


Evaluating and Refining the Construct of Sexual Quality With Item Response Theory: Development of the Quality of Sex Inventory

Amanda M. Shaw¹ · Ronald D. Rogge¹ 

Received: 31 May 2014 / Revised: 22 September 2015 / Accepted: 14 October 2015 / Published online: 4 January 2016
© Springer Science+Business Media New York 2015

Abstract This study took a critical look at the construct of sexual quality. The 65 items of four well-validated self-report measures of sexual satisfaction (the Index of Sexual Satisfaction [ISS], Hudson, Harrison, & Crosscup, 1981; the Global Measure of Sexual Satisfaction [GMSEX], Lawrance & Byers, 1995; the Pinney Sexual Satisfaction Inventory [PSSI], Pinney, Gerrard, & Denney, 1987; the Young Sexual Satisfaction Scale [YSSS], Young, Denny, Luquis, & Young, 1998) and an additional 74 potential sexual quality items were given to 3060 online participants. Using Item Response Theory (IRT), we demonstrated that the ISS, YSSS, and PSSI scales provided suboptimal levels of precision in assessing sexual quality, particularly given the length of those scales. Exploratory factor analyses, IRT, differential item functioning analyses, and longitudinal responsiveness analyses were used to develop and evaluate the Quality of Sex Inventory. Results suggested that, in comparison to existing scales, the QSI (1) offers investigators and clinicians more theoretically focused scales, (2) distinguishes sexual satisfaction from sexual dissatisfaction, and (3) offers greater precision and power for detecting differences with (4) comparably high levels of responsiveness for detecting change over time despite being notably shorter than most of the existing scales. The QSI-satisfaction subscales demonstrated strong convergent validity with other measures of sexual satisfaction and excellent con-

struct validity with anchor scales from the nomological net surrounding that construct, suggesting that they continue to assess the same theoretical construct as prior scales. Implications for research are discussed.

Keywords Sexual satisfaction · Sexual quality · Couples · Marriage · Item response theory

Introduction

As sex is a defining characteristic of most romantic relationships, couples and sex researchers have generally been interested in the quality of sex within relationships. As a result, a wide range of scales have been developed to assess sexual quality (primarily focusing on sexual satisfaction). Although a number of these scales (e.g., the Index of Sexual Satisfaction [ISS], Hudson, Harrison, & Crosscup, 1981; the Global Measure of Sexual Satisfaction [GMS-EX], Lawrance & Byers, 1995; the Pinney Sexual Satisfaction Inventory [PSSI], Pinney, Gerrard, & Denney, 1987; the Young Sexual Satisfaction Scale [YSSS], Young, Denny, Luquis, & Young, 1998) have become widely used and validated, as a set they tend to (1) be notably heterogeneous in content, varying in how they define the limits of this construct, and (2) conceptualize the quality of sexual activity as a single dimension (ranging from dissatisfaction to satisfaction on a single continuum). The current study sought to advance recent measurement evaluation work in this area (e.g., Mark, Herbenick, Fortenberry, Sanders, & Reece, 2014) by (1) using advanced statistics to evaluate a set of sexual satisfaction measures (both for their cross-sectional precision and power as well as their responsiveness to detecting change over time), (2) refining the conceptual definition of this construct, and (3) developing a new psychometrically optimized and bi-dimensional measure of sexual quality.

✉ Ronald D. Rogge
rogge@psych.rochester.edu

¹ Department of Clinical and Social Sciences in Psychology,
University of Rochester, RC Box 270266, Rochester,
NY 14627-0266, USA

The Importance of Sexual Quality

Defining Terms

When assessing global evaluations (e.g., the global subjective experience of a romantic relationship or a sexual relationship), researchers commonly used the terms “satisfaction,” “happiness,” and “quality” interchangeably (see Fincham & Rogge, 2010 for a discussion). This is largely due to the fact that items containing those various terms tend to correlate very strongly with one another and tend to show some of the highest loadings on the corresponding global evaluation being assessed, suggesting that they all represent prototypical items of such constructs (e.g., Funk & Rogge, 2007). However, in the interest of conceptual clarity, in the current paper we used the term *sexual quality* to refer to the overarching construct of individuals’ global positive and negative evaluations of their sex lives. Within this framework, we conceptualized sexual satisfaction and dissatisfaction as inter-related components of sexual quality representing those global positive and negative evaluations of one’s sex life. We further argue that satisfaction and dissatisfaction might actually represent distinct attitudes that, although correlated, could change independently of one another over time. Thus, in the current paper we primarily used the term *sexual quality* (rather than *sexual satisfaction*) to refer to the overarching construct of interest as that allows for both *sexual satisfaction* and *sexual dissatisfaction* to represent distinct components of that construct.

Prior Work

Sexual quality is a widely studied construct that is often central to an individual’s sexual life. The breadth of research that has assessed sexual quality is extensive and reaches into a variety of different domains. Numerous empirical articles have examined the links between sexual quality and relationship quality (Butzer & Campbell, 2008; Byers, 2005; Haavio-Mannila & Kontula, 1997; Oggins, Leber, & Veroff, 1993), infidelity (Mark, Janssen, & Milhausen, 2011; Plack, Kroger, Allen, Baucom, & Hahlweg, 2010; Shaw, Rhoades, Allen, Stanley, & Markman, 2013), body image (Daniel & Bridges, 2013; Pujols & Meston, 2009), physical health and illness (Connell, Coates, Doherty-Poirier, & Wood, 2013; Foster & Byers, 2013; Yount, 2013), childhood sexual abuse (Easton, Coohy, O’Leary, Zhang, & Hua, 2011; Meston, Lorenz, & Stephenson, 2013; Stroebel et al., 2012), and mental health (Holmberg, Blair, & Phillips, 2010; Suvak, Brogan, & Shipherd, 2012; Syme, Klonoff, Macera, & Brodine, 2013), suggesting that the quality of an individual’s sexual life can be strongly linked to many other areas of individual functioning. Additionally, in sex therapy and sexual medicine, sexual quality is often used as a gauge of sexual and relationship health. Hence, precise and sensitive measures of sexual quality are needed, and the current study sought to evaluate and potentially enhance those measures in support of this growing body of research.

Defining Sexual Quality

Previous work on sexual quality has been largely based on extremely broad definitions of the construct, leading researchers to include diverse item content that is representative of the various types of complaints that couples might have with their sex lives. Although we agree with the pragmatic utility of creating such broad-band scales, this conceptual heterogeneity has been criticized (e.g., Mark et al., 2014) as it confounds the measurement of sexual quality (the criterion of interest) with the measurement of distinct constructs that help to predict sexual quality. This predictor–criterion overlap (as described by Mark et al.) can, therefore, yield scales so broad that they run the risk of obscuring interesting results by collapsing the measurement of as many as five different constructs into a single scale with a single-scale score, as was found for the ISS. To address this concern, the current study sought to clarify and refine the theoretical definition of this construct. We strongly agree with Mark et al. that items assessing topics such as frequency of sexual activity, sexual functioning and/or dysfunctioning, sexual communication, orgasmic frequency, emotional closeness, sexual awareness, desires for sexual activity, and attitudes toward sex (see Derogatis & Melisaratos, 1979; Meston & Trapnell, 2005; Stulhofer, Busko, & Brouillard, 2010) could be conceptualized as distinct constructs from sexual quality, thereby opening up further areas of research. To exemplify this at a concrete level, an individual could have extremely enjoyable and fulfilling sex with his or her partner, but could desire that sex to occur more frequently than it currently does. Although the levels of one predictor of sexual quality might be a little lower (desire for more sex), that individual might still report high levels of sexual quality due to the higher levels of other predictors (e.g., sexual functioning, partner’s sexual responsiveness/skill, orgasmic frequencies). Distinctions like this (and the useful insights they could provide researchers and clinicians) would be completely obscured if a measure of sexual quality contained items tapping all of those various constructs only to be collapsed into a single score. We, therefore, conceptualized these constructs as independent predictors of sexual quality—worthy of being measured and studied *independently* of sexual quality. This is a critical distinction, because it is only possible to determine the relative importance of those various factors for determining sexual quality across the course of a relationship if those various predictors are measured as separate constructs. We further posited that each of those predictors of sexual quality, although likely related to one another, could change independently of one another over time, again suggesting the importance of conceptualizing them and measuring them as separate constructs. Finally, we suspected that each of those predictors of sexual quality could serve as distinct points of intervention for improving an individual’s sex life, once again underscoring the value of treating them as conceptually distinct factors. Thus, including such diverse items within measures of sexual quality would confound the assessment of global evaluations (i.e., how happy the

respondent is with his/her sex life) with predictors of those evaluations, potentially obscuring meaningful results that might have been obtained if each of those constructs were assessed separately. Following conceptual and measurement work in the couples literature (e.g., Fincham & Bradbury, 1987; Norton, 1983; Schumm, Nichols, Schectman, & Grinsby, 1983), we chose to conceptually limit our theoretical definition of sexual quality to individuals' positive and negative subjective global evaluations of their sex lives in an effort to create more conceptually distinct and precise scales for future use.

Existing Sexual Quality Measures

Our more specific conceptual definition of sexual quality (limiting our interest to global positive and negative sentiment toward one's sexual relationship with a partner) led us to exclude a number of scales from the current project as they were either (1) based on a far broader and diverse definition of sexual satisfaction/quality, (2) focused more on sexual dysfunction than global sentiment, or (3) were multidimensional in nature and, therefore, measured a variety of constructs beyond just global sentiment. These excluded scales included the Monash Women's Health Program Female Sexual Satisfaction Questionnaire (Davison, Bell, La Chine, Holden, & Davis, 2008), the Derogatis Sexual Functioning Inventory (Derogatis & Melisaratos, 1979), Eysenck's Inventory of Attitudes to Sex (Eysenck, 1970), the Quality of Sexual Function scale (Heinemann, Potthoff, Heinemann, Pauls, Ahlers, & Saad, 2005), the Interpersonal Exchange Model of Sexual Satisfaction (Lawrance, Byers, & Cohen, 2011), the Sexual Interaction Inventory (LoPiccolo & Steger, 1974), the Sexual Satisfaction Scale for Women (Meston & Trapnell, 2005), the Modified Rush Sexual Inventory (Rao, Zajacka, & Skubiak, 2005), the Golombok-Rust Inventory of Sexual Satisfaction (Rust & Golombok, 1986), the Multidimensional Sexuality Questionnaire (Snell, Fisher, & Walters, 1993), the New Sexual Satisfaction Scale (Stulhofer et al., 2010), and the Whitley Inventory of Sexual Satisfaction (Whitley & Poulsen, 1975). Thus, in an effort to reduce participant burden and to focus on our more specific conceptual definition, the current study examined four short and well-validated measures of sexual quality, focusing on scales that are primarily used in relationship research and are freely available for research use. The 25-item ISS (Hudson et al., 1981) is one of the most widely cited measures of sexual quality within the relationship literature (see Mark et al., 2014) and was designed to be used as a unidimensional instrument. As the ISS was originally designed to assess response to treatment for sexual problems, the authors developed the items based on common complaints from clients regarding their sexual lives. The 11-item YSSS (Young et al., 1998) is also a widely cited measure of sexual quality within research about committed relationships and was developed to assess several aspects of sexual quality including pleasure, attraction, intensity, and fulfillment among married individuals. The 24-item PSSI (Pinney et al., 1987) was

originally developed to assess sexual quality in women (either married or single) and to improve upon the psychometric weaknesses of other sexual quality scales including the ISS. The final measure we chose to assess, the 5-item GMSEX, is substantially shorter than the other measures (Lawrance & Byers, 1995). Although not as widely cited, the GMSEX was included as it represents a novel approach to measuring sexual quality given its semantic differential format which assesses sexual quality by asking respondents to rate the quality of their sex lives on a 7-point scale spanning pairs of opposing adjectives (i.e., good–bad, pleasant–unpleasant, positive–negative).

Psychometrics of Existing Unidimensional Sexual Quality Scales

In one of the few studies of its kind, Mark et al. (2014) directly compared the psychometrics of the ISS, the NSSS-S (a shortened and presumably unidimensional form of that scale), the GMSEX, and a single-item assessment for assessing sexual quality. These scales were given to 425 respondents and 2-month follow-up data were collected from 139 of the original respondents. Using classical test theory analyses (i.e., correlational methods including internal consistency, test–retest reliability, convergent validity correlations, and exploratory factor analysis), the authors demonstrated that the GMSEX demonstrated better psychometric properties than the ISS or NSSS-S. The current study sought to extend this work by (1) looking at an expanded set of scales in a much larger sample, (2) exploring the generalizability of the Mark et al. findings by examining test–retest reliabilities and internal consistencies of the scales across a range of diverse subsamples, (3) augmenting the correlational analyses of Mark et al. by using Item Response Theory analyses (IRT; a large-sample statistical technique) to more precisely explore the information provided by the current scales when used as unidimensional scales, (4) extending the test–retest correlations of Mark et al. to more precisely examine how responsive each scale is to detecting meaningful change over time, and (5) using Exploratory Factor Analysis (EFA) and then separate IRT analyses on the satisfaction and dissatisfaction items identified within a 139-item pool to create the psychometrically optimized subscales of a bi-dimensional measure of sexual quality (the Quality of Sex Inventory).

