

Revised Scoring and Improved Reliability for the Communication Patterns Questionnaire

Alexander O. Crenshaw
University of Utah

Andrew Christensen
University of California, Los Angeles

Donald H. Baucom
University of North Carolina, Chapel Hill

Norman B. Epstein
University of Maryland, College Park

Brian R. W. Baucom
University of Utah

The Communication Patterns Questionnaire (CPQ; Christensen, 1987) is a widely used self-report measure of couple communication behavior and is well validated for assessing the demand/withdraw interaction pattern, which is a robust predictor of poor relationship and individual outcomes (Schrodt, Witt, & Shimkowski, 2014). However, no studies have examined the CPQ's factor structure using analytic techniques sufficient by modern standards, nor have any studies replicated the factor structure using additional samples. Further, the current scoring system uses fewer than half of the total items for its 4 subscales, despite the existence of unused items that have content conceptually consistent with those subscales. These characteristics of the CPQ have likely contributed to findings that subscale scores are often troubled by suboptimal psychometric properties such as low internal reliability (e.g., Christensen, Eldridge, Catta-Preta, Lim, & Santagata, 2006). The present study uses exploratory and confirmatory factor analyses on 4 samples to reexamine the factor structure of the CPQ to improve scale score reliability and to determine if including more items in the subscales is warranted. Results indicate that a 3-factor solution (constructive communication and 2 demand/withdraw scales) provides the best fit for the data. That factor structure was confirmed in the replication samples. Compared with the original scales, the revised scales include additional items that expand the conceptual range of the constructs, substantially improve reliability of scale scores, and demonstrate stronger associations with relationship satisfaction and sensitivity to change in therapy. Implications for research and treatment are discussed.

Keywords: assessment, marriage, couples, communication, Communication Patterns Questionnaire (CPQ)

Supplemental materials: <http://dx.doi.org/10.1037/pas0000385.supp>

Effective communication between partners is widely considered to be an essential part of successful romantic relationship functioning, and dissatisfaction with communication is the most com-

mon reason couples seek therapy (e.g., Doss, Simpson, & Christensen, 2004). Communication within romantic relationships encompasses a wide range of behaviors and behavioral patterns, and a large accumulation of evidence suggests that both negative and positive behaviors contribute to relationship satisfaction and outcomes (e.g., Karney & Bradbury, 1995). Assessment of these behavioral patterns is an integral part of couple communication research as well as of the practice of couple therapy. The ability to measure communication patterns using a self-report measure is particularly important for couple therapists, given the time and resource requirements of other methods. Unfortunately, the most widely used self-report measure of communication, the Communication Patterns Questionnaire (CPQ; Christensen, 1987), has significant psychometric limitations (e.g., Christensen, Eldridge, Catta-Preta, Lim, & Santagata, 2006) that prevent researchers and therapists from optimally assessing and tracking communication patterns in couples. However, these are not limitations of the CPQ itself, but rather limitations of the current scoring used to compute its subscales. They can be addressed by reanalyzing the CPQ's factor structure and including additional, unused items in order

This article was published Online First September 12, 2016.

Alexander O. Crenshaw, Department of Psychology, University of Utah; Andrew Christensen, Department of Psychology, University of California, Los Angeles; Donald H. Baucom, Department of Psychology, University of North Carolina, Chapel Hill; Norman B. Epstein, Department of Family Science, University of Maryland, College Park; Brian R. W. Baucom, Department of Psychology, University of Utah.

This article was supported in part by start-up funding from the University of Utah awarded to Brian R. W. Baucom. The randomized clinical trial data on which this article is based was supported by grants from the National Institute of Mental Health awarded to Andrew Christensen at UCLA (MH56223) and Neil S. Jacobson at the University of Washington (MH56165). We thank Dr. Lisa Harris and Dr. James Shenk, former students of Dr. Christensen, for use of their data that comprised the divorcing sample.

Correspondence concerning this article should be addressed to Alexander O. Crenshaw, Department of Psychology, University of Utah, 380 South 1530 East, Room 502, Salt Lake City, UT 84112. E-mail: alexander.crenshaw@utah.edu

to improve psychometric properties and the utility of the scale in both research and clinical practice settings.

Couple Communication

Communication behavior in couples can be categorized into two main types: positive behaviors and negative behaviors (Woodin, 2011). Within these clusters, constructive communication (positive) and demand/withdraw behavior (negative) are particularly strongly associated with a wide range of relationship functioning variables (e.g., K. J. Baucom, Baucom, & Christensen, 2015; Schrodt, Witt, & Shimkowski, 2014). Constructive communication is an inclusive term for a host of positive behaviors that serve to promote a collaborative approach to problem solving and engender trust and understanding. Examples include making suggestions (in contrast to demands), compromising, perspective taking, and expressing feelings. Constructive communication is strongly and positively associated with marital satisfaction (Heavey, Larson, Zumtobel, & Christensen, 1996; Litzinger & Gordon, 2005), is associated with more forgiveness in heterosexual couples (Fincham & Beach, 2002), and is believed to buffer the detrimental effect of poor sexual satisfaction on overall marital satisfaction (Litzinger & Gordon, 2005).

In contrast to the relationship enhancing nature of constructive communication, demand/withdraw behavior and mutual avoidance are patterns of behavior that sustain and intensify conflict and are associated with negative affect during and following interaction between partners (e.g., McGinn, McFarland, & Christensen, 2009). Mutual avoidance describes a process in which both partners avoid the conflict altogether, for example, by becoming silent, changing the subject, or walking away from each other (Christensen & Shenk, 1991). In mutual avoidance, withdrawal by one partner is not contested by the other, as he or she is also seeking to withdraw. In contrast, demand/withdraw behavior is a dyadic pattern in which one partner nags, criticizes, complains, or otherwise attempts to initiate change, while the other partner avoids, terminates, or withdraws from the interaction (Christensen, 1987).

