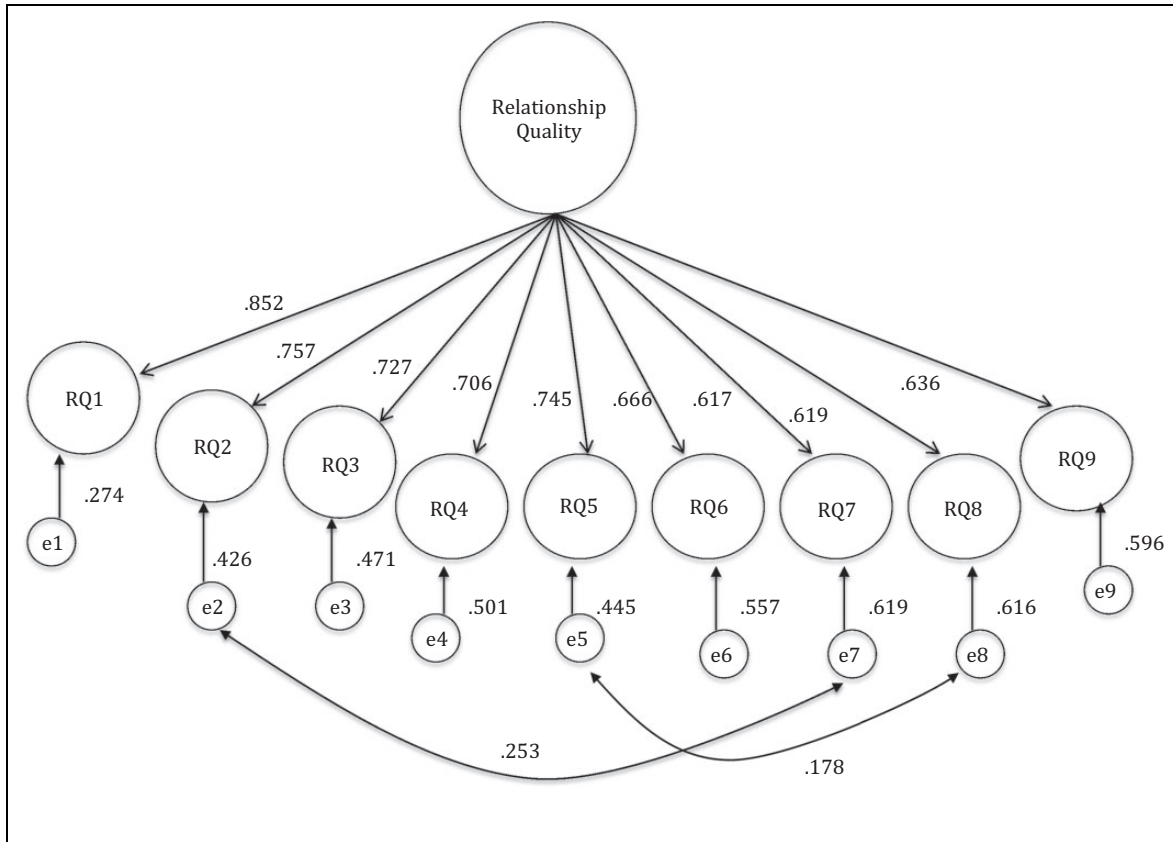***p < .001.

Figure 1. Confirmatory factor analysis: Item loadings (*n* = 4,066). Items correspond to item list in Table 5.

Table 5. Final Relationship Quality (RQ) Scale.

Item Label	Items
RQ1	I am content in our relationship.
RQ2	This is the relationship I always dreamed of.
RQ3	We have grown apart over time.[a]
RQ4	I am totally committed to making this relationship work.
RQ5	We enjoy each other's company.
RQ6	My partner is usually aware of my needs.
RQ7	I think of my partner as my soul mate.
RQ8	My partner makes me laugh.
RQ9	We have shared values.

[a]Reverse scored.

p < .001, and Cohen's *d* = 0.25. Those without children (*M* = 38.39, *SD* = 5.44) reported higher relationship scores than those respondents with children (*M* = 36.94, *SD* = 6.54).

Consistent with the literature, parents are found to report lower relationship satisfaction (Twenge et al., 2003); thus, this finding provides evidence of known-groups validity. Lastly, no significant differences in RQ scores, *t* = .31, *df* = 7,307, *p* = .753, and Cohen's *d* = 0.04, were reported between men (*M* = 37.76, *SD* = 5.95) and women (*M* = 37.71, *SD* = 5.98). This is also consistent with the literature that suggests relationship satisfaction is not different by gender (Jackson et al., 2014).