Distinguishing Sexual Satisfaction from Sexual Dissatisfaction

The measures of sexual quality that primarily focus on global evaluations and were, therefore, included in the current study (i.e., the ISS, YSSS, PSSI, and GMSEX) treat this construct as a single dimension (i.e., unidimensional) with high levels of sexual satisfaction at one end of the continuum and high levels of sexual dissatisfaction at the opposite end. Thus, each of the existing scales yields a single

score representing respondents' levels of satisfaction-to-dissatisfaction. This assumes that individuals are either satisfied or dissatisfied with their sex lives, rather than allowing for the possibility that individuals may be ambivalent (potentially feeling both sexually satisfied and dissatisfied at the same time) or indifferent (feeling low levels of both sexual satisfaction and dissatisfaction) about their sexual relationships with their partners. Recent work on global evaluations of romantic relationship quality has suggested that relationship satisfaction and relationship dissatisfaction can effectively be treated as distinct dimensions (e.g., Fincham & Rogge, 2010; Mattson, Rogge, Johnson, Davidson, & Fincham, 2012) and that disentangling those dimensions might uncover distinct treatment effects obscured by unidimensional scales (e.g., Rogge, Cobb, Lawrence, Johnson, & Bradbury, 2013). To build on this work, the current study used factor analyses in a diverse item pool to examine the possibility that sexual dissatisfaction might represent a distinct dimension.

Heterogeneity of Item Content

Consistent with the broad-band conceptual definitions used by previous researchers in developing measures of sexual quality, the existing scales have markedly heterogeneous item content. For instance, the PSSI contains ten items beginning with the words "I wish," which may be measuring sexual desires rather than sexual quality (e.g., "I wish my partner were more loving and caring when we make love," "I wish I were less inhibited when I make love"). Similarly, the ISS contains items that may tap into sexual attitudes (e.g., "I feel that sex is dirty and disgusting," "I think sex is wonderful,") as well as sexual hygiene (e.g., "My partner observes good personal hygiene"). The YSSS also contains diverse items that measure sexual communication (e.g., "I have good communication with my spouse about sex") and physical attraction (e.g., "I am sexually attracted to my spouse"). Consistent with our concerns about the appropriateness of including such diverse items into sexual quality scales, the authors of the ISS highlighted that four of their items cross-load on factors measuring constructs other than sexual quality including *relationship* quality and sexual attitudes (Hudson et al., 1981), and a recent factor analysis of the ISS revealed five distinct factors (Mark et al., 2014). The EFA and item response theory analyses employed in the current study allowed us to objectively evaluate the appropriateness of including such items on measures of global sexual quality.

The Current Study

Item Response Theory

The current study sought to evaluate existing measures of sexual quality and to develop a new sexual quality scale by augmenting more traditional measurement analyses (e.g., internal consistency, test–retest correlations, EFA) with Item Response Theory (IRT, Hambleton, Swaminathan, & Rogers, 1991). IRT is a large-

sample, statistical technique (requiring a minimum of 500 subjects; Clark & Watson, 1995) that offers researchers a higher degree of precision in evaluating items for self-report scales. It accomplishes this by estimating latent scores for each respondent on the construct being assessed (termed theta or θ scores in the IRT equations), and then examining how tightly responses to each item are linked to respondents' actual levels of that construct. Thus, IRT considers items to be informative when respondents with lower levels of the construct reliably select lower-ranked responses whereas respondents with higher levels of the construct select higher-ranked responses. In fact, IRT models those response distributions in order to quantify the nature of information provided by each item in an analysis. Although IRT was originally developed to evaluate dichotomous items (e.g., correct vs. incorrect, true vs. false; typically using a 2-parameter model), it is possible to use a simple elaboration of those equations (the Graded Response Model, GRM; Samejima, 1969) to model the quality of information provided by items using Likert response scales. The GRM elaboration of IRT estimates a set of parameters for each item to approximate the response curves for that item (i.e., proportions of subjects giving each possible response at various levels of θ). IRT then synthesizes the information from those item response curves into Item Information Curves (IICs), visually depicting how much information or precision each item offers across the entire range of the latent trait being measured (from 3 SDs below the mean to 3 SDs above the mean). These IICs can then be summed to create Test Information Curves (TICs) to reveal the information provided by different sets of items when combined into specific scales. While the IRT IICs provide researchers with information above and beyond traditional correlational statistics (e.g., item-to-total correlations, factor loadings), IRT analyses often yield a notably different ranking of item quality than correlational methods (e.g., Funk & Rogge, 2007), particularly among items that demonstrate fairly strong correlations with the construct being measured (e.g., factor loadings $\geq .65$). Thus, IRT offers a powerful tool for creating psychometrically optimized scales. In fact, it has been used to create psychometrically optimized measures of adult attachment (Fraley, Waller, & Brennan, 2000) and relationship satisfaction (Funk & Rogge, 2007) that offer higher precision and power than existing scales. Recent work has suggested that the increased cross-sectional precision of IRT-derived scales can offer researchers greater longitudinal precision for detecting change over time, creating scales that are not only more responsive to detecting significant individual change over time but that also yield stronger treatment effects (Rogge, Crasta, Maniaci, Funk, & Lee, 2015).

To apply IRT to the task of optimizing the assessment of sexual quality, a pool of 139 potential items (including the items of the existing measures) were given to an online sample of 3060 respondents. We evaluated the entire item pool, including the existing scales, using both EFA and IRT analyses.

Hypothesis 1: We hypothesized that given their heterogeneity of item content, the existing scales would contain

items fairly uncorrelated with the construct of sexual quality. Any problematic items (as assessed with internal consistency analyses, EFA, and IRT) would largely add error variance (or worse, confounding variance) to the existing scales, weakening the quality of information provided by them.

Hypothesis 2: We further hypothesized that sexual quality might best be represented by two distinct dimensions of sexual satisfaction and sexual dissatisfaction. Based on these first two hypotheses, the study also sought to use the EFA and two separate sets of subsequent IRT analyses to create the two subscales of a new measure of sexual quality—the Quality of Sex Inventory or QSI. To accommodate a wide range of applications (from treatment studies requiring the increased precision offered by longer 12-item subscales to shorter studies requiring the brevity of 6-item subscales), we created two versions of the QSI: a 24-item scale (with two separate 12-item subscales measuring sexual satisfaction and dissatisfaction) and a 12-item scale (with two separate 6-item subscales measuring sexual satisfaction and dissatisfaction).

Hypothesis 3: We hypothesized that the IRT-optimized QSI subscales would offer greater cross-sectional precision and power for detecting subtle group differences compared to the existing measures.

Hypothesis 4: We also hypothesized that the QSI subscales would demonstrate stronger psychometric properties (i.e., test–retest reliabilities and internal consistencies) across a variety of subsamples.

Hypothesis 5: Finally, given the anticipated use of these scales to examine change in sexual relationships, we sought to extend previous measurement work by directly examining the responsiveness of the various sexual quality measures for detecting change over time. We hypothesized that the QSI scales would demonstrate high levels of responsiveness to change over time given their increased cross-sectional precision.

Method

Participants

A total of 3060 respondents currently in sexually active romantic relationships completed an online survey. The participants were predominantly female (65 %) and Caucasian (75 %) with 5.5 % African American, 6.1 % Latino, 8.0 % Asian, and 5.9 % other/multi-racial. The mean age was 27.0 years old ($SD = 9.4$). The average income was \$21,270 per year ($SD = \$21,272$). A majority of the participants attended some type of college (46 % some college or trade school, 29 % bachelor's degree, 12 % graduate degree) while 13 % completed high school or less. Close to half of the respondents (47 %) reported currently living with their romantic partners, with another 14 % frequently spending nights with their partners and 12 % spending most nights with their partners. Consistent with this, 54 % were dating their partners exclusively

(for an average of 1.8 years, $SD = 2.3$), 29 % were married (together for an average of 10.4 years, $SD = 8.3$ and married an average of 7.8 years, $SD = 8.5$), 8.3 % were engaged (together an average of 3.4 years, $SD = 3.0$), and 6.7 % were dating casually (dating more than one person for an average of 0.7 years, $SD = 1.1$). The sample was modestly happy with mean relationship satisfaction (CSI-16; Funk & Rogge, 2007) scores of 62.9 (on a 0–81 scale, $SD = 13.8$) for exclusively dating participants, 60.0 ($SD = 16.7$) for married participants, 65.9 ($SD = 13.4$) for engaged participants, and 49.4 ($SD = 13.9$) for casually dating participants. Using a cut score of 51.5 for identifying clinically meaningful relationship dissatisfaction (see Funk & Rogge, 2007), the CSI-16 scores revealed that 22 % of the total sample and more specifically, 53 % of the casually dating, 19 % of the exclusively dating, 14 % of the engaged, and 25 % of the married respondents were significantly dissatisfied in their relationships, suggesting that the sample contained relationships representing a wide range of relationship quality.

Procedure

Recruitment

All procedures and materials for this study were approved by an Institutional Review Board. Respondents had to be at least 18 years of age and currently in sexually active romantic relationships to participate. The survey was presented online via Survey Gizmo.com and took approximately 20–25 min to complete. The study was advertised as “The Quality of Sex Study,” and the recruitment web-postings and emails emphasized that it was reasonably short and voluntary, offered individuals normative feedback on their responses, and included an optional follow-up survey. Participants were recruited from online websites (40 %; e.g., Craigslist, University of Hanover, TheNest.com), the Mechanical Turk service of Amazon.com (32 %), the psychology undergraduate subject pool (20 %), and recommendations from other participants (8.4 %). All respondents received feedback on their relationships at the end of the initial survey as the primary recruitment incentive. Respondents who participated through Mechanical Turk (Mturk) also received 85 cents to \$1.52 of store credit for Amazon.com, and respondents from the undergraduate subject pool received extra credit toward their psychology courses through an online interface (SONA). These additional recruitment incentives were distributed entirely via ID codes within those two systems (Mturk and SONA) in an effort to protect the privacy of the respondents. Thus, the only personal information collected in the survey was the email addresses that the respondents could choose to provide. Those email addresses were promptly deleted after all respondents had been invited to the follow-up surveys, thereby de-identifying the data to ensure the ongoing confidentiality of participants' responses.

Two-Month Follow-Up

Participants were given the option of providing email addresses in the initial survey so that they could be invited to complete an optional 2-month follow-up assessment, and 2978 respondents (97%) did so. They were sent a series of up to 4 invitations to complete the follow-up (spaced roughly 7 days apart), and 869 participants (29% of those providing email addresses) did so. ANOVA analyses failed to uncover any relationship satisfaction (CSI) or sexual satisfaction (QSI) differences between participants providing follow-up data and those who did not. The analyses suggested small demographic differences, indicating that the participants providing follow-up responses tended to be slightly older, $F(1, 3031) = 53.0, p < .001, d = 0.29$, with slightly higher education, $F(1, 3047) = 58.1, p < .001, d = 0.31$, and income, $F(1, 2918) = 7.6, p < .006, d = 0.11$, and slightly longer relationships, $F(1, 3038) = 26.5, p < .001, d = 0.21$.

Measures

Couples Satisfaction Index (CSI-16)

The CSI-16 (Funk & Rogge, 2007) is a 16-item measure of relationship satisfaction. The items assess global evaluations of relationship quality (see Table 1), and were rated on 6- and 7-point scales. Responses were totaled so that higher scores reflected higher levels of relationship satisfaction ($\alpha = .97$).

Index of Sexual Satisfaction (ISS)

Roughly half of the ISS items are worded in a positive direction (e.g., “I feel that my partner enjoys our sex life”) and the rest are worded in a negative direction (e.g., “My partner is too rough or brutal when we have sex”). The items were rated on a 6-point response scale (“Not at all TRUE” to “Completely TRUE”). In contrast to its name, the items of the ISS are typically combined into a single score (ranging from 0 to 100) such that higher scores reflect higher levels of *dissatisfaction*. However, in the current study, the items were recorded and summed to create a total (still ranging from 0 to 100) so that higher scores reflect higher sexual *satisfaction* to match the direction of the other existing sexual quality measures ($\alpha = .93$).

Global Measure of Sexual Satisfaction (GMSEX)

The GMSEX (Lawrance & Byers, 1995) contains 5 semantic differential items on a 7-point scale (e.g., “bad–good,” “unpleasant–pleasant,” “negative–positive,” “unsatisfying–satisfying,” and “worthless–valuable”). Responses were summed so that higher scores reflected higher levels of sexual satisfaction ($\alpha = .95$).

Pinney Sexual Satisfaction Inventory (PSSI)

The PSSI (Pinney et al., 1987) is a 24-item measure of sexual satisfaction (e.g., “I feel that nothing is lacking in my sex life,” “I am satisfied that my physical needs are completely met during lovemaking,” “I am satisfied with the amount of foreplay involved in my lovemaking”). The items were rated on a 7-point response scale (“Strongly Disagree” to “Strongly Agree”). Responses were summed so that higher scores indicated higher levels of sexual satisfaction ($\alpha = .94$).