A large body of evidence links demand/withdraw behavior to numerous individual and relationship sequelae. Higher levels of demand/withdraw behavior are associated with greater relationship distress among both satisfied and unsatisfied couples (see Eldridge & Baucom, 2012), a finding that has been replicated in opposite-sex couples from numerous countries (e.g., United States, Taiwan, Brazil, Switzerland, and Pakistan; see B. R. Baucom, McFarland, & Christensen, 2010; Christensen et al., 2006), and in same-sex couples in the United States (e.g., Kurdek, 2004). Higher levels of demand/withdraw behavior are also associated with greater likelihood of divorce (Gottman & Levenson, 2000), infidelity (Balderrama-Durbin, Allen, & Rhoades, 2012), and intimate partner violence (Holtzworth-Munroe, Smutzler, & Stuart, 1998). Demand/withdraw is also associated with a host of negative individual outcomes, including depression (Rehman, Ginting, Karimiha, & Goodnight, 2010), alcoholism (Kelly, Halford, & Young, 2002), and decreased subjective well-being (Schrodt et al., 2014).

These behavioral patterns are most commonly assessed in two ways: observational coding and self-report. Observational coding involves having trained (e.g., Heavey, Gill, & Christensen, 1998) or untrained (e.g., K. J. Baucom, Baucom, & Christensen, 2012) raters view video recordings of couples engaging in a discussion,

and rate the strength and/or frequency of certain behaviors. Observational coding is commonly used in laboratory-based research because it is objective. However, because of its time- and resource-consuming nature, observational coding tends to be restricted to research contexts and with small to moderate sample sizes. Large-scale survey research, Internet-based research, and clinical settings are much more reliant on self-report measures to assess communication patterns.

Communication Patterns Questionnaire (CPQ)

The CPQ is freely available and one of the most commonly used self-report measures for assessing communication patterns in romantic couples. (Christensen, 1987; Schrodt et al., 2014). Based in part on an original measure developed by Sullaway and Christensen (1983), the CPQ consists of 35 Likert-scale items that assess dyadic patterns in ways that couples typically deal with relationship problems at three time periods: when a problem arises, during discussion of the problem, and after the discussion of the problem. The items of the CPQ are most commonly used to generate four subscales: Constructive Communication (seven items), Mutual Avoidance (three items), and two demand/withdraw scales (Self-Demand/Partner Withdraw and Partner-Demand/Self-Withdraw; three items each).

The CPQ scoring has undergone several revisions since its creation. It was originally conceptualized as having three scales—Mutual Constructive Communication, Demand/Withdraw Behavior, and Demand/Withdraw Roles (Christensen, 1987). Using this scoring, stronger demand/withdraw behavior has been linked with lower relationship satisfaction and greater asymmetry in level of intimacy and independence desired by partners (Christensen, 1987). The same study also found the Constructive Communication scale to be inversely related to Demand/Withdraw Behavior. However, this early scoring system grouped items into scales on conceptual grounds and examined psychometric properties using only a within-couple intraclass correlation (ICC), finding moderate agreement between males and females.

Christensen and Shenk (1991) revised the CPQ on conceptual grounds to include a fourth scale: Mutual Avoidance. Accounting for the fact that an individual can occupy both demanding and withdrawing roles in a relationship, even if those two behaviors are mutually exclusive at any given time point, Christensen and Shenk modified the demand/withdraw scales by removing the Demand/Withdraw Roles scale and separating demand/withdraw behavior into Male-Demand/Female-Withdraw and Female-Demand/Male-Withdraw subscales. In a sample of 62 couples, they found that all four scales of their revised CPQ distinguished distressed from nondistressed couples. Another study examining psychometric properties of the CPQ in a sample of 96 married community couples found that 29 of its 35 items were individually able to distinguish couples with respect to marital adjustment (Noller & White, 1990). In addition, Noller and White (1990) used exploratory factor analysis to examine the CPQ's factor structure, finding four factors somewhat different from previous scoring systems: Coercion, Mutuality, Postconflict Distress, and Destructive Processes. Using this scoring system, they found well-adjusted couples reported higher levels of mutuality, and poorly adjusted couples reported higher levels of destructive process, coercion, and postconflict distress.

Taken together, there is strong evidence for the utility of the CPQ in assessing couple communication behavior, despite the fact that a number of different scoring systems have been used. Currently, the Christensen and Shenk (1991) scoring system for the CPQ is the most commonly used, although the Constructive Communication scale has since been revised to include seven items, and includes items assessing both positive communication and negative communication (Heavey et al., 1996). However, this scoring was constructed on theoretical grounds without the use of factor analytic techniques, which raises concerns about psychometric properties of the subscales. Indeed, psychometric properties of the CPQ are highly inconsistent across studies. For example, Christensen et al. (2006) reported interitem ICCs (Cronbach's alpha) between .73 and .78 for constructive communication and female-demand/male-withdraw, but also reported an ICC of .58 for male-demand/female-withdraw among Americans, and ICCs ranging from .21 to .81 in samples from Taiwan, Brazil, and Italy. Another cross-cultural study found ICCs ranging from .44 to .80 (Bodenmann, Kaiser, Hahlweg, & Fehm-Wolfsdorf, 1998).