Discussion and Applications to Practice

Results of our study provide evidence for the initial validation of the RQ scale. Designed for and tested with a sample of individuals in an enduring relationship, this new scale shows evidence of factorial validity, convergent construct validity, and known-groups validity. This scale is also short and easy to administer with strong reliability. For these reasons, this

scale may be useful in survey research on couple relationships. The RQ also advances contemporary research interests in diverse couples (e.g., cohabitators) using the word "partner" instead of spouse/husband/wife. This new scale builds on established measures, but avoids some of the problematic aspects of those scales, including conceptual overlap with other relationship issues, such as conflict, communication, or parenthood. These variables should be studied as factors *related to* RQ instead of components *of it*. Thus, other standardized scales or single-item indicators can assess these issues to determine how other relationship issues impact overall RQ. This is an improvement over existing scales of relationship satisfaction (e.g., MAT, DAS, and CSI) that conflate these concepts.

The new RQ scale also, and importantly, represents a *strengths-based approach* to the measurement of RQ in that items are focused on positive elements of the relationship instead of a problems-focused agenda. By focusing on elements of the relationship that may be working, the scale summary score is indicative of the degree to which positive aspects of the relationship are present. The focus on everyday relationship practices as the means through which couples sustain their long-term partnerships thus shifts the emphasis away from regular markers of RQ (such "good" communication or regular and mutually "enjoyable" sexual intimacy) and traditional, culturally inscribed understandings of what makes a relationship work. This has the capacity to extend understandings of how RQ is manifest, in an everyday sense, and to enrich knowledge on what constitutes RQ in a working relationship. As such, it has the potential to make a significant contribution to and have practical applications in the fields of relationship support and intervention.

Utilizing a community sample, instead of one comprised of individuals/couples engaged in relationship therapy, means that the RQ scale has the capacity to be used with a wider and/or general population. Future research would be needed to determine if it could be used specifically with those engaged in couples work. For example, given the items on the RQ scale, it may be useful as an initial assessment tool to determine where the couple presently are in their relationship, and provide some indication of where they may want to aim toward, in the future. Relatedly, future research is needed to determine if the RQ scale can discriminate between distressed and nondistressed couples; this would expand the usefulness of this scale to a clinical setting. Further exploration of the RQ scales criterion-related and construct validity may reveal clinical utility and potential uses by practitioners.

The RQ scale provides an indication of RQ at one point in time. This may be helpful for both researchers and practitioners who seek to obtain an overall assessment of individual perceptions of relationships and determine the role of other factors that may be influencing RQ (e.g., communication). However, to fully comprehend RQ, a past point of reference is necessary and thus longitudinal data are needed to determine any change over time (Bradbury, Fincham, & Beach, 2000). Future research could seek to use the RQ in longitudinal studies with both individuals in the community and those who are help seeking, to determine its sensitivity to changes in RQ over time.

The results of our study should be considered within the framework of its limitations. The survey was completed by respondents representing 60 different countries; however, the sample was collected primarily from 3 countries. The vast majority of these respondents were from Euro-centric countries with historic ties to British colonialism. Furthermore, the sample was exceedingly well educated and female. Given the online nature of the survey, a high level of education and greater participation by women can be expected. Notwithstanding these limitations, our sample did achieve some degree of diversity in terms of other sociodemographic characteristics. Just over 37% of the sample was cohabitating, in civil union/domestic partnership, or were noncohabitating long-term partners; 12% of respondents reported their sexual orientation as lesbian, gay, or bisexual. There was also a good distribution of age.

Generalization of the results should be done with caution, and future research with the RQ scale should employ methods to obtain more diverse samples of individuals, especially in terms of education, socioeconomic background, and cultural diversity. The split-half factor validation process was exploratory and as such the RQ measure was first identified through EFA and later error terms were allowed to correlate in the CFA model where appropriate. Though the use of the split-half factor validation process conducted with the present sample adds confidence to the factorial validity of the RQ scale, psychometric studies are sample dependent, and additional validation studies of the RQ scale are warranted, including further investigation into how the RQ correlates to other standardized measures of relationship/marital satisfaction.

In sum, the findings from our preliminary study indicate initial validation of the RQ scale. Based on a large community-dwelling sample from multiple countries, the RQ showed good reliability and evidence of validity. The RQ addressed some limitations found in other relationship scales, such as anachronistic items, limiting terms (e.g., "marital"), inconsistency in response options, and includes a focus on relationship strengths without the inclusion of additional RQs (e.g., communication). Additional psychometric studies with community samples can expand the utility of this scale, which may include application in practice as an assessment tool.