Young Sexual Satisfaction Scale (YSSS)

The YSSS (Young et al., 1998) is an 11-item measure of sexual satisfaction (e.g., “I am satisfied with my spouse as a sexual partner,” “After sex I feel relaxed, fulfilled,” “I have satisfying orgasms”). The items were rated on a 5-point response scale (“Strongly Disagree” to “Strongly Agree”). Responses were summed so that higher scores indicated higher levels of sexual satisfaction ($\alpha = .91$).

Additional 74 Sexual Quality Items

To diversify the item pool beyond the contents of the existing scales, the authors wrote an additional 74 items that made use of a 6-point response scale (“Not at all TRUE” to “Completely TRUE”). A majority of these items were written from the context of our conceptual definition, and so we aimed to write simple and clear items that might serve as prototypical items for both positive (e.g., *sexual activity with my partner is... fulfilling, fantastic, rewarding*) and negative global sentiments (e.g., *sexual activity with my partner is... not fun, a turn off, not very exciting*). We also kept the items reasonably short using no more than an 8th grade reading level in an effort to minimize their burden on respondents and maximize comprehension. When writing these items, we drew upon the second author’s expertise in developing and validating global evaluative scales (e.g., Funk & Rogge, 2007; Rogge et al., 2015). We made the decision to write a larger pool of items and then let our EFA and IRT analyses in the 3060 online respondents be the deciding factor in whether any of those items were worthy of being included in the final QSI scales, obviating the need for any further vetting process. Although this required more time on the part of our respondents (completing the entire pool of 74 extra items), it ensured that any decisions to retain or drop items would be based solely on psychometrics, as even decisions by seasoned experts are subject to potential biases and errors. This was a highly effective process as 21 of the top 24 items that made it onto the QSI scales came from this item pool.

Table 1 Pattern matrix from an exploratory factor analysis

Item text	QSI subscales		Boundary constructs	
	F1	F2	F3	F4
F1: QSI sexual satisfaction subscale (provided as samples of items loading on this factor)				
My sex life is fulfilling	.77	−.03	−.13	.13
I am happy with my sex life with my partner	.71	−.05	−.16	.15
I am satisfied with our sexual relationship	.73	−.02	−.20	.11
I am happy with the quality of sexual activity in our relationship	.74	−.03	−.13	.13
Sexual activity with my partner is fantastic	.76	−.01	−.02	.14
I am happy with my partner as a lover	.65	−.11	−.04	.22
Sexual activity with my partner is rewarding	.70	−.06	.00	.20
Sexual activity with my partner is enjoyable	.72	−.12	.06	.15
My sex life is very exciting	.79	.03	−.14	.03
Sexual activity with my partner is everything I could hope for	.74	.05	−.11	.15
Sex is fun for my partner and me	.68	−.15	−.02	.10
F2: QSI sexual dissatisfaction subscale (provided as samples of items loading on this factor)				
Sexual activity with my partner is not fun	−.09	.74	.11	−.03
Sexual activity with my partner is a turn off	.10	.78	.04	−.03
Sexual activity with my partner is not worth the time or effort	.07	.79	.07	−.05
I do NOT enjoy sexual activity with my partner	−.06	.73	−.02	−.04
Sexual activity with my partner leaves me empty	.07	.73	.11	−.16
Sexual activity with my partner is not very exciting	−.23	.58	.22	−.04
I would rather not engage in sexual activity with my partner	.05	.82	−.05	−.03
I don't look forward to sexual activity with my partner	−.05	.69	−.02	.00
My sex life with my partner has become somewhat dull	−.26	.49	.32	−.07
I am tired of engaging in sexual activity with my partner	−.07	.71	−.04	−.05
Sexual activity with my partner leaves me feeling distant and alone	.13	.72	.13	−.17
I am very disappointed with my sex life with my partner	−.20	.50	.33	−.10
F3: sexual desires items (NOT part of the QSI, included to enforce a conceptual boundary)				
I want more sexual activity with my partner	.09	−.06	.67	.01
I wish my partner initiated sex more often (PSSI)	−.04	−.09	.64	−.02
My partner does not want sex when I do (ISS)	.03	.20	.64	−.12
I wish my partner was more passionate about sexual activity	−.12	.19	.61	−.07
I feel that I should have sex more often (ISS)	−.08	.04	.60	.00
I wish my partner was more attentive to my sexual needs	−.24	.24	.53	−.04
I wish sexual activity with my partner had more variety	−.27	.22	.48	.05
My partner is not responsive to my sexual needs or desires	−.09	.41	.46	−.11
I wish my partner could communicate more openly about what he/she wants in our sexual encounters (PSSI)	−.15	.00	.44	−.03
I am satisfied with the frequency with which I have sexual intercourse (PSSI)	.46	.15	−.45	.05
I am pleased with the frequency with which my spouse and I engage in sexual activity (YSSS)	.50	.15	−.43	.07
F4: sample relationship satisfaction items (NOT part of the QSI, a conceptual boundary)				
Our relationship is strong	.01	.03	.13	.89
I have a warm and comfortable relationship with my partner	.01	.01	.06	.88
In general, how satisfied are you with your relationship?	.10	.04	.05	.87
My relationship with my partner makes me happy	.07	.00	.09	.86
Please rate your romantic relationship from Bad to Good	−.04	−.02	.03	.86
How rewarding is your relationship with your partner?	.11	.03	.08	.85
Please rate your romantic relationship from Sturdy to Fragile	−.14	−.04	.05	.85
I really feel like part of a team with my partner	.00	.04	.08	.83

Table 1 continued

Item text	QSI subscales		Boundary constructs	
	F1	F2	F3	F4
Correlations among the Factors				
QSI subscales				
Factor 1: sexual satisfaction	–			
Factor 2: sexual dissatisfaction	–.44	–		
Conceptual boundary constructs (not on the QSI but included to distinguish the QSI scales from these related constructs)				
Factor 3: sexual desires	–.32	.11	–	
Factor 4: relationship satisfaction	.56	–.37	–.31	–

To facilitate presentation, only the items ultimately selected for the QSI scale (by IRT analyses) are presented for Factors 1 and 2, and only the first 8 items loading on Factor 4 are presented. However, this EFA was conducted on a set of 155 items. All correlations above .30 are bolded to facilitate interpretation

Frequency of and Desire for More Physical Affection

We used five items to assess physical affection (e.g., “Cuddling,” “Holding one another,” “Deep kissing,” “Kissing/caressing non-genital areas of partner’s body,” “Partner kissing/caressing non-genital areas of your body”). Respondents gave ratings on how frequently each activity occurred over the 2 weeks prior to completion of the survey and on how frequently they would have liked for it to happen on a 7-point scale (0 times to 13 or more times). The actual frequency ratings summed to create a frequency of affection composite ($\alpha = .94$) where higher scores indicated more frequent physical affection. The differences between actual and desired levels of each behavior were summed to create a desire for more physical affection composite ($\alpha = .92$) in which higher scores indicated a desire for more affection.

Frequency of and Desire for More Sexual Activity

We used 5 items to assess sexual activity (e.g., “Giving oral sex,” “Receiving oral sex,” “Vaginal and/or anal sex,” “Climaxing with partner,” “Partner climaxing with you”). Respondents gave ratings on how frequently each activity occurred over the 2 weeks prior to completion of the survey and on how frequently they would have liked for it to happen on a 7-point scale (0 times to 13 or more times). The actual frequency ratings were summed to create a frequency of sexual activity composite ($\alpha = .90$) in which higher scores indicated more frequent activity. The differences between actual and desired levels of each behavior were summed to create a desire for more sexual activity composite ($\alpha = .89$).

Sociosexual Orientation Inventory—Revised (SOI-R)

The SOI-R (Penke & Asendorpf, 2008) is a 9-item measure of attitudes toward (e.g., “Sex without love is ok”), desires for, (e.g.,

“How often do you have fantasies about having sex with someone you are not in a committed romantic relationship with?”), and experience with (e.g., “With how many different partners have you had sexual intercourse on one and only one occasion?”) casual sex. The items were rated on 9-point scales and were totaled so that higher scores reflected higher interest in casual sex ($\alpha = .86$).

Libido

Four items were used to assess individuals’ sex drives (e.g., “I tend to be horny most of the time,” “I find myself craving sex often,” “My mind often wanders to sex,” “I can get turned on very quickly”). The items were rated on a 5-point response scale (“Not at all TRUE” to “Very TRUE”) and were totaled so that higher scores reflected higher libido ($\alpha = .90$). These items were written to be prototypical of the construct of interest (conceptually focusing on the drive to have sex; see Baumeister, Catanese, & Vohs, 2001), were part of a larger set of 9 items that were piloted in a sample of 385 online respondents, and were the items showing the strongest corrected item-to-total and squared multiple correlations in that earlier sample (suggesting that they were also very internally consistent in addition to being conceptually focused).

Physical Attraction to Partner

A single item was used to assess each respondent’s physical attraction to their romantic partner (e.g., “How physically attractive do you find your partner?”). Respondents answered this question on a 6-point scale (“Not at all” to “Extremely”).

Negative Conflict Behavior

We used four items to assess negative conflict behaviors (e.g., “Swear at your partner,” “Yell and scream at your partner,”

“Use profanity toward your partner,” “Purposefully insult your partner”). The items were rated on a 6-point scale (“Never” to “All the time”) and were totaled so that higher scores reflected more negative conflict ($\alpha = .89$). These items were part of a larger 100-item pool evaluated in 3303 online respondents, and were specifically items that loaded cleanly on a negative conflict factor (as opposed to psychological or physical aggression factors), offered the highest information as assessed by subsequent IRT analyses, and as a scale, demonstrated acceptably high correlations with standard measures of negative conflict like the Marital Coping Inventory—conflict subscale ($r = .75$; Walsh, Rodrigues, & Rogge, 2015).

Emotional Support

We used four items from the 25-item SIRRS (Dehle, Larson, & Landers, 2001) to assess emotional support received from a romantic partner (e.g., “Said he/she thought I handled a situation well,” “Expressed confidence in my ability to handle a situation,” “Said it was ok to feel the way I was feeling,” “Took my side when discussing a situation”). These four items have been previously identified as forming an internally consistent emotional support factor within the SIRRS (e.g., Barry, Bunde, Lawrence, & Brock, 2009). Respondents reported how often their partner engaged in each support behavior in the last 2 weeks prior to completing the survey on an 8-point scale (0 times to 7 or more times). Responses were summed so that higher scores reflected more emotional support ($\alpha = .91$).

Results

Examining Existing Sexual Quality Scales

A majority of the items of the existing scales were worded to specifically assess sexual satisfaction (as reflected by their scale names) and were designed to yield a single score—treating sexual quality as a single dimension. As a result, when evaluating the existing sexual quality scales, we focused on sexual satisfaction as the underlying construct. To comprehensively evaluate the quality of information provided by the sexual satisfaction items as a set, we performed an IRT analysis of the 65 items comprising the existing scales along with 41 additional items identified as assessing sexual satisfaction by the EFA analyses (described below). Evaluating these 106 items simultaneously (1) improved the quality of the IRT analysis by providing a large number of items for estimating subjects’ latent levels of sexual satisfaction (stabilizing the solution by improving those theta estimates), and (2) helped ensure that the resulting item parameters (as well as the item and test information curves derived from those parameters)

were on a common scale, allowing for direct comparison. To perform the IRT analyses, GRM parameters for the 106 items were simultaneously estimated with Multilog 7.0 (Thissen, Chen, & Bock, 2002) using marginal maximum likelihood estimation.

Information Provided by Existing Scales

As mentioned above, IRT analyses generate item information curves (IICs) which can be added together to create test information curves (TICs) for sets of items. In this analysis, the information curves visually displayed how much information a set of items provided for assessing sexual satisfaction across a broad range of that construct (from 3 standard deviations below to 3 standard deviations above the mean). Information curves that were further away from the x-axis indicated that the scale in question provided more information/precision for assessing sexual satisfaction for individuals in that range. As seen in Fig. 1a, although the IICs identified many of the items of the existing scales as effective for assessing sexual satisfaction (e.g., the ISS item “My sex life is very exciting”), they also suggested that many of the items of existing scales offered little to no information for assessing sexual satisfaction (e.g., the ISS item “I feel that sex is something that has to be endured in our relationship,” the PSSI item “I wish I were less inhibited when I make love”). These IRT results suggested that including such items on sexual quality scales would largely contribute error variance (i.e., noise) to the scale. Consistent with this, the TICs for the existing scales (represented by the various dashed lines in the right panel of Fig. 1a) suggested that the 25-item ISS and the 24-item PSSI provided only modest amounts of information for assessing sexual satisfaction despite containing the largest numbers of items (leading to the expectation that they should be providing the largest amounts of information). These results offered support for Hypothesis 1, suggesting that the existing scales might contain suboptimal items contributing primarily error variance to the assessment of sexual quality. In contrast, the IRT analyses suggested that the 5-item GMSEX was a particularly effective measure of sexual satisfaction for a majority of the range of this construct, offering researchers a little more than half the information of the ISS and PSSI with the use of 1/5th the number of items. In comparison, the TIC for the YSSS suggested that the 11-item scale offered less information for assessing sexual satisfaction across most of the range compared to the 5-item GMSEX scale. Taken as a set, these analyses call into question the quality of the ISS, YSSS, and PSSI scales as they failed to markedly outperform the 5-item GMSEX. The information curves as a set also suggested that the amount of information provided sharply drops off at the highest levels of sexual satisfaction. This is most likely due to a ceiling effect as people with latent sexual satisfaction greater than 1.5 standard deviations above the mean generally tend to have nearly perfect scores on these measures—reducing the ability of the measures to distinguish such respondents from one another. Taken as a

set, the IRT analyses of the existing sexual quality measures suggested the potential for being able to improve upon these scales.