The lack of consistency in internal reliabilities of subscale scores for the CPQ across various samples raises concerns about the extent of its ability to validly describe communication across a range of populations. With such inconsistent reliabilities reported, one may question whether the CPQ is measuring the same constructs in different populations and among couples at different levels of functioning. In addition, although items measuring the same construct tend to produce a strong ICC, a strong ICC by itself is not sufficient for determining whether items measure the same construct. Further, the one study that utilized factor analytic methods for determining the factor structure of the CPQ (Noller & White, 1990) examined only 96 married couples, did not sample across a range of couple functioning, and did not replicate their exploratory results with a priori confirmatory techniques. In addition, they only utilized the Kaiser-Guttman "Eigenvalue > 1" criterion for deciding the number of factors, a technique that is inadequate by modern standards and that tends to overestimate the number of factors (Tabachnick & Fidell, 2013).

Another problem with the current scoring of the CPQ is that it makes use of only 16 of 35 total items, even though several unused items are conceptually consistent with some of the subscales. This fact is especially problematic for the demand/withdraw scales, which contain only three items each, despite the fact that the CPQ includes several additional items that assess conceptually similar behavior. Put in broader terms, demand/withdraw could be described as a behavior pattern in which one person actively approaches a problem while the other actively avoids the problem, discounts it as a problem, or responds with passivity. Thus, any item that describes an asymmetrical behavior pattern in which one partner has a negatively valenced approach orientation to the partner or problem while the other partner has an avoidant orientation toward the partner or problem may capture demand/withdraw behavior and is likely to be a good candidate for the Demand/Withdraw scale. For example, Item 17 ("I threaten negative consequences and my partner gives in or backs down") appears to be an especially destructive type of demand/withdraw behavior, but it is currently unused in any scale.

Thus, there is strong reason to believe that the CPQ could be improved considerably through using items that are conceptually consistent with its subscales but not currently included in the

scoring system. However, no study to date has examined the factor structure of the CPQ using methods that meet modern analytic standards, nor has any study confirmed the hypothesized scales on a replication sample using an a priori approach such as confirmatory factor analysis (CFA). The present study uses modern factor analytic techniques to reexamine the factor structure of the CPQ, determine whether additional items should be included in its subscales, and examine replicability of the factor structure on three separate samples representing a wide range of couple functioning. Specifically, we hypothesized that exploratory factor analysis (EFA) conducted on a sample of treatment-seeking couples will replicate the four-factor solution used by Christensen and Shenk (1991) and modified by Heavey et al. (1996). Second, we hypothesized that the EFA would result in several currently unused but conceptually consistent items loading strongly onto the subscales. Third, we hypothesized that using CFA, the factor structure would replicate across three additional samples representing a wide range of relationship functioning, and factor loadings for revised subscales would not significantly differ across men and women. Fourth, inclusion of additional items was hypothesized to result in improved internal reliability of subscale scores, improving power to detect associations with other important variables. Finally, we hypothesized that revised CPQ subscales would show improved construct validity by having significantly stronger associations with relationship satisfaction and by demonstrating greater sensitivity to change produced by couple therapy.

Method

Participants

The current investigation utilized four separate samples of heterosexual married couples. The first sample (clinical trial) consists of 134 couples that took part in a multisite, randomized clinical trial of two behaviorally based couple therapies (Christensen et al., 2004). All couples had to be legally married, living together, and meet criteria for serious and stable marital distress prior to treatment. Both partners had to be between the ages of 18 and 65 years, fluent in English, and have at least a high school or equivalent education (see Christensen et al., 2004, for a complete description of sample characteristics). Mean marital satisfaction in this sample, as measured by the Dyadic Adjustment Scale (DAS; Spanier, 1976), was 84.5 ($SD = 15.0$) for men and 84.7 ($SD = 14.0$) for women. DAS scores in this sample fell below the well-accepted and widely used cutoff of 97.5 for clinically significant distress, which is one standard deviation below the population mean (e.g., Christensen et al., 2004).

Sample 2, the community sample, was a subset ($n = 359$) of couples with complete CPQ data from a sample of 386 married couples from communities in North Carolina and the Maryland/Washington, DC area as part of a larger study. Couples were recruited to match the U.S. population on key demographic variables, including age, income, and ethnic status (see D. H. Baucom, Epstein, Rankin, & Burnett, 1996, for a complete description of the community sample). Mean DAS scores in this sample were 111 ($SD = 15.4$) for men and 112 ($SD = 14.9$) for women, well above the distress cutoff of 97.5.

Sample 3, the clinic sample, was a subset ($n = 60$) of couples with complete CPQ data from a sample of 85 couples presenting

for marital therapy to either an independent practice or university psychology clinic in southern California. Couples completed a series of questionnaires, including the CPQ and DAS, during a pretreatment evaluation. Average DAS scores were 96.10 ($SD = 12.85$) for men and 90.39 ($SD = 18.17$) for women, slightly below the distress cutoff.

Sample 4, the divorcing sample, was a subset ($n = 52$) of couples with complete CPQ data from a sample of 60 couples recruited from a conciliation court (for couples unable to reach a custody agreement) in southern California as part of a larger study (Harris, 1992). Table 1 presents descriptive statistics for demographic characteristics for all four samples.

Measures

CPQ. The CPQ (Christensen, 1987) is a self-report measure of communication behavior in romantic couples. It contains 35 Likert-scale items assessing how couples typically deal with problems in their relationship: four items assessing behavior when a problem arises, 18 items assessing behavior during a discussion of a problem, and 13 items assessing behavior that occurs after discussion of a problem. Each item assesses partners' perception of how likely a certain type of behavior (e.g., both members avoid discussing the problem) occurs when faced with a relationship problem, from 1 (*very unlikely*) to 9 (*very likely*). Of the 35 items on the CPQ, 16 are currently used to form four subscales: Constructive Communication (seven items), Self-Demand/Partner-Withdraw (three items), Partner-Demand/Self-Withdraw (three items), and Mutual Avoidance (three items).