Declaration of Conflicting Interests

Funding

References

Bertrando, P. (2008). Emotional dances: Therapeutic dialogues as embodied systems. *Journal of Family Therapy*, *30*, 362–372.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bradbury, T. N., Fincham, F. D., & Beach, S. R. H. (2000). Research on the nature and determinants of marital satisfaction: A decade in review. *Journal of Marriage and Family*, 62, 964–980.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Brownlie, J. (2014). *Ordinary relationships: A sociological study of emotions, reflexivity and culture*. Basingstoke, England: Palgrave Macmillan.

Burkitt, I. (2014). *Emotions and social relations*. London, England: Sage.

Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly*, 73, 641–678.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2007). *Internet, mail, and mixed-mode surveys: The tailored design method*. London, England: Wiley.

Fincham, F. D., & Bradbury, T. N. (1987). The assessment of marital quality: A reevaluation. *Journal of Marriage and the Family*, 49, 797–809.

Fredman, G. (2004). *Transforming emotion: Conversations in counselling and psychotherapy*. London, England: Whurr Publishers.

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*, 21, 572–583.

Gabb, J., & Fink, J. (2015a). *Couple relationships in the 21st century*. London, England: Palgrave Macmillan.

Gabb, J., & Fink, J. (2015b). Telling moments and everyday experience: Multiple methods research on couple relationships and personal lives. *Sociology*, 49, 970–987.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist*, 59, 93–104.

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, 39–48.

Hendrick, S. (1988). A generic measure of relationship satisfaction. *Journal of Marriage and the Family*, 50, 93–98.

Hewson, C. (2003). Conducting research on the Internet. *The Psychologist*, 16, 290–293.

Hewson, C., & Laurent, D. (2012). Research design and tools for Internet research. In J. Hughes (Ed.), *Sage Internet research methods* (Vol. 1) (pp. 165–104). London, England: Sage.

Heyman, R. E., Sayers, S. L., & Bellack, A. S. (1994). Global marital satisfaction versus marital adjustment: An empirical comparison of three measures. *Journal of Family Psychology*, 8, 432–446.

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, 1–55.

Huston, T. L., & Vangelisti, A. L. (1991). Socioemotional behavior and satisfaction in marital relationships: A longitudinal study. *Journal of Personality and Social Psychology*, 61, 721–733.

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting practice in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14, 6–23.

Jackson, J. B., Miller, R. B., Oka, M., & Henry, R. G. (2014). Gender differences in marital satisfaction: A meta-analysis. *Journal of Marriage and Family*, 76, 105–129.

Jamieson, L. (1998). *Intimacy: Personal relationships in modern societies*. Cambridge, England: Polity Press.

Karney, B. R., & Bradbury, T. N. (1997). Neuroticism, marital interaction, and the trajectory of marital satisfaction. *Journal of Personality and Social Psychology*, 72, 1075–1092.

Kiecolt-Glaser, J. K., & Newton, T. L. (2001). Marriage and health: His and hers. *Psychological Bulletin*, 127, 472–503.

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

Locke, H. J., & Wallace, K. M. (1959). Short marital-adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living*, 21, 251–255.

Morgan, D. H. J. (1996). *Family connections: An introduction to family studies*. Cambridge, England: Polity Press.

Morgan, D. H. J. (2011). *Rethinking family practices*. Basingstoke, England: Palgrave Macmillan.

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

Proulx, C. M., Helms, H. M., & Buehler, C. (2007). Marital quality and personal well-being: A meta-analysis. *Journal of Marriage and Family*, 69, 576–593.

Sabatelli, R. M. (1988). Measurement issues in marital research: A review and critique of contemporary survey instruments. *Journal of Marriage and the Family*, 4, 891–915.

Schumm, W. R., Nichols, C. W., Schectman, K. L., & Grigsby, C. C. (1983). Characteristics of responses to the Kansas marital satisfaction scale by a sample of 84 married mothers. *Psychological Reports*, 53, 567–572.

Smart, C. (2007). *Personal life*. Cambridge, England: Polity Press.

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, 15–28.

Twenge, J. M., Campbell, W. K., & Foster, C. A. (2003). Parenthood and marital satisfaction: A meta-analytic review. *Journal of Marriage and Family*, 65, 574–583.

Vaughn, M. J., & Baier, M. E. M. (1999). Reliability and validity of the relationship assessment scale. *The American Journal of Family Therapy*, 27, 137–147.

Walker, J., Barrett, H., Wilson, G., & Chang, Y. S. (2010). *Understanding the needs of adults (particularly parents) regarding relationship support* (Research brief DCSF-RBX-10-01). London, England: DCFS.

Walker, R. B., & Luszcz, M. A. (2009). The health and relationship dynamics of late-life couples: A systematic review of the literature. *Ageing and Society*, 29, 455–480.