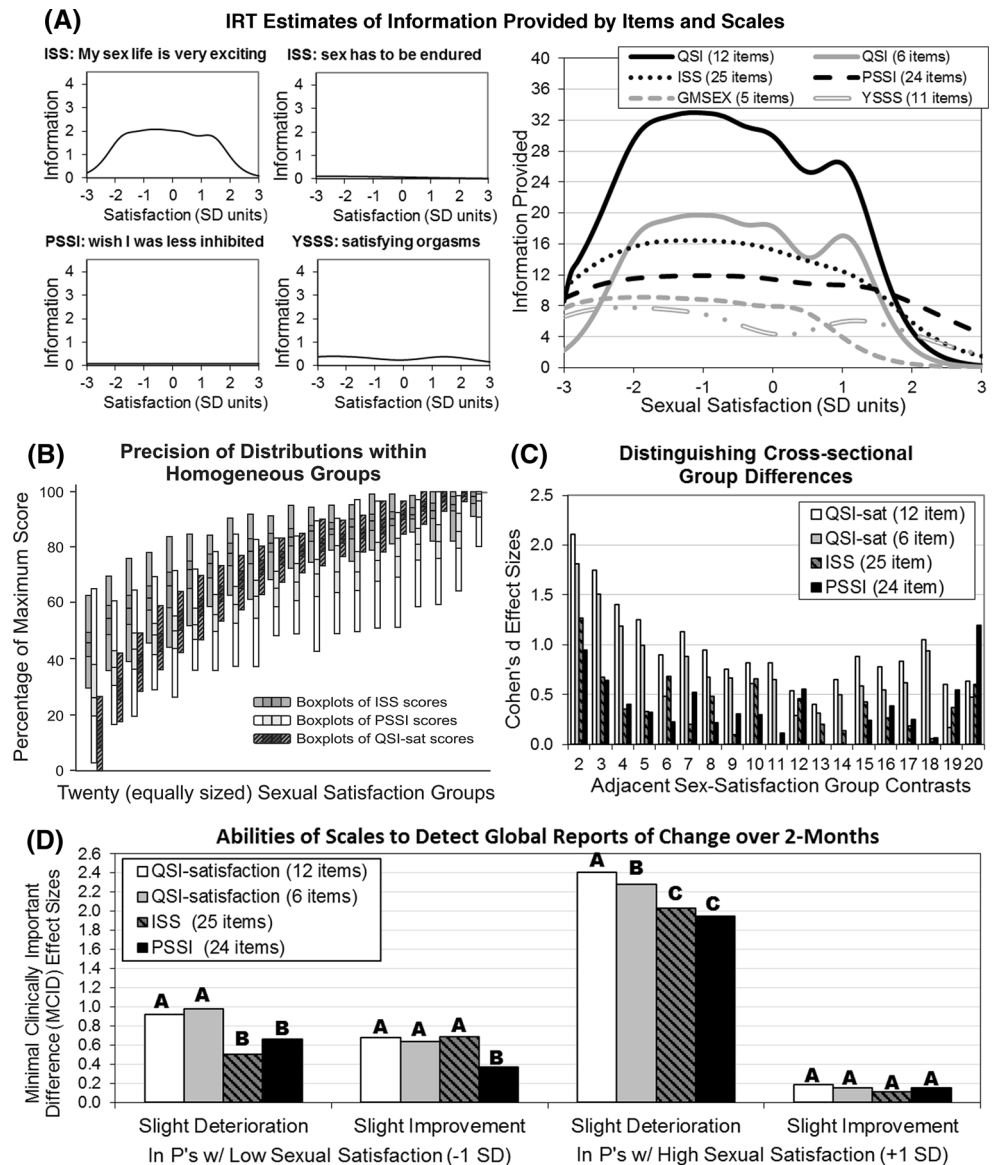
Equivalence of Items Across Subgroups

To determine if any of the items might function differently in subgroups of respondents, we ran separate IRT models in subpopulations (men vs. women; married/engaged vs. dating). Although we found the item parameters to be reasonably stable for a majority of the items ($r = .90$ for men vs. women and $r = .88$ for married-engaged vs. dating), a number of items seem to have been interpreted more stringently/leniently across subsamples, thereby demonstrating differential item functioning (DIF).

DIF Across Gender

The IRT analyses suggested that there were 5 ISS items demonstrating DIF across gender groups such that men tended to be at least a standard deviation less satisfied than women (i.e., having estimated beta thresholds between response options an average of 1 SD or more lower in men than in women) before selecting comparable answer choices indicating lower satisfaction (e.g., “I feel that my partner wants too much sex from me,” “My partner dwells on sex too much,” “I feel that sex is wonderful”—reversed). The IRT analyses suggested men in the sample had to be notably lower than women on sexual satisfaction before they endorsed stronger agreement with such items. The IRT analyses also identified 6 ISS, 2 PSSI, and 2 YSSS items for which women had to be notably lower than men on sexual satisfaction before they endorsed

Fig. 1 Comparison of Sexual Satisfaction Scales on **a** IRT estimates of Information Provided, **b** Precision, **c** Corresponding Power, and **d** Responsiveness to Change over Time. *QSI-sat* Quality of Sex Inventory, satisfaction subscale, *ISS* Index of Sexual Satisfaction, *PSSI* Pinney Sexual Satisfaction Inventory, *YSSS* Young Sexual Satisfaction Scale, *GMSEX* Global Measure of Sexual Satisfaction, *SD* Standard Deviation, *P’s w/* Participants with. In **d**, different letters indicate significant differences as tested via the approach of Meng, Rosenthal, and Rubin (1992)



the less-satisfied responses on those items (e.g., “*My partner does not want sex when I do*,” “*I wish my partner initiated sex more often*”). Although all of these items seemed to be measuring sexual satisfaction to the same degree in both men and women (giving roughly similar item discrimination parameters), the DIF findings suggested that men and women with the same levels of sexual satisfaction tended to select different responses on these items. Thus, the results suggested that such items run the risk of creating artificial differences in sexual satisfaction across men and women if included in sexual satisfaction measures. In constructing the QSI scales, we avoided including any items demonstrating DIF across genders. As a result, the QSI scales should be free from DIF, and so scores in male respondents should be conceptually identical to those obtained from female respondents.

DIF Across Relationship Stages

The IRT subgroup analyses also suggested that there were 6 PSSI, 5 ISS, and 4 YSSS items for which dating respondents had to be markedly lower on sexual satisfaction (when compared to married and engaged respondents) to endorse the less-satisfied responses (e.g., “*My partner does not want sex when I do*,” “*I feel that I should have sex more often*”). The IRT subgroup analyses further identified 7 ISS items for which married respondents had to be notably lower to endorse the less-satisfied responses (e.g., “*I feel that my partner wants too much sex from me*,” “*I feel that sex is dirty and disgusting*,” “*My partner dwells on sex too much*”). Taken as a set, the IRT analyses in subsamples primarily identified the items with the most heterogeneous content (potentially assessing distinct constructs from sexual satisfaction like desires/preferences and sexual attitudes/expectations) as having the most problematic levels of DIF. Thus, in addition to potentially being problematic at a conceptual/theoretical level, including such items might also adversely affect researchers’ ability to compare scores across these subgroups.

Evaluating Dimensionality in the Full Item Pool

As a first step in developing a psychometrically optimized measure of sexual quality, we ran an EFA on the pool of 139 sexual quality items using principle axis factoring with oblimin rotation (allowing the factors to correlate). Given the heterogeneity of item content in the existing scales, this EFA allowed us to clarify how many distinct constructs were represented in the item pool. We also included the 16 items of the CSI in this EFA to extract a distinct relationship satisfaction factor. By including these items assessing a distinct construct (not intended to be a part of the final QSI scale) and subsequently extracting that as a separate factor, we were able to examine how each item in the pool correlated with romantic satisfaction in comparison to how it correlated specifically with sexual satisfaction or dissatisfaction. This allowed

us to screen out items that might actually be assessing relationship satisfaction rather than sexual satisfaction or dissatisfaction, thereby providing a tool to discriminate sexual quality items from the closely related construct of relationship quality (directly enforcing that conceptual boundary). The scree plot suggested four clear factors with eigenvalues of 64.3, 10.6, 6.4, and 5.9, yielding a four-factor solution that accounted for 55 % of the variance. As seen in Table 1, the items of the CSI extracted as a single-relationship satisfaction factor (Factor 4) and the remaining items from the pool split into three distinct factors—sexual satisfaction, sexual dissatisfaction, and sexual desire. Consistent with Hypothesis 2, the quality of sexual relationships was represented by separate positive (sexual satisfaction—Factor 1) and negative (sexual dissatisfaction—Factor 2) dimensions that only demonstrated a moderate negative correlation with one another ($r = -.44$). In fact, the reverse-worded items of the ISS demonstrated clear loadings on the sexual dissatisfaction dimension despite the fact that those items are routinely combined with the ISS sexual satisfaction items to generate a single-total score. This finding mirrored similar findings on global relationship quality (Mattson et al., 2012) and positive/negative mood (e.g., the PANAS; Watson, Clark, & Tellegen, 1988), suggesting either that individuals can hold both positive and negative feelings/evaluations simultaneously and somewhat independently or at least that individuals can vacillate between states of satisfaction and dissatisfaction somewhat independently. Consequently, it is likely more informative to keep those dimensions separate rather than collapsing them into a single-global evaluation, as is the current practice. Consistent with concerns over the heterogeneity of content in existing sexual quality scales, a third-distinct factor emerged representing individuals’ desires for more frequent, passionate, and/or responsive sex with their partners. Notably, this factor included items from the PSSI, the YSSS, and the ISS, suggesting that the total scores for those scales were somewhat contaminated by the variance of this distinct dimension. The EFA analyses further revealed that 2 of the items of the PSSI actually loaded more strongly on the relationship satisfaction factor than on the sexual satisfaction factor, suggesting that those items more appropriately belonged on measures of relationship quality, not sexual quality. Thus, by using factors 3 and 4 to identify items not purely measuring sexual satisfaction or dissatisfaction, we were able to enforce the conceptual boundaries we established at the start of the study.

To ensure that this factor structure was comparable across men and women and across relationships stages (e.g., dating vs. engaged or married), we ran secondary EFAs in those subsamples. All of these analyses yielded 4 factor solutions with nearly identical factor loading patterns suggesting that the factor structure was comparable across those subgroups. To be clear, the QSI consists of only two factors—sexual satisfaction and sexual dissatisfaction. As mentioned above, the relationship satisfaction factor and the sexual desire factor served as conceptual boundaries to ensure that the QSI did *not* include heterogeneous items measuring either of those distinct constructs.