Dyadic Adjustment Scale (DAS). The DAS (Spanier, 1976) is a 32-item measure of relationship satisfaction, in which higher scores indicate higher satisfaction. Scores below 97.5 indicate clinically significant relationship distress (e.g., Christensen et al., 2004).

Analyses

EFA were conducted in SPSS 21 on the clinical trial sample using the common factor model with maximum likelihood estimation (Schmitt, 2011; Tabachnick & Fidell, 2013). An oblique (Promax) rotation was used to allow for correlation between factors. We determined the clinical trial sample to be the most appropriate for the EFA because the CPQ is most commonly used to measure communication in distressed (vs. nondistressed) couples and in treatment-seeking couples. We wanted to ensure that the revised scales were most appropriate for the population for which it is used, and the clinical trial sample was the only sample in which all couples were clinically distressed, treatment seeking couples. There were two central aims of the EFA. The first was to determine whether an analysis of all 35 items would produce four factors consistent with the current scoring of the CPQ, or whether a different solution was more appropriate. The second aim was to maximize conceptual clarity, interpretability, and theoretical meaningfulness of the subscales by examining whether inclusion of other theoretically similar but previously unused items broadened the content domain of each scale.

To accomplish both aims, EFAs were first conducted using a data-driven, empirical approach in order to narrow the field of possible factor solutions. This approach began with examination of scree plots, separately for men and women, based on eigenvalues from an initial, unrestricted (i.e., number of factors extracted was set equal to number of items) extraction. The scree test identifies the optimal number of factors in EFA as being the number of eigenvalues above the "elbow" in the plot, which is the point at which the slope of the line decreases most sharply (Tabachnick & Fidell, 2013). Rather than rely solely on the scree test, it was used to form an initial hypothesis about the number of factors present and to determine a range of other plausible factor solutions. All plausible factor solutions were then examined, and results were compared in terms of variance explained, conceptual interpretability,

Table 1
Sample Characteristics for Each of the Four Samples

Variable	Clinical trial ($n = 134$)		Community ($n = 359$)		Clinic ($n = 60$)		Divorcing ($n = 52$)	
	Male	Female	Male	Female	Male	Female	Male	Female
Race/ethnicity (%)								
Caucasian	79.1	76.1	89	89	98.8	95.3	37.3	40.0
African American	6.7	8.2	11	11	1.2	1.2	30.5	30.0
Latino/Latina	5.2	5.2	—	—	—	1.2	30.5	28.3
Asian/Pacific Islander	6.0	4.5	—	—	—	2.4	1.7	1.7
Native American/Alaskan	.7	—	—	—	—	—	—	—
	<i>M (SD)</i>		<i>M (SD)</i>		<i>M (SD)</i>		<i>M (SD)</i>	
	Male	Female	Male	Female	Male	Female	Male	Female
DAS	84.5 (15.0)	84.7 (14.0)	111 (15.4)	112 (15.0)	96.1 (12.9)	90.4 (18.2)	n/a	n/a
Age	43.5 (8.7)	41.6 (8.6)	44.2 (13.1)	42.2 (12.6)	38.7 (8.8)	35.3 (7.4)	37.6 (7.7)	34.5 (6.5)
Years education	17.0 (3.2)	17.0 (3.2)	15.7 (3.3)	15.1 (2.7)	n/a	n/a	14.2 (2.4)	14.9 (2.6)
Annual income (median)	\$48,000	\$36,000	\$50,000–\$70,000		n/a		\$10,000–\$50,000	
Marriage length (years)	10.0 (7.6)		17.5 (13.2)		7.69 (7.3)		n/a	

Note. Sample sizes are numbers of couples. Median annual income was reported at the individual level in the clinical trial sample and at the couple level in the community and divorcing sample. Education and income were not available in the clinic sample, and DAS and marriage length were not available in the divorcing sample. DAS = Dyadic Adjustment Scale (Spanier, 1976).

ity of factors (i.e., did the items within each factor appear to measure a single identifiable construct), and consistency of item loadings across both men and women. All items with standardized loadings above .3 on a given subscale were considered possible candidates for inclusion in that subscale.

Once a factor solution had been determined, CFAs were conducted separately for men and women on each of the three replication samples (community, clinic, and divorcing). CFAs were also conducted separately for each subscale in order to examine fit for each scale, sex, and sample combination individually. Items within subscales were not expected to correlate after accounting for shared factor variance, so residual correlations were fixed to zero.

Although the sample sizes of the clinic and divorcing samples were smaller than is typically recommended for standard CFA (i.e., maximum likelihood estimation; Kline, 2015; B. Muthén & Asparouhov, 2012), we decided to perform CFAs separately on each sample and each subscale, rather than combining them, for several reasons. First, an important question from both a theoretical and measurement perspective is whether the CPQ can validly capture communication behavior across a wide range of couple functioning. A related but separate question is whether communication behavior assessed via the CPQ can be described by the same set of dimensions (i.e., factor structure) across levels of relationship functioning. Both of these questions should be answered in order to determine whether the CPQ can be used validly across the spectrum of relationship quality. We also chose to conduct CFAs separately for each subscale in order to be able to identify specific sources of misfit in the model if poor fit were to arise.

Although standard structural equation modeling (SEM) typically calls for large sample sizes of approximately 200 or higher, Bayesian SEM (B. Muthén, 2010; B. Muthén & Asparouhov, 2012) can be used with sample sizes as small as 2 or 3 times the number of unknown parameters, especially when good priors are provided (Lee & Song, 2004). In the smallest sample used in this study (divorcing), the ratio of sample size to number of unknown parameters was 2.9 (52 individuals divided by 18 parameters—nine loadings and nine error variances)¹ for the Constructive Communication scale and 3.7 (52 divided by 14) for the demand/withdraw scales. As a result, Bayesian SEM was appropriate for estimating a separate model for each of the three replication samples, despite the small clinic and divorcing samples.

Analyses were performed using the Bayes estimator in Mplus 7.31 (L. K. Muthén & Muthén, 1998–2012), with the number of iterations set at 30,000. Rather than viewing parameters as constants, Bayesian analysis makes use of predetermined values, or priors, to estimate the parameter distribution (B. Muthén & Asparouhov, 2012). Priors can be diffuse (noninformative) or based on previous theory or empirical results (informative). As we used the CFA to validate the measurement model specified in the EFA, unstandardized factor loadings from the EFA were used as informative priors in the CFA model. Bayesian SEM also requires a value be set for the variance of each prior, and B. Muthén (2010) recommends testing several prior variance values and selecting the value with the lowest deviance information criterion (DIC). A prior variance of .1 resulted in the lowest DIC for all three samples, so all prior variances were set at .1. All models were then rerun using noninformative priors to examine the model's sensitivity to priors.

Evaluation of model fit was done via the posterior predictive p value, which is the standard fit index used for Bayesian SEM (B. Muthén & Asparouhov, 2012). The posterior predictive p is similar to a chi-square test of model fit in that a “nonsignificant” p value indicates good model fit, but it does not behave in the same way as a chi-square test, in that the expected Type I error is not .05 for a fitting model (see B. Muthén & Asparouhov, 2012). However, the posterior predictive p does appear sensitive to sample size, though the extent of its sensitivity to sample size does not yet appear fully resolved (see B. Muthén & Asparouhov, 2012). Consistent with B. Muthén and Asparouhov (2012), we selected a posterior predictive p cutoff of .05 for the current study. As the purpose of the current study was to revise and improve an existing measure rather than test a new measurement model, subscales were not rejected based only on a posterior predictive p below the cutoff. In addition to posterior predictive p , we considered the statistical significance ($p < .05$) of individual item loadings, consistency of factor loadings with EFA results and across samples, and sensitivity of results to choice of priors.

Once subscales were finalized, a test of “weak” factorial invariance (equivalency of item loadings; Kline, 2015) was conducted through multiple group analysis on the clinical trial sample to test whether item loadings could be treated as equivalent across men and women. These analyses were performed using the clinical trial sample because it was not part of the CFA and it also allowed examination of equivalency of item loadings on a sample for which the CPQ is most often used. Using the maximum likelihood (ML) estimator in Mplus in order to allow for statistical comparison of nested models, models for each subscale were run in which item loadings were constrained to be equal for men and women and again without such restriction. A chi-square difference test was used to determine whether the constrained and unconstrained models were significantly different.

We also examined changes in internal reliability for each subscale score in each sample when moving from the original to revised scoring. In addition to being necessary for ensuring that items on a scale are in fact measuring a single construct, having high internal reliability is also important for maximizing statistical power in empirical studies. Given a true correlation, ρ , between a CPQ subscale and another measure of interest, the observed correlation, r , will be reduced by the degree to which α of each measure is below 1 (Kline, 2015). Thus, by improving the internal reliability of a measure's scores, power in any analysis using that measure is necessarily improved. The R package cocron (Diedenhofen, 2016) was used to test significant differences in the internal reliability of subscales using the revised and original scoring.

Lastly, we examined convergent validity of the revised subscales and compared them with the original subscales. First, we examined correlations between relationship satisfaction and original and revised CPQ subscales in the clinical trial, community, and clinic samples. Satisfaction data were not available in the divorcing sample. The Fisher's r -to- z transformation in which two correlations share the same variable (i.e., DAS) was used to determine whether the differences in pairs of correlations with the DAS were significant (Lee & Preacher, 2013). Next, the ability of

¹ This calculation reflects the number of parameters for the final version of each scale, after poorly fitting items were dropped in the CFA step.

the CPQ subscales to detect change over a course of couple therapy was examined in the clinical trial sample in a series of multilevel models (MLMs). MLMs were estimated at sample sizes ranging from the full sample ($N = 134$) to the size of the average published outcome study of couple therapy ($n = 30$) using a bootstrap resampling procedure.

Results

Exploratory Factor Analysis

All item loadings reported are standardized unless stated otherwise. Scree plots (see Figure A1 in the online supplemental materials) for both men and women suggested three clear factors, as indicated by a clear “elbow” in both plots at the fourth eigenvalues (Tabachnick & Fidell, 2013). As a result, we identified a three-factor solution as the most likely solution, but chose to also examine four- and five-factor solutions in order to rule out these alternative possibilities. Consequently, subsequent extractions were performed using three-, four-, and five-factor solutions, followed by a comparison of the possible solutions in terms of variance explained, conceptual clarity, and interpretability.

Variance Explained

A three-factor solution resulted in rotated eigenvalues of 3.71, 3.60, and 3.21 for men; eigenvalues for women were 4.23, 3.20, and 3.48. The three-factor solution explained 29.35% of the variance in CPQ responses for both men and women. By comparison, the rotated eigenvalues for a four-factor solution were 3.70, 3.62, 3.25, and 1.69 (men), and 4.23, 3.20, 3.35, and 2.17 (women). The four-factor solution explained 34.04% of the variance in CPQ responses for men and 33.56% for women. Finally, a five-factor solution resulted in rotated eigenvalues of 3.54, 3.47, 3.17, 2.67, and 2.12 (men), and 4.26, 3.23, 3.31, 2.04, and 1.39 (women). The five-factor solution explained 37.82% (men) and 37.21% (women) of the variance in CPQ responses. Extractions with more factors will necessarily explain more variance than extractions with fewer factors, but the addition of a fourth and fifth factor in this sample did not explain substantially more variance compared with the three-factor solution.