Creating a Two-Dimensional Measure of Sexual Satisfaction

Selecting Items with EFA and IRT

To identify separate sets of positive and negative items based on the EFA, we identified the 81 sexual satisfaction items and the 31 sexual dissatisfaction items that correlated at least .40 with their respective factors and more strongly with their respective factors than with any of the other factors. We chose a minimum structural coefficient of .40 to follow current best practices for interpreting EFAs (see Clark & Watson, 1995; Floyd & Widaman, 1995) as that ensured that the items retained share at least 16 % of variance with the construct being measured. As mentioned above, the positive items were combined with the items of the existing scales and evaluated in one comprehensive 106-item IRT analysis. Although this included some of the negatively worded items of existing scales that formed a separate dissatisfaction factor, researchers are currently using those items as sexual satisfaction items when they are combined into totals on the existing unidimensional scales. Thus, to fully evaluate the existing scales and to estimate the information assessing sexual satisfaction that those negatively worded items offer *as they are currently used in the field*, we included those negatively worded items in the sexual satisfaction IRT analysis. Supporting this strategy, the negatively worded items demonstrated moderate correlations (e.g., .40 to .60) with the sexual satisfaction factor despite the fact that they correlated more strongly with the sexual dissatisfaction factor, suggesting that they at least partially were assessing sexual satisfaction. We ran a separate IRT analysis on the 36 dissatisfaction items: the 13 sexual dissatisfaction items of the ISS and the 23 additional sexual dissatisfaction items identified in the EFA. Thus, in the separate dissatisfaction IRT analysis, we gave the negatively worded items from the ISS a second chance—re-evaluating them alongside items truly measuring the same construct. This strategy modeled those items both as they are currently being used (as satisfaction items) and as our EFAs suggested they should be used (as dissatisfaction items). IICs from the IRT models were evaluated to identify the 12 sexual satisfaction items and the 12 sexual dissatisfaction items offering the highest levels of information (i.e., greatest precision, lowest noise) for assessing these dimensions (see Table 1 for the actual items selected). Provided that scale developers select high-quality and non-redundant items, longer scales offer greater precision (e.g., lower noise, higher power) to see differences between respondents cross-sectionally (e.g., Funk & Rogge, 2007). That cross-sectional precision can translate into a greater responsiveness to detecting meaningful change over time, even yielding stronger effect sizes in small treatment studies (e.g., Rogge et al., 2015). Based upon this work, we chose to create and validate two 12-item subscales for the QSI to offer researchers measures that were not too lengthy but that would offer greater amounts of precision—particularly in small samples. A majority (21) of the

24 items on the QSI scales came from the additional item pool written by the authors, with just one sexual satisfaction item (“*My sex life is exciting*”) from the ISS and two semantic differential items (“*bad–good*,” “*unpleasant–pleasant*”) from the GMSEX. Recognizing that not all researchers would have the same needs, we also chose to create and validate shorter 6-item versions of the QSI subscales, offering researchers a much briefer method of assessing sexual quality. Thus, once the 24 items of the QSI had been identified, shorter versions of the measure were created by identifying the 6 items within each 12-item subscale that provided the largest amount of information—thereby creating a 12-item version of the scale (presented as the first 6 items for Factors 1 and 2 in Table 1). The items of the shorter QSI were all from the item pool generated by the authors, sharing no items in common with the existing scales. Thus, the item overlap between the existing scales and the new QSI scales was minimal. After the items of the QSI scales had been selected, responses were given point values from 0 to 5 indicating higher levels of satisfaction or dissatisfaction for the appropriate subscales, and were then summed to create total scores.

Equivalence Across Subgroups

IRT analyses in distinct subgroups of respondents suggested that the items of the QSI-satisfaction and the QSI-dissatisfaction subscales operated comparably across men and women and across different relationship stages.

Information Provided by the QSI-Satisfaction Subscales

The right panel of Fig. 1a presents TICs comparing the QSI sexual satisfaction subscales to the existing scales of sexual satisfaction (similar comparisons were not possible with the QSI sexual dissatisfaction subscales as no prior measures existed). These TICs suggested that the 12-item version of the QSI sexual satisfaction subscale (QSI-satisfaction) provided markedly greater amounts of information than the existing measures for all but the highest levels of satisfaction (i.e., people greater than 1.8 SDs above the mean). This suggested that the IRT-optimized 12-item QSI-satisfaction subscale offered more information than existing scales that have twice as many items. Even the 6-item QSI-satisfaction subscale provided impressive amounts of information, surpassing the information contributed by the 24-item PSSI, the 5-item GMSEX, and the 11-item YSSS, and rivaling the information contributed by the 25-item ISS, for most of the range of latent scores examined (roughly from 2.1 SDs below the mean to 1.6 SDs above the mean, covering roughly 93 % of the sample) despite being much shorter. Thus, the QSI-6 provided equal if not better information for assessing sexual quality than the existing scales for all but roughly 7 % of the most extreme respondents (where having as many as four times the items offered a slight advantage).

Precision of the QSI Sexual Satisfaction Subscale

To further examine the increased precision afforded by the QSI scales, we grouped respondents into 20 groups based on their IRT-derived latent sexual satisfaction (θ) scores. As the respondents in each of these groups had highly similar levels of satisfaction, any scatter in their satisfaction scores would have largely been due to measurement error or noise in the scale being used. As seen in Fig. 1b, the distributions of 12-item QSI-satisfaction scores in each satisfaction group were not only tighter than those seen for the ISS and the PSSI (containing less noise), but as a set the distributions of QSI-satisfaction scores did a better job of spanning the entire numeric range of the measure than the ISS and PSSI distributions, offering a better range of discrimination at all levels of sexual satisfaction. These results were consistent with Hypothesis 3, visually representing the higher levels of noise in the existing scales due to the inclusion of less effective items.

Power Afforded by the QSI-Satisfaction Subscales

To determine if this increased precision of the QSI scales would translate to increased power for detecting subtle group differences, we calculated the effect sizes yielded by each measure (Cohen's d) for detecting differences between each sexual satisfaction group and the satisfaction group just below it. As seen in Fig. 1c, the 12-item QSI-satisfaction subscale offered markedly stronger effect sizes than the 6-item QSI-satisfaction, the ISS or the PSSI across almost the entire range of satisfaction groups. Converting these Cohen's d s into correlation coefficients and then testing for significant differences (using equations from Meng, Rosenthal, & Rubin, 1992) suggested that the 12-item QSI-satisfaction subscale yielded significantly stronger effect sizes for adjacent satisfaction group differences than the other scales on 56 of the 57 comparisons examined (yielding average z -scores of 10.75, suggesting robust average differences). Similarly, the 6-item QSI-satisfaction subscale yielded significantly stronger effect sizes than the ISS and PSSI on 33 of the 38 comparisons examined (yielding average z -scores of 5.68, suggesting strong average differences). These results suggested, consistent with Hypothesis 3, that both the 12- and 6-item QSI-satisfaction subscales provided greater precision of measurement (lower levels of noise) and correspondingly higher levels of power for detecting differences than the notably longer ISS and PSSI scales. The results further suggested that the longer 12-item QSI-satisfaction scale might offer researchers more precision and power than the 6-item QSI-satisfaction scale. Thus, researchers considering using the QSI as a primary outcome measure—particularly in small samples—might consider using the longer version of the scale, given the extra power it affords.

Test–Retest Reliabilities and Internal Consistencies of Scales

The IRT analyses presented in Fig. 1 offered compelling evidence to suggest that the QSI-satisfaction scale represents a far more effective measure than existing scales, offering researchers higher precision and power for detecting differences. We sought to build on the recent work of Mark et al. (2014) by extending our IRT analyses with the more traditional psychometric analyses of test–retest reliabilities and internal consistencies to further evaluate the QSI subscales in comparison to the existing scales.

Test–Retest Reliabilities

A scale can be considered reliable if individuals tend to receive highly similar scores upon re-testing. Mark et al. (2014) demonstrated that the ISS had slightly lower levels of test–retest reliability over 2 months. As seen in Table 3, analyses in the 696 respondents providing 2-month follow-up data in the current project provided similar results. An omnibus test (Woodruff & Feldt, 1986) for differences across the six test–retest correlations of the satisfaction scales revealed a significant difference, $\chi^2(5) = 111.5$, $p < .0001$, and subsequent pairwise comparisons (Abd-El-Fattah & Hassan, 2011) revealed that the 12-item QSI-satisfaction subscale had higher levels of test–retest reliability than the existing scales, despite being much shorter in length. Similar analyses in the 419 respondents reporting little to no change in their sexual satisfaction over time yielded stronger test–retest correlations for all of the scales (the follow-up survey did not include comparable questions about change in sexual dissatisfaction, and so it was not possible to calculate similar statistics for the QSI-dissatisfaction scales). The omnibus test for differences remained significant ($\chi^2(5) = 194.0$, $p < .0001$) and pairwise comparisons suggested that the 12-item QSI-satisfaction subscale continued to demonstrate higher test–retest reliability than the other scales. Thus, consistent with Hypothesis 4, the combined use of EFA and IRT to increase the precision of the QSI scales seemed to improve their test–retest reliabilities in comparison to existing scales, offering higher precision and stability across time.

Generalizability of Internal Consistencies

Although Mark et al. (2014) demonstrated that the GMSEX might be more internally consistent than other existing scales, their sample was predominantly White (89%), heterosexual (93%), and college educated (96%), preventing them from examining how the internal consistencies of existing scales might generalize across diverse populations. As shown in Table 3, the large size of the cur-

rent sample allowed us to address this by examining the internal consistencies of the sexual quality scales across a range of demographic subpopulations. As a set, the scales demonstrated high levels of internal consistency across all of the demographic groups examined, suggesting that all of the scales functioned well across those groups. Thus, these results help to validate the use of these scales across a wide range of sample diversity. However, comparable omnibus (Woodruff & Feldt, 1986) and pairwise comparisons (Abd-El-Fattah & Hassan, 2011) of these alpha coefficients suggested that the 12-item QSI-satisfaction subscale offered higher levels of internal consistency than the existing satisfaction scales in 18 of the 19 subgroups examined. Similarly, even the 6-item QSI-satisfaction subscale offered higher internal consistency than the existing satisfaction scales in 13 of the 19 subgroups, despite being markedly shorter in length. Thus, consistent with Hypothesis 4, the increased precision of the QSI-satisfaction scales (see Fig. 1) translated into higher levels of internal consistency, offering researchers psychometrically optimized scales. Although the internal consistencies of the QSI-dissatisfaction subscales are also presented, they were not statistically compared to the existing scales because they measure a distinct construct.

Responsiveness to Change over Time

Detecting Individual Change

To determine if the cross-sectional precision and power afforded by the QSI-satisfaction subscales translated into an effective ability to detect naturally occurring change over time, we evaluated the responsiveness of all the sexual satisfaction scales. Following current guidelines (e.g., Husted, Cook, Farewell, & Gladman, 2000), we first calculated the Standard Error of Repeated Measurement (SE_{RM}) for each of the scales. To do this, we began by identifying a ‘stable’ population: the 278 respondents who reported no change in their sexual satisfaction across the 2-month follow-up on two items (*Since the last assessment, how much has each of these changed, if at all: “Quality of sex” and “Overall satisfaction with your sexual relationship with your partner”*; $\alpha = .89$). We then ran repeated measures ANOVAs in those stable respondents to obtain estimates of noise over time for each scale within this set of respondents who reported no actual change (as suggested by Guyatt, Walter, & Norman, 1987). Specifically, we calculated SE_{RM} using the Mean Squared Error over time (MSE_{TIME}) estimates ($SE_{RM} = \text{SQRT}(2 \times MSE_{TIME})$). These SE_{RM} estimates were then transformed into Minimal Detectable Change (MDC_{95} ; Stratford et al., 1996) estimates with the following equation: $MDC_{95} = 1.962 \times SE_{RM}$. Based on the Reliable Change Index (Jacobson & Truax, 1991), MDC_{95} coefficients tell researchers and clinicians how many points an individual must change on a scale between two assessments for that change to be statistically significant. Thus, these coefficients allow individuals to be grouped into “significantly

improved,” “no change,” and “significantly deteriorated” categories based solely on pre–post or wave₁–wave₂ assessments, thereby offering researchers and clinicians powerful tools when using scales in longitudinal studies. As seen in Table 2, both the 6-item and 12-item versions of the QSI sexual satisfaction subscales required individuals to change just under 1 standard deviation between assessments for that change to be statistically significant, demonstrating responsiveness comparable to the existing measures despite being far shorter in length. Consistent with Hypothesis 5, this suggested that the increased cross-sectional precision of the QSI-satisfaction scales translated into high levels of responsiveness to individual change.

Responsiveness to Global Perceptions of Change

To further evaluate the responsiveness of various sexual satisfaction scales to detect change over time—this time at a group level—we calculated Minimal Clinically Important Difference (MCID; Guyatt et al., 1987) coefficients for the satisfaction scales in the study. Specifically, we ran regressions allowing self-reports of global change (on the two items described above, rated on a 7-point scale: *−3 has gotten MUCH worse, −2 has gotten somewhat worse, −1 has gotten a little worse, 0 stayed the same, +1 has gotten a little better, +2 has gotten somewhat better, +3 has gotten MUCH better*) to predict raw change scores on each scale. This allowed us to estimate the average number of points of change that each scale demonstrated in response to a single point of global change on that *−3* to *+3* response scale. To allow for the possibility that detecting improvement might differ from detecting deterioration, we included a dichotomous variable coding direction of change as a moderator. As it might also be easier to detect improvement in respondents with low levels of initial sexual satisfaction and to detect deterioration in respondents with high levels of initial sexual satisfaction, we also included T0 sexual satisfaction as a predictor of the contrast between detecting deterioration and improvement. The unstandardized β coefficients from these regressions were all statistically significant and yielded estimates of how many points of change could be expected on each scale for one point (i.e., “slight” amounts) of improvement versus deterioration in respondents starting with low versus high initial sexual satisfaction (i.e., 1 SD below vs. above the mean). To place these predicted average raw change scores onto a common metric, we converted them into effect sizes for each scale by dividing by the scale’s SE_{RM} . As seen in Fig. 1d, the QSI-satisfaction subscales had comparable MCID effect sizes to those of the ISS and PSSI despite being markedly shorter in length. In fact, when the magnitude of these MCID effects were statistically compared (as indicated by the letters above the effects, with different letters indicating significant differences), the QSI-satisfaction subscales—even just the 6-item version—yielded significantly stronger MCID effects for detecting slight deterioration

Table 2 Psychometric properties of the sexual quality scales and the conceptually distinct anchor scales from the theoretical/empirical nomological net surrounding sexual quality

Sexual quality scales	Descriptive characteristics						α	MDC ₉₅	Correlations among scales (in men: above the diagonal, in women: below the diagonal)								
	Possible range	In men		In women		Raw units			SD units	1	2	3	4	5	6	7	8
		M	SD	M	SD												
1. ISS (25)	0–100	75.6	16.1	78.6	15.7	.93	14.6	0.9	–	.78	.78	.71	.83	.81	–.78	–.72	
2. YSSS (11)	1–5	3.9	0.8	4.0	0.7	.91	0.85	1.1	.82	–	.83	.76	.87	.86	–.48	–.43	
3. PSSI (24)	24–168	114.2	26.8	116.8	27.4	.94	30.2	1.1	.78	.83	–	.74	.83	.82	–.47	–.42	
4. GMSEX (5)	5–35	29.3	6.7	29.9	6.3	.95	7.9	1.2	.80	.80	.74	–	.83	.83	–.45	–.39	
5. QSI-satisfaction (12)	0–60	40.5	15.0	41.5	15.1	.98	12.7	0.8	.90	.87	.81	.86	–	.99	–.47	–.41	
6. QSI-satisfaction (6)	0–30	20.0	8.0	20.6	8.0	.96	7.2	0.9	.88	.87	.81	.85	.99	–	–.46	–.40	
7. QSI-dissatisfaction (12)	0–60	6.9	10.8	4.9	9.3	.95	7.8 ^A	0.7	–.79	–.62	–.53	–.67	–.68	–.67	–	.97	
8. QSI-dissatisfaction (6)	0–30	3.0	5.4	2.1	4.5	.91	5.8 ^A	1.1	–.75	–.59	–.49	–.64	–.64	–.63	.98	–	
Anchor scales	Descriptive characteristics						Correlations with anchor scales in all respondents										
	Possible range	In men		In women		Sexual satisfaction scales	Sexual dissatisfaction	1	2	3	4	5	6	7	8		
		M	SD	M	SD												
Factor 3—sexual desires	0 to 30	13.5	7.7	10.2	6.9	–.57	–.53	–.70	–.49	–.51	–.52	.40		.36			
Factor 4—relationship satisfaction (CSI-16)	0 to 81	60.4	15.1	61.9	15.1	.63	.62	.60	.65	.67	.65	–.47		–.43			
Frequency of physical affection	0 to 65	32.6	18.9	35.4	19.7	.44	.45	.46	.43	.46	.45	–.26		–.23			
Desire for more physical affection	–65 to 65	9.9	14.5	10.6	14.9	–.18	–.22	–.31	–.23	–.20	–.21	.08		.06			
Frequency of sexual activity	0 to 65	17.4	14.2	17.2	14.3	.37	.46	.44	.39	.45	.45	–.14		–.10			
Desire for more sexual activity	–65 to 65	17.5	15.6	12.2	12.9	–.12	–.16	–.24	–.14	–.13	–.15	.05		.04			
Sociosexual orientation (SOI)	0 to 72	39.7	14.5	30.7	13.4	–.30	–.24	–.23	–.22	–.28	–.27	.26		.26			
Libido	0 to 16	10.7	3.6	8.6	4.4	.23	.27	.15	.20	.27	.26	–.09		–.09			
Physical attraction to partner	1 to 6	4.9	1.1	5.0	1.0	.54	.49	.38	.46	.55	.53	–.43		–.41			
Negative conflict	0 to 20	4.3	4.6	4.2	4.2	–.35	–.23	–.29	–.19	–.22	–.20	.22		.20			
Emotional support	0 to 28	13.3	7.5	12.6	8.0	.31	.37	.36	.32	.37	.36	–.11		–.10			
Living together	0 to 1	49%	–	46%	–	–.17	–.17	–.19	–.17	–.19	–.20	.13		.08			

α Cronbach's alpha coefficient, MDC₉₅ Minimum Detectable Change Index, ISS Index of Sexual Satisfaction, YSSS Young Sexual Satisfaction Scale, PSSI Pinney Sexual Satisfaction Inventory, GMSEX Global Measure of Sexual Satisfaction, QSI Quality of Sex Inventory, CSI Couples Satisfaction Index

^A The follow-up survey did not contain all of the items of the QSI-dissatisfaction scales. As a result, these two MDC₉₅ estimates were derived using equations recommended by Jacobson and Truax (1991)

than did the ISS and PSSI scales and outperformed the PSSI at detecting improvement in respondents starting the study with lower levels of sexual satisfaction. Taken as a set, these results continued to support Hypothesis 5, suggesting that the increased cross-sectional precision of the QSI-satisfaction subscales translated into high levels of responsiveness for detecting change.

Convergent and Discriminant Validity of the QSI Scales

The IRT, DIF, EFA, MDC₉₅, and MCID analyses suggested that the QSI scales might offer researchers and clinicians more precise and responsive tools for evaluating sexual quality. We conducted a final set of analyses examining the construct validity of the QSI scales to accomplish the following objectives: (1) to more firmly ground the QSI scales within the existing literature, (2) to help confirm that the QSI sexual satisfaction subscales continue to

measure the same basic construct assessed by previous scales, and (3) to offer evidence supporting our assertion that the QSI sexual dissatisfaction subscales are measuring a novel construct.

QSI Sexual Satisfaction Subscales

As seen in Tables 2 and 3, the QSI scales demonstrated excellent internal consistency. More importantly, the QSI scales demonstrated strong convergent validity with the existing measures of sexual satisfaction, showing appropriately strong correlations with those measures (Table 2) despite only sharing three items (out of 24 total items) with those existing scales. Furthermore, the QSI scales demonstrated nearly identical patterns of association with anchor scales to those seen with the existing measures of sexual satisfaction. Those correlations not only demonstrated discriminant validity from those anchor constructs, but they also suggest that the QSI

Table 3 Examining test–retest reliabilities and internal consistencies of the sexual quality scales across meaningful subsamples

Statistic by subsample	Subsample <i>N</i>	ISS	YSSS	PSSI	GMSEX	QSI-satisfaction		QSI-dissatisfaction	
						12-Item	6-Item	12-Item	6-Item
Test–retest reliabilities									
Full follow-up sample	696	.78 ^{BC}	.76 ^C	.79 ^{AB}	.72 ^D	.81 ^A	.80 ^B	–	–
No change group	419	.84 ^C	.81 ^{CD}	.83 ^C	.79 ^D	.89 ^A	.87 ^B	–	–
Cronbach's alpha coefficients (internal consistencies)									
By gender									
Men	1075	.92 ^E	.91 ^E	.94 ^D	.94 ^C	.96 ^A	.95 ^B	.93	.90
Women	1985	.92 ^E	.90 ^F	.93 ^D	.94 ^C	.97 ^A	.96 ^B	.94	.91
By living arrangements									
Living with partner	1425	.92 ^E	.91 ^F	.94 ^D	.95 ^C	.97 ^A	.96 ^B	.93	.90
Living separately	1352	.92 ^D	.89 ^E	.92 ^{CD}	.93 ^C	.96 ^A	.95 ^B	.94	.92
Long-distance relationships	270	.92 ^C	.89 ^D	.92 ^C	.93 ^{BC}	.96 ^A	.95 ^B	.94	.92
By race/ethnicity									
Caucasian	2271	.92 ^E	.90 ^F	.93 ^D	.94 ^C	.97 ^A	.96 ^B	.93	.89
Asian	245	.91 ^C	.92 ^C	.91 ^C	.94 ^B	.96 ^A	.94 ^B	.93	.92
Hispanic/Latino	188	.92 ^{CD}	.90 ^D	.94 ^C	.95 ^{BC}	.97 ^A	.96 ^B	.94	.91
African American	167	.93 ^C	.93 ^C	.94 ^{BC}	.95 ^B	.97 ^A	.96 ^B	.95	.92
By relationship stage									
Married	871	.92 ^E	.91 ^F	.94 ^D	.95 ^C	.97 ^A	.96 ^B	.94	.90
Engaged	253	.92 ^D	.89 ^E	.93 ^{CD}	.94 ^C	.97 ^A	.96 ^B	.92	.90
Dating	1936	.92 ^D	.90 ^E	.93 ^D	.94 ^C	.96 ^A	.95 ^B	.94	.91
By education level									
High school or less	403	.93 ^D	.91 ^E	.94 ^C	.95 ^C	.97 ^A	.96 ^B	.94	.91
Some college/trade school	1398	.92 ^E	.90 ^F	.93 ^D	.94 ^C	.97 ^A	.95 ^B	.93	.89
Bachelor's degree	881	.92 ^D	.91 ^E	.93 ^D	.94 ^C	.97 ^A	.96 ^B	.94	.92
Graduate degree	367	.92 ^{DE}	.91 ^E	.94 ^D	.95 ^C	.97 ^A	.96 ^B	.94	.91
By sexual orientation									
Heterosexual	2875	.92 ^E	.90 ^F	.93 ^D	.95 ^C	.97 ^A	.96 ^B	.94	.91
Gay males	64	.93 ^{BC}	.91 ^C	.94 ^{BC}	.95 ^B	.97 ^A	.95 ^B	.94	.91
Lesbians	121	.91 ^B	.85 ^C	.92 ^B	.93 ^{AB}	.95 ^A	.92 ^B	.91	.88

Differences in test–retest correlations and Cronbach's alphas among the six sexual satisfaction scales (the ISS, YSSS, PSSI, GMSEX, and QSI 12- and 6-item satisfaction scales) were first evaluated with an omnibus test recommended in Woodruff and Feldt (1986). These omnibus tests were all significant at $p < .0001$ (with $\chi^2(5)$'s ranging from 53 to 3981). We then conducted two sets of 15 pairwise comparisons of the test–retest correlations and 19 sets of 15 pairwise comparisons among alphas (one set within each subsample) using a test of dependent alphas (Abd-El-Fattah & Hassan, 2011). To control experiment-wide alpha levels, we used a Bonferroni correction, only considering pairwise test–retest contrasts significant at $p < .001966$ and only considering pairwise alpha contrasts significant at $p < .000196$. Alphas with different superscript letters represent significant differences from those pairwise comparisons

scales offer researchers conceptual equivalents to scales like the ISS and PSSI, assessing the same construct with dramatically enhanced precision. In fact, the correlation matrices presented in Table 2 demonstrated a slight short-coming of classical test theory that can be addressed by the addition of IRT analyses. At a test-level of analysis, the measures of sexual satisfaction were virtually indiscernible from one another, producing nearly identical correlational results within the nomological net surrounding sexual satisfaction. Although these highly similar correlations could have been due in part to shared items between the

existing and new scales, that explanation was less likely as only three of the final 24 items of the QSI scales came from existing measures. In contrast to the correlational results suggesting the equivalence of all of the scales, the IRT results presented in Fig. 1 told a very different story. By carefully modeling each item's performance at different levels of satisfaction, the IRT analyses revealed dramatic differences in the precision and power of measurement offered by the different satisfaction scales, providing compelling evidence of the advantages of using the QSI scales.

Table 4 Distinctiveness of the change groups revealed by the QSI-satisfaction and dissatisfaction subscales in comparison to the change identified by the ISS

Type of statistic presented	2-month change identified by the QSI sexual satisfaction subscale		
	Sig. worse	No change	Sig. better
2-month change identified by the QSI sexual dissatisfaction subscale			
Significantly worse			
# identified by QSI	16	26	–
% also identified by ISS	16 (100 %)	9 (35 %)	–
No change			
# identified by QSI	30	516	22
% also identified by ISS	18 (60 %)	473 (92 %)	14 (64 %)
Significantly better			
# identified by QSI	–	3	6
% also identified by ISS	–	2 (67 %)	5 (83 %)

Individuals were classified into specific change groups (significantly worse, no change, significantly better) by comparing their change scores on the QSI-satisfaction and dissatisfaction subscales to the MDC-95 coefficients for those scales. Individuals were also classified into change groups based on their ISS change scores to examine the degree to which the ISS classification was able to capture the information on outcomes provided by the QSI scales

QSI-Dissatisfaction Subscales

As seen in Table 2, the QSI-dissatisfaction subscales only demonstrated modest negative correlations with the QSI-satisfaction subscales (ranging from $r = -.40$ to $-.47$ in men and from $-.63$ to $-.68$ in women). This continued to suggest that sexual satisfaction and sexual dissatisfaction are likely measuring different constructs. Consistent with this, the QSI-dissatisfaction subscales showed distinct patterns of association with the anchor scales from the pattern seen for sexual satisfaction scales.