Conceptual Interpretability and Consistency of Loadings Across Sex

Across sex, the three-factor solution had the same conceptual interpretation and yielded very similar solutions in terms of item loadings (see Table 2 for all factor loadings). We interpreted the three factors to be Constructive Communication, Self-Demand/Partner-Withdraw, and Partner-Demand/Self-Withdraw. These factors are conceptually the same as three of the previous CPQ factors, except that the three items from the original Mutual Avoidance scale (both avoid discussing problem, both withdraw after discussion, and neither gives in after discussion) loaded negatively on the Constructive Communication scale in these solutions.

The four-factor solution yielded the same three conceptual factors as the three-factor solution, with an additional factor, inconsistent across sex. For men, the set of constructive com-

munication items was split such that the fourth factor included three items previously on the constructive communication factor (both try to discuss the problem, both express feelings, both suggest solutions) in addition to two unrelated items (partner hits me, I'm nice after discussion while partner is distant). For women, the fourth factor consisted of the CPQ's two violence items (“I push, shove, slap, hit, or kick my partner”; and the partner version of the same item).

The five-factor solution yielded the same three conceptual factors as the three-factor solution, with the addition of a factor containing the two violence items mentioned previously that was consistent across sex and a fifth factor that was inconsistent across sex. For men, the fifth factor was uninterpretable, including the following items: both try to discuss the problem, both express feelings, both suggest solutions, I call partner names, and I'm nice after discussion while partner is distant. For women, the fifth factor contained only two postdiscussion items: I feel guilty while my partner feels hurt, and I try to be nice while my partner is distant.

Taken together, the EFA results suggest that a three-factor solution provides an optimal description of the dimensional structure of CPQ items. The three-factor solution is highly similar for men and women, it is conceptually clear and interpretable, and the presence of a Constructive Communication and two demand-withdraw subscales is consistent with how the CPQ has been used in previous research. A four-factor solution yielded a fourth factor that was inconsistent between men and women and included only three and two items, respectively. A five-factor solution yielded a fourth factor (Violence) that was consistent between men and women, but it contained only two items, which is below the required three items for retaining a factor (e.g., Kline, 2015). Furthermore, the fifth factor was uninterpretable for men and contained only two items for women. It is worth noting that none of the solutions that were explored yielded anything close to the original Mutual Avoidance subscale used in the previous scoring of the CPQ. Instead, those items loaded negatively on the Constructive Communication factor.

CFA

Tables A1 to A3 (in the online supplemental materials) present means, standard deviations, and correlations for all CPQ items in the CFA samples, and Figure A2 shows the CFA model specification. Initial analyses identified four items that loaded poorly across the replication samples and resulted in poor model fit. Therefore, Items 15 and 16 (I express feelings while my partner offers reasons and solutions, and the partner version of the same item) were removed from the Constructive Communication scale, and Items 21 and 22 (I push, shove, slap, hit, or kick partner, and the partner version of the same item) were removed from the Self-Demand/Partner-Withdraw scale and Partner-Demand/Self-Withdraw scale, respectively. The modified solution provided a substantially better fit for the data overall, and the removal of two items that describe the behavior of only one member of the couple from the demand/withdraw scales resulted in a conceptually clearer dyadic representation of demand/withdraw behavior.

Table 2
Standardized Item Loadings from Exploratory Factor Analysis for Three-Factor Solution
(Clinical Trial Sample)

Item	Constructive communication		Self-demand/ P.-withdraw		P.-demand/ Self-withdraw	
	M	F	M	F	M	F
1. Both avoid discussing ^a	-.389	-.318				
2. Both try to discuss ^a	.551	.667				
6. Both express feelings ^b	.359	.550				
8. Both suggest compromises	.597	.714				
23. Both feel understood ^c	.636	.667				
24. Both withdraw ^c	-.466	-.500				
25. Both feel resolved ^c	.675	.712				
26. Neither gives in ^c	-.625	-.419				
27. Both are especially nice ^c	.620	.633				
3. I start discussion/P avoids ^a	<i>-.373</i>		.510	.560		
9. I nag & demand/P withdraws ^b			.717	.782		
11. I criticize/P defends ^b			.499	.587		
13. I pressure to change/P resists ^b			.606	.609		
17. I threaten/P gives in ^b			.697	.353		
19. I call names, swear, etc. ^b		<i>-.305</i>	.613	.519		
32. I pressure to apologize/P resists ^c			.464	.568		
4. P starts discussion/I avoid ^a				<i>-.372</i>	.263	.352
10. P nags & demands/I withdraw ^b				<i>-.381</i>	.616	.570
12. P criticizes/I defend ^b					.507	.594
14. P pressures to change/I resist ^b					.517	.555
18. P threatens/I give in ^b		<i>.312</i>			.600	.646
20. P calls names, swears, etc. ^b					.689	.553
33. P pressures to apologize/I resist ^c					.595	.537
15. I express feelings/P offers solutions	<i>.303</i>	<i>.395</i>				
16. P expresses feelings/I offer solutions	<i>.166</i>	<i>.366</i>				
28. I feel guilty/partner feels hurt ^c	<i>.335</i>	<i>.331</i>	<i>.369</i>			
29. P feels guilty/I feel hurt ^c	<i>.358</i>					
5. Both blame, accuse, criticize ^b				<i>.357</i>	<i>.335</i>	<i>.337</i>
7. Both threaten each other ^b			<i>.400</i>		<i>.410</i>	
21. I push, shove, slap (etc) partner ^b			<i>.359</i>	<i>.033</i>	<i>.099</i>	<i>.340</i>
22. P pushes, shoves, slaps (etc) me ^b			<i>.203</i>	<i>-.059</i>	<i>.388</i>	<i>.440</i>
30. I'm nice/P distant ^c						
31. P nice/I'm distant ^c						
34. I seek support from others ^c						
35. P seeks support from others ^c						

Note. Item content shortened for readability. Loadings under .3 omitted for readability unless included for conceptual reasons. Bolded items were included in final scales; italicized items loaded over .3 but were ultimately excluded. Items 15, 16, 21, and 22 were removed after CFAs. M = male; F = female; P = partner. ^a“When some problem in the relationship arises”. ^b“During discussion of a relationship problem”. ^c“After a discussion of a relationship problem”.