Distinct Dimensions of Change over Time

To examine how keeping sexual satisfaction and dissatisfaction as separate outcomes might yield unique information for researchers, we examined how these two dimensions changed over time across individuals. Specifically, we compared “significantly improved,” “no change,” and “significantly deteriorated” individuals based on the QSI-satisfaction and dissatisfaction subscales to the significant individual change identified by the ISS. As seen in Table 4, use of the MDC_{95} coefficients of the QSI revealed 7 clear change categories in which individuals changed in the same direction on one or both of the subscales. A sizable number of the respondents significantly changed on just one of the dimensions (satisfaction or dissatisfaction but not both), suggesting that sexual satisfaction and sexual dissatisfaction can change independently over time, potentially yielding distinct treatment effects that would have been obscured if these subscales had been collapsed into a single dimension.

When the ISS individual change categories were compared to these seven QSI change categories, the ISS showed excellent

agreement with the joint QSI categories on the diagonal of the table, identifying 100 % of the individuals showing deterioration on both QSI subscales as “significantly worse,” 92 % of the individuals showing no change on both QSI subscales as “no change,” and 83 % of the individuals showing improvement on both QSI subscales as “significantly better.” However, the ISS was only able to identify 35–67 % of the cases that had been found to be significantly better or worse on just one of the QSI subscales. This further helped to underscore how the use of unidimensional scales could potentially obscure interesting results. As a majority of the follow-up respondents demonstrated no significant change on any of the scales (given the relatively short follow-up interval of 2 months), these results might have represented an underestimate of the diversity of information that could be obtained using the QSI over a longer follow-up interval.

Discussion

The present study used IRT to evaluate the quality of the information provided by a set of well-validated measures of sexual quality. Consistent with Hypothesis 1, the results suggested that the existing scales (the ISS, the PSSI, and the YSSS) are not as informative or precise as they could be. Specifically, the IRT and EFA analyses suggested that each of those existing scales contained multiple items that contributed mostly error variance to the assessment of sexual quality, thereby lowering their precision through the inclusion of heterogeneous items. Thus, by using advanced statistical techniques in one of the first large-scale studies to directly compare the relative quality of these scales, the results suggested the need for an optimized measure of sexual quality. Consistent with

Hypothesis 2, the EFA results further suggested that sexual satisfaction and sexual dissatisfaction might represent distinct (yet related) components of sexual quality. These findings are consistent with recent work on other global evaluative dimensions like relationship quality (Fincham & Linfield, 1997; Fincham & Rogge, 2010; Mattson et al., 2012), highlighting the benefits of assessing positive and negative evaluations separately as well as the risks of obscuring meaningful results by collapsing global evaluations into a single dimension. Based on these findings, we conducted separate IRT analyses on a large and diverse pool of satisfaction and dissatisfaction items, selecting items offering the greatest precision in assessing these two dimensions of sexual quality to create the Quality of Sex Inventory (QSI). Consistent with Hypothesis 3, these scales offered markedly increased precision and power in assessing sexual satisfaction over the existing measures while retaining strong convergent and construct validity with existing measures. Consistent with Hypothesis 4, the QSI may offer researchers scales with optimized psychometric properties across a diverse range of subpopulations. Finally, consistent with Hypothesis 5, our longitudinal analyses suggested that despite being notably shorter in length, the QSI scales continued to demonstrate high levels of responsiveness to change in sexual quality over time when compared to the existing measures. The longitudinal analyses also helped to highlight the added information that could be obtained on change over time by keeping sexual satisfaction and dissatisfaction as distinct outcomes.

Implications

Limitations of Excessively Heterogeneous Measures

Most of the existing scales were originally conceptualized and designed as broad-band measures of sexual quality. As a result, the existing scales have notably heterogeneous item content. In fact, the EFA results suggested that the existing scales contained items from at least four distinct constructs: sexual satisfaction, sexual dissatisfaction, sexual desire, and relationship satisfaction. Given the more specific conceptual definition that guided the current study, only the sexual satisfaction and dissatisfaction items were included in the QSI. Thus, the factors for relationship satisfaction and sexual desire served as conceptual boundaries to remove heterogeneous items from the current measure. Heterogeneous item content is problematic for researchers interested in treating sexual satisfaction as a distinct construct from those other constructs in their models, as the assessment of sexual quality (with the existing scales) is confounded with items tapping these other constructs, muddling the conceptual clarity of what is being measured. Unexpectedly, the IRT-based DIF results shed additional light on this issue as it was largely the items loading on conceptually distinct factors like sexual desires that showed the greatest amounts of DIF. Thus, the inclusion of those

heterogeneous items might not only serve to obscure or blur the construct being assessed but may also make scores less comparable across gender and relationship stage subgroups.

Benefits of Increased Precision

The cross-sectional analyses suggested that the IRT-optimized QSI-satisfaction subscales offered greater precision in measurement and correspondingly higher levels of power for detecting subtle group differences than the existing measures. In fact, the results suggested that three widely used measures of sexual quality—the ISS, the YSSS, and the PSSI—contained surprisingly low amounts of information and relatively high levels of measurement error or noise, particularly for their respective lengths. Having said that, there is little doubt (and a large body of literature indicating) that these scales do in fact assess sexual quality, and do so well enough to reveal meaningful associations with other variables. The current results qualify that validity by revealing that the existing measures assessed sexual quality in a somewhat imprecise manner, reducing their power for detecting differences and even potentially reducing their power for detecting change in sexual quality over time. In contrast, the increased precision of the QSI scales might offer researchers a method of drastically reducing that measurement error and for increasing power while actually decreasing the total number of items necessary.

These results were similar to those obtained with IRT-optimized measures of relationship satisfaction (Funk & Rogge, 2007). Thus, by starting with a large and diverse item pool and by augmenting traditional measurement statistics like EFA with more sophisticated techniques like IRT, the results suggested that we were effective at creating a psychometrically optimized measure of sexual quality. These results more specifically suggested that the QSI scales would likely offer greater power for detecting a cross-sectional difference between groups (e.g., Fig. 1c), meaningful individual change over time (i.e., the MDC95 results), and differences between clinically distinct groups (i.e., the MCID results). Recent work has further suggested that higher cross-sectional and longitudinal precision of this form can translate into offering larger treatment effects in clinical trials (e.g., Rogge et al., 2015). As experimental and intervention studies (particularly those in clinical populations) tend to be labor-intensive and costly to conduct, they typically make use of smaller sample sizes. As a result, the greater precision and power offered by the QSI scales could offer researchers a critical edge in such studies, enabling them to detect significant results in smaller samples.

Distinction Between Satisfaction and Dissatisfaction

The EFA and longitudinal results suggested the potential utility of conceptually distinguishing between sexual satisfaction and sexual dissatisfaction. Specifically, these constructs emerged as meaningfully distinct at a correlational level in the initial data

and demonstrated that, although they were modestly correlated with one another, they were able to change independently of one another over time. In fact, these results began to suggest that collapsing positive and negative evaluations into a single dimension (as is current practice with the existing scales) might obscure meaningful longitudinal results. Consistent with this, a recent study examining relationship enhancing interventions in a sample of 174 couples followed over the first 3 years of marriage demonstrated significant outcome differences across intervention groups for positive but not negative relationship evaluations, and those differences were obscured when a unidimensional scale was used to model relationship quality (Rogge et al., 2013). The current results suggested that the use of unidimensional scales like the ISS, YSSS, or PSSI could potentially be obscuring results in a similar manner.

Detecting Improvement Versus Deterioration

One of the unique contributions of the current study was extending measurement work longitudinally to evaluate how the various sexual quality measures were responsive to change over time. Although these analyses continued to support the use of the QSI-satisfaction subscales, they also yielded an unexpected result. The MCID analyses suggested that all of the sexual satisfaction scales showed far more robust and consistent shifts in scores in response to *decreases* in sexual satisfaction (particularly for respondents starting out with high levels of sexual satisfaction) than in response to *increases* in sexual satisfaction. This is potentially problematic for treatment researchers interested in improving couples' sexual quality, as it suggests that detecting improvement is a bit more challenging across all of the self-report scales examined. One possible explanation for this result would be that couples might experience reductions in sexual quality in similar manners but might experience improvements in sexual quality in different (i.e., more heterogeneous) ways, making it harder to assess that type of change over time. Future studies should extend the longitudinal analysis of these scales to help maximize their responsiveness to change. For example, most globally evaluative scales use present tense when wording their items and typically do not offer respondents a specific timeframe. These practices, although widespread, might actually encourage respondents to report on more stable, trait-like subjective evaluations, rendering those reports less responsive to slight fluctuations over time. In contrast, measures of clinical symptoms like depression or anxiety (and items used in daily diary studies) routinely use past tense and give highly specific timeframes (e.g., "*in the last day*," "*in the last week*") to enhance the responsiveness of self-reports. Drawing upon that work, it might be possible to improve the responsiveness of all these sexual quality scales simply by putting their items into the past tense and giving respondents a concrete timeframe (e.g., "*In the last 2 weeks...*") to encourage respondents to consider them as state-like evaluations rather than trait-like evaluations.

Limitations

Despite the robust findings supporting the efficacy and validity of the QSI scales, the interpretation of these results was qualified by some limitations. To begin, the study was conducted entirely online. Although this provided a highly efficient and cost-effective method of amassing the large sample necessary for IRT analyses, participation required a computer and access to the internet, possibly filtering out respondents from the lowest socio-economic statuses. Future studies examining the QSI should strive to include a more diverse subject pool. A second limitation of the current study was that only one member of each dyad participated in the study, prohibiting analyses examining the level of agreement in partners' perceptions of sexual quality. Future studies will need to examine the QSI scales within dyads to fully examine the dependency of that data. A third limitation was that the current study was conducted in a community sample rather than a clinical population. IRT, when conducted in a sufficiently large and diverse sample, yields sample-independent results that allow researchers to anticipate how scales will operate in a variety of future samples. In fact, the IRT results presented suggested that the QSI scales should offer excellent levels of information and precision in clinical samples (e.g., in the range from 1 to 3 SDs below the mean). In addition, roughly 22 % of the sample reported significant relationship dissatisfaction on the CSI, thereby suggesting reasonable variance on relationship quality (and by extension sexual quality) in the sample. However, given the nature of the current sample, the results primarily offered insights into how the examined measures of sexual quality might operate in a non-clinical population, and future studies should evaluate the relative utilities of these scales in populations of couples undergoing treatment for sexual problems. Another limitation was that frequency of and desire for sexual activity and physical affection were self-reported for the 2-week period prior to when the participants completed the survey, which, for a subsample of women may have included the timeframe of their menstrual cycles. This may have influenced their or their partners' sexual desire or frequency, so future studies should consider accounting for menstrual cycle when assessing these constructs. Along these same lines, the current study did not measure amount of time spent with partner over the 2-week period assessed, which could influence frequency and quality of sexual activity and physical affection. Future studies should account for time spent with partner and related variables when evaluating sexual quality to help delineate the boundaries of those predictors. Finally, the study made use of a newly created 4-item scale assessing sex drive. These items were written to be conceptually focused (see Baumeister et al., 2001), contained highly similar content to more established measures of sex drive (e.g., the Sex Drive Questionnaire, Ostovich & Sabini, 2004; Sex Drive Scale, Lippa, 2006; and the Arizona Sexual Experiences Scale, McGahuey et al., 2011), were piloted in a sample of 385 online respondents, and demonstrated high internal consistency in the pilot sample and the current sample. However, future research could extend the current work by using

these more established scales to examine the links between sex drive and sexual quality.

Conclusions

Despite these limitations, the results presented here provided critical information on the shortcomings of the current measures of sexual quality as well as considerable support for the use of the QSI scales as optimized measures. Combining EFA, IRT, and responsiveness (i.e., MDC₉₅ & MCID) analyses in a large item pool given to a large online sample allowed us to develop the QSI scales—psychometrically optimized measures of sexual satisfaction and dissatisfaction. The results suggested that the QSI scales might offer researchers greater precision (lower noise) for detecting cross-sectional differences between groups and might offer comparable if not higher levels of responsiveness for detecting change over time despite being notably shorter than most of the existing measures examined.

Longer Scales Provide Greater Precision/Power

Consistent with previous work (e.g., Fraley et al., 2000; Funk & Rogge, 2007; Rogge et al., 2015), our results suggested that longer scales generally offered greater precision both cross-sectionally and longitudinally compared to shorter scales. Thus, whenever possible, we would recommend that researchers should use the longest version of the QSI that their studies can accommodate. However, recognizing that researchers might have a wide range of needs across various studies, we developed and validated two different versions of the QSI. The longer 24-item scale (made up of two 12-item subscales assessing sexual satisfaction and dissatisfaction separately) would offer researchers the highest precision and power for detecting meaningful effects. Thus, it would be of greatest use in studies where sexual quality was a primary outcome—particularly if those studies were being conducted in smaller samples such as one is likely to find when piloting new treatments. The shorter 12-item scale (made up of two 6-item subscales assessing sexual satisfaction and dissatisfaction separately) would continue to offer researchers reasonably high levels of precision with far fewer items. Thus, these shorter scales may be more appropriate in large samples (where power is not as serious an issue), in studies where sexual quality is not as central to the hypotheses being examined, or in studies where the length of the survey is a critical issue (e.g., daily diary studies, telephone interviews). If survey length truly only allowed for a handful of items (e.g., national surveys), researchers could even choose to include just one of the 6-item subscales of the 12-item QSI.

Distinguishing Dissatisfaction from Satisfaction

The results presented helped to validate sexual dissatisfaction and satisfaction as distinct yet related constructs. More specifically,

the analyses presented in Table 4 were consistent with recent treatment findings (e.g., Rogge et al., 2013) suggesting that unidimensional measures of global sentiment can potentially obscure meaningful treatment results revealed by bi-dimensional measures like the QSI. Thus, although researchers could choose to include a single subscale of the QSI in their studies, we would recommend that they include both sexual satisfaction and dissatisfaction as distinct outcomes in their studies, modeling them separately so as to gain a more comprehensive understanding of couples' sexual quality.

In summary, the current results suggested that the QSI scales offer researchers a measure of sexual quality (1) with greater conceptual clarity, (2) shifting the theoretical definition of this construct to a bi-dimensional model, (3) offering researchers greater precision and power, and (4) offering a flexible array of lengths.

References

- Abd-El-Fattah, S. M., & Hassan, H. K. (2011). Dependent-alpha calculator: Testing the differences between dependent coefficients alpha. *Journal of Applied Quantitative Methods*, 6, 59–61.
- Barry, R. A., Bunde, M., Lawrence, E., & Brock, R. (2009). Validity and utility of a multidimensional model of received support in intimate relationships. *Journal of Family Psychology*, 23, 48–57.
- Baumeister, R. F., Catanese, K. R., & Vohs, K. D. (2001). Is there a gender difference in strength of sex drive? Theoretical views, conceptual distinctions, and a review of relevant evidence. *Personality and Social Psychology Review*, 5, 242–273.
- Butzer, B., & Campbell, L. (2008). Adult attachment, sexual satisfaction, and relationship satisfaction: A study of married couples. *Personal Relationships*, 15, 141–154.
- Byers, S. E. (2005). Relationship satisfaction and sexual satisfaction: A longitudinal study of individuals in long-term relationships. *Journal of Sex Research*, 42, 113–118.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309–319.
- Connell, K. M., Coates, R., Doherty-Poirier, M., & Wood, F. M. (2013). A literature review to determine the impact of sexuality and body image changes following burn injuries. *Sexuality and Disability*, 31, 403–412.
- Daniel, S., & Bridges, S. K. (2013). The relationships among body image, masculinity, and sexual satisfaction in men. *Psychology of Men & Masculinity*, 14, 345–351.
- Davison, S. L., Bell, R. J., La China, M., Holden, S. L., & Davis, S. R. (2008). Assessing sexual function in well women: Validity and reliability of the Monash Women's Health Program Female Sexual Satisfaction Questionnaire. *Journal of Sexual Medicine*, 5, 2575–2586.
- Dehle, C., Larsen, D., & Landers, J. E. (2001). Social support in marriage. *American Journal of Family Therapy*, 29, 307–324.
- Derogatis, L. R., & Melisaratos, N. (1979). The DSFI: A multidimensional measure of sexual functioning. *Journal of Sex and Marital Therapy*, 5, 244–281.
- Easton, S. D., Coohy, C., O'Leary, P., Zhang, Y., & Hua, L. (2011). The effect of childhood sexual abuse on psychosexual functioning during adulthood. *Journal of Family Violence*, 26, 41–50.
- Eysenck, H. J. (1970). Personality and attitudes to sex: A factorial study. *Personality*, 1, 355–376.
- Fincham, F. D., & Bradbury, T. N. (1987). The assessment of marital quality: A reevaluation. *Journal of Marriage and the Family*, 49, 797–809.

- Fincham, F. D., & Linfield, K. J. (1997). A new look at marital quality: Can spouses feel positive and negative about their marriage? *Journal of Family Psychology, 11*, 489–502.
- Fincham, F. D., & Rogge, R. D. (2010). Understanding relationship satisfaction: Theoretical challenges and new tools for assessment. *Journal of Family Theory and Review, 2*, 227–242.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286–299.
- Foster, L. R., & Byers, E. S. (2013). Comparison of the sexual well-being of individuals with and without a herpes and/or human papillomavirus infection diagnosis. *International Journal of Sexual Health, 25*, 148–162.
- Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology, 78*, 350–365.
- 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.
- Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. *Journal of Chronic Disease, 40*, 171–178.
- Haavio-Mannila, E., & Kontula, O. (1997). Correlates of increased sexual satisfaction. *Archives of Sexual Behavior, 26*, 399–419.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Heinemann, L. A. J., Potthoff, P., Heinemann, K., Pauls, A., Ahlers, C. J., & Saad, F. (2005). Scale for Quality of Sexual Function (QSF) as an outcome measure for both genders. *Journal of Sexual Medicine, 2*, 82–95.
- Holmberg, D., Blair, K. L., & Phillips, M. (2010). Women's sexual satisfaction as a predictor of well-being in same-sex versus mixed sex relationships. *Journal of Sex Research, 47*, 1–11.
- Hudson, W. W., Harrison, D. F., & Crosscup, P. C. (1981). A short-form scale to measure sexual discord in dyadic relationships. *Journal of Sex Research, 17*, 157–174.
- Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: A critical review and recommendations. *Journal of Clinical Epidemiology, 53*, 459–469.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19.
- Lawrance, K., & Byers, E. S. (1995). Sexual satisfaction in long-term heterosexual relationships: The interpersonal exchange model of sexual satisfaction. *Personal Relationships, 2*, 267–285.
- Lawrance, K., Byers, E. S., & Cohen, J. N. (2011). Interpersonal exchange model of sexual satisfaction questionnaire. In T. D. Fisher, C. M. Davis, W. L. Yarber, & S. L. Davis (Eds.), *Handbook of sexuality-related measures* (pp. 525–530). New York: Routledge.
- Lippa, R. A. (2006). Is high sex drive associated with increased sexual attraction to both sexes? It depends on whether you are male or female. *Psychological Science, 17*, 46–52.
- LoPiccolo, J., & Steger, J. C. (1974). The Sexual Interaction Inventory: A new instrument for assessment of sexual dysfunction. *Archives of Sexual Behavior, 3*, 585–595.
- Mark, K. P., Herbenick, D., Fortenberry, J. D., Sanders, S., & Reece, M. (2014). A psychometric comparison of three scales and a single-item measure to assess sexual satisfaction. *Journal of Sex Research, 51*, 159–169.
- Mark, K. P., Janssen, E., & Milhausen, R. R. (2011). Infidelity in heterosexual couples: Demographic, interpersonal, and personality-related predictors of extradyadic sex. *Archives of Sexual Behavior, 40*, 971–982.
- Mattson, R. E., Rogge, R. D., Johnson, M. D., Davidson, E. K. B., & Fincham, F. D. (2012). The positive and negative semantic dimensions of relationship satisfaction. *Personal Relationships, 20*, 328–355.
- McGahuey, C. A., Gelenberg, A. J., Laukes, C. A., Moreno, F. A., Delgado, P. L., McKnight, K. M., & Manber, R. (2011). The Arizona Sexual Experience Scale (ASEX): Reliability and validity. *Journal of Sex and Marital Therapy, 26*, 25–40.
- Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111*, 172–175.
- Meston, C. M., Lorenz, T. A., & Stephenson, K. R. (2013). Effects of expressive writing on sexual dysfunction, depression, and PTSD in women with a history of childhood sexual abuse: Results from a randomized clinical trial. *Journal of Sexual Medicine, 10*, 2177–2189.
- Meston, C., & Trapnell, P. (2005). Development and validation of a five-factor sexual satisfaction distress scale for women: The Sexual Satisfaction Scale for Women (SSS-W). *Journal of Sexual Medicine, 2*, 66–81.
- Norton, R. (1983). Measuring marital quality: A critical look at the dependent variable. *Journal of Marriage and the Family, 45*, 141–151.
- Oggins, J., Leber, D., & Veroff, J. (1993). Race and gender differences in black and white newlyweds' perceptions of sexual and marital relations. *Journal of Sex Research, 30*, 152–160.
- Ostovich, J. M., & Sabini, J. (2004). How are sociosexuality, sex drive, and lifetime number of sexual partners related? *Personality and Social Psychology Bulletin, 30*, 1255–1266.
- Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: A more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology, 95*, 1113–1135.
- Pinney, E. M., Gerrard, M., & Denny, N. W. (1987). The Pinney Sexual Satisfaction Inventory. *Journal of Sex Research, 23*, 233–251.
- Plack, K., Kroger, C., Allen, E. S., Baucom, D. H., & Hahlweg, K. (2010). Risk factors for infidelity: Why do we have affairs? *Zeitschrift für Klinische Psychologie und Psychotherapie: Forschung und Praxis, 39*, 189–199.
- Pujols, Y., & Meston, C. M. (2009). The association between sexual satisfaction and body image in women. *Journal of Sexual Medicine, 7*, 905–916.
- Rao, D., Zajacka, J., & Skubiak, T. (2005). The Modified Rush Sexual Inventory: Preliminary psychometric findings. *Psychiatry Research, 137*, 175–181.
- Rogge, R. D., Cobb, R. J., Lawrence, E., Johnson, M. D., & Bradbury, T. N. (2013). Is skills training necessary for the primary prevention of marital distress and dissolution: A 3-year experimental study of three interventions. *Journal of Consulting and Clinical Psychology, 81*, 949–961.
- Rogge, R. D., Crasta, D., Maniaci, M. R., Funk, J. L., & Lee, S. (2015). *How well can we detect shifts in relationship satisfaction over time? Evaluating responsiveness to change in relationship satisfaction scales*. Manuscript submitted for publication.
- Rust, J., & Golombok, S. (1986). The GRISS: A psychometric instrument for the assessment of sexual dysfunction. *Archives of Sexual Behavior, 15*, 157–165.
- Samejima, F. (1997). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85–100). New York: Springer.
- Schumm, W. A., Nichols, C. W., Schectman, K. L., & Grinsby, C. C. (1983). Characteristics of responses to the Kansas Marital Satisfaction Scale by a sample of 84 married mothers. *Psychological Reports, 53*, 567–572.
- Shaw, A. M., Rhoades, G. K., Allen, E. S., Stanley, S. M., & Markman, H. J. (2013). Predictors of extradyadic sexual involvement in unmarried opposite-sex relationships. *Journal of Sex Research, 50*, 598–610.
- Snell, W. E., Fisher, T. D., & Walters, A. S. (1993). The Multidimensional Sexuality Questionnaire: An objective self-report measure of psychological tendencies associated with human sexuality. *Annals of Sex Research, 6*, 27–55.
- Stratford, P. W., Finch, E., Solomon, P., Binkley, J., Gill, C., & Moreland, J. (1996). Using the Roland-Morris Questionnaire to make decisions about individual patients. *Physiotherapy Canada, 48*, 107–110.
- Stroebe, S. S., O'Keefe, S. L., Beard, K. W., Kuo, S., Swindell, S. V. S., & Kommor, M. J. (2012). Father-daughter incest: Data from an anonymous computerized survey. *Journal of Child Sexual Abuse, 21*, 176–199.

- Stulhofer, A., Busko, V., & Brouillard, P. (2010). Development and bicultural validation of the New Sexual Satisfaction Scale. *Journal of Sex Research, 47*, 257–268.
- Suvak, M. K., Brogan, L. A., & Shipherd, J. C. (2012). Predictors of sexual functioning in a sample of U.S. marines: An 11-year follow-up study. *International Journal of Sexual Health, 24*, 26–44.
- Syme, M. L., Klonoff, E. A., Macera, C. A., & Brodine, S. K. (2013). Predicting sexual decline and dissatisfaction among older adults: The role of partnered and individual physical and mental health factors. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 68B*, 323–332.
- Thissen, D., Chen, W. H., & Bock, D. (2002). *Multilog user's guide: Multiple, categorical item and test scoring using item response theory (Version 70)* [Computer software]. Lincolnwood, IL: Scientific Software International.
- Walsh, C., Rodrigues, A., & Rogge, R. D. (2015). *Exploring the underlying structure of aversive behavior in relationships*. Manuscript submitted for publication.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PAN-AS scales. *Journal of Personality and Social Psychology, 54*, 1063–1070.
- Whitley, M. P., & Poulsen, S. B. (1975). Assertiveness and sexual satisfaction in employed professional women. *Journal of Marriage and Family, 37*, 573–581.
- Woodruff, D. J., & Feldt, L. S. (1986). Tests for equality of several alpha coefficients when their sample estimates are dependent. *Psychometrika, 51*, 393–413.
- Young, M., Denny, G., Luquis, R., & Young, T. (1998). Correlates of sexual satisfaction in marriage. *Canadian Journal of Human Sexuality, 7*, 115–127.
- Yount, S. M. (2013). The impact of pelvic floor disorders and pelvic surgery on women's sexual satisfaction and function. *Journal of Midwifery & Women's Health, 58*, 538–545.