Table 3 displays results from the CFAs, including standardized factor loadings and posterior predictive *p* values for each subscale/sample combination. Overall, factor loadings in all three replication samples for both males and females were significant, nearly all above or substantially above .3, and highly similar to those in the clinical trial sample. Of the 138 factor loadings reported, all but one loaded significantly on their respective scales (male Item 24 in divorcing sample was nonsignificant, *p* > .05). Posterior predictive *p* values suggest that the model in 12 of the 18 scale-sample-sex combinations has less than perfect model fit (*p* < .05).

CFAs were then repeated using noninformative priors in order to examine sensitivity of the measurement model to priors. For the community and clinic samples, the direction, magnitude, and significance of factor loadings were largely unchanged for all factor loadings across all scales. For the

divorcing sample, factor loadings were more sensitive to specification of priors. Specifically, using noninformative priors there were several instances on each scale for men, and on CC for women, in which the magnitudes of factor loadings were substantially smaller relative to when using informative priors (analyses available from the first author).

A test of “weak” factorial invariance (equivalence of unstandardized item loadings; Kline, 2015) was then conducted on the clinical trial sample to test equivalency of item loadings across men and women. As shown in Table A4, all factor loadings for the constrained model were significant at (*ps* < .01), and none of the chi-square difference tests comparing the constrained with the unconstrained model were significant (Constructive Communication, $\chi^2[9] = 6.01, p = .739$; Self-Demand/Partner-Withdraw, $\chi^2[7] = 5.01, p = .659$; Partner-Demand/Self-Withdraw, $\chi^2[7] = 10.47, p = .163$). Results indicate that

Table 3
Standardized Factor Loadings From Bayesian Confirmatory Factor Analysis Using Empirical Priors for Replication Samples (Community, Clinic, Divorcing)

Item	Community		Clinic		Divorcing	
	M	F	M	F	M	F
Constructive communication						
1. Both avoid discussing ^a	.419	.563	.379	.319	.298	.258
2. Both try to discuss ^a	.588	.602	.617	.588	.651	.516
6. Both express feelings ^b	.505	.631	.385	.438	.464	.535
8. Both suggest solutions & compromises ^b	.694	.643	.652	.652	.668	.755
23. Both feel understood ^c	.768	.796	.579	.657	.677	.673
24. Both withdraw ^c	.644	.663	.516	.728	.185 ^d	.364
25. Both feel resolved ^c	.735	.744	.574	.701	.527	.519
26. Neither gives in ^c	.588	.611	.694	.618	.314	.324
27. Both are especially nice ^c	.543	.504	.649	.392	.522	.543
Posterior predictive <i>p</i>	<.001	<.001	<.001	.002	.027	.113
Self-demand/partner-withdraw						
3. I start discussion/P avoids ^a	.402	.359	.543	.576	.381	.439
9. I nag & demand/P withdraws ^b	.631	.682	.700	.761	.627	.629
11. I criticize/P defends ^b	.778	.831	.571	.608	.654	.630
13. I pressure to change/P resists ^b	.692	.742	.619	.741	.500	.607
17. I threaten/P gives in ^b	.553	.602	.466	.479	.447	.482
19. I call names, swear, etc. ^b	.649	.648	.548	.337	.476	.561
32. I pressure to apologize/P resists ^c	.528	.524	.617	.643	.333	.496
Posterior predictive <i>p</i>	<.001	<.001	.067	<.001	.073	.665
Partner-demand/self-withdraw						
4. P starts discussion/I avoid ^a	.472	.559	.338	.531	.243	.564
10. P nags & demands/I withdraw ^b	.746	.677	.589	.588	.518	.731
12. P criticizes/I defend ^b	.680	.724	.669	.593	.565	.638
14. P pressures to change/I resist ^b	.671	.754	.708	.744	.350	.651
18. P threatens/I give in ^b	.589	.592	.700	.450	.679	.525
20. P calls names, swears, etc. ^b	.642	.582	.658	.464	.714	.494
33. P pressures to apologize/I resist ^c	.462	.571	.502	.554	.606	.491
Posterior predictive <i>p</i>	<.001	<.001	.285	.008	.209	.018

Note. Item content shortened to improve readability. M = male; F = female; P = partner.

^a "When some problem in the relationship arises . . .". ^b "During discussion of a relationship problem . . .". ^c "After a discussion of a relationship problem . . .". ^d Factor loading was not statistically significant ($p > .05$). All other loadings were significant at $p < .05$.

factor loadings for CPQ subscales are not significantly different for men and women.

Reliability and Power Improvement

Table 4 presents old and new reliabilities, reliability increase, proportional increase in reliability, and a chi-square test for difference in Cronbach's alpha separately for men and women in all samples. Chi-square tests of differences in Cronbach alphas (Feldt, Woodruff, & Salih, 1987) found that 18 of 24 reliabilities using the revised scoring were significantly larger than when using the original scoring. Further, six of 24 reliabilities using the original scoring were above .7, a generally recognized cutoff for good internal reliability, whereas 22 of 24 reliabilities were above .7 using the revised scoring. In most cases, proportion increase in reliability, which translates most closely to expected power improvement, was substantial. Using the original clinical trial sample as an example, the largest proportion increase was for men's Constructive Communication subscale, which increased from $\alpha = .566$ to $\alpha = .801$ (a 41.5% increase). The smallest proportion increase was for women's Self-Demand/Partner-Withdraw subscale, which increased from $\alpha = .645$ to $\alpha = .770$ (a 19.4% increase). Of the 24 reliabilities computed, one de-

creased slightly from the original to the revised scoring; the men's Self-Demand/Partner-Withdraw subscale in the divorcing sample changed from .634 (original) to .617 (revised), a decrease of 2.7%.

The observed reliability improvements translate into substantially improved power for detecting meaningful relationships with other variables, reducing the sample size needed to find statistical significance. To examine the extent to which the revised scoring system improves power in studies that use the CPQ, we used the original (clinical trial) sample to calculate the sample size needed to achieve .8 power in a two-tailed correlation analysis between the CPQ and a hypothetical other measure. Using G*Power 3.1, we examined a range of values for the reliability of the other measure scores and the true population correlation (ρ) between the CPQ scale and the other measure. We used two subscales from the clinical trial sample: male-reported Constructive Communication, which showed the greatest improvement in reliability, and female-reported Self-Demand/Partner-Withdraw, which showed the lowest improvement in reliability. Table 5 displays sample sizes needed to achieve .8 power across various ρ and other-scale α values. For the greatest power improvement (male-reported Constructive Communication), the sample size needed to achieve .8 power was 29.4% to 31.3%

Table 4
Cronbach's Alpha in All Four Samples Using Original and Revised Scoring for the Communication Patterns Questionnaire

Sample	Males					Females				
	Old α	New α	Diff.	χ^2	Prop. incr.	Old α	New α	Diff.	χ^2	Prop. incr.
Clinical trial										
CC	.566	.801^a	.235	22.9***	.415	.650	.808^a	.158	18.6***	.243
SD/PW	.618	.782^a	.164	26.6***	.265	.645	.770^a	.125	17.3***	.194
PD/SW	.541	.736^a	.195	17.4***	.360	.615	.751^a	.136	14.0***	.221
Community										
CC	.790	.845	.055	14.0***	.070	.781	.863	.082	29.3***	.105
SD/PW	.679	.805^a	.126	65.0***	.186	.657	.813^a	.156	128.7***	.237
PD/SW	.722	.820	.098	58.1***	.136	.699	.821^a	.122	88.6***	.175
Clinic										
CC	.757	.803	.046	1.0	.061	.720	.812	.092	3.5 ⁺	.128
SD/PW	.647	.765^a	.118	9.2**	.182	.689	.795^a	.106	9.6**	.154
PD/SW	.654	.804^a	.150	12.8***	.229	.532	.726^a	.194	11.4***	.365
Divorcing										
CC	.562	.666	.104	1.2	.185	.569	.720^a	.151	3.2 ⁺	.265
SD/PW	.634	.617	-.017	.1	-.027	.642	.802^a	.160	16.0***	.249
PD/SW	.622	.794^a	.172	12.5***	.277	.758	.814	.056	2.7	.074

Note. Bolded numbers represent cases in which the intraclass correlation coefficient (ICC; Cronbach's α) is at or above .7. Old α = ICC of original scoring; New α = ICC of revised scoring; Diff. = difference (subtracted) between α of revised scoring and original scoring; χ^2 = chi-square test of difference of old and new Cronbach's α ; Prop. incr. = proportion increase in α when moving from original to revised scoring; CC = constructive communication; SD = self-demand/partner-withdraw; PD/SW = partner-demand/self-withdraw.

^a Went from below .7 (original scoring) to above .7 (revised scoring).

⁺ $p < .10$. ** $p < .01$. *** $p < .001$.

lower when using the revised scoring, compared with the original scoring. For the smallest power improvement (female-reported Self-Demand/Partner-Withdraw), the sample size needed to achieve .8 power was 16.1% to 17.9% lower when using the revised scoring. Thus, across all scales in the clinical trial sample, sample sizes needed to achieve .8 power when using the CPQ are reduced by 16.1% to 31.3% when using the revised, compared with original, scoring.

Associations With Relationship Satisfaction and Sensitivity to Change

We next tested convergent validity by examining correlations between relationship satisfaction and CPQ subscales using original and revised scoring in the clinical trial, community, and clinic samples (satisfaction data were not available in the divorcing sample; see Table A5). Using the original scoring, 25 of

Table 5
Sample Sizes Needed to Achieve .8 Power, Using Example of Largest and Smallest Alpha Improvement in Clinical Trial Sample

ρ	Internal reliability (Cronbach's α) of other measure											
	.60		.65		.70		.75		.80		.85	
	Old	New	Old	New	Old	New	Old	New	Old	New	Old	New
Male-reported constructive communication												
.25	367	259	339	238	314	221	293	206	275	193	258	182
.30	179	254	234	165	217	153	203	142	190	133	179	125
.35	131	186	171	120	159	111	148	104	139	97	130	91
.40	99	142	131	91	121	85	113	79	106	74	99	69
.45	78	111	103	72	95	66	89	62	83	58	78	54
.50	63	90	72	58	67	53	62	49	58	46	54	43
Female-reported self-demand/partner-withdraw												
.25	269	322	297	248	275	230	257	215	241	201	226	189
.30	186	223	205	171	190	159	178	148	166	139	156	130
.35	136	163	150	125	139	116	130	108	121	101	114	95
.40	103	124	114	95	106	88	99	82	92	77	87	72
.45	81	97	90	75	83	69	77	64	72	60	68	56
.50	65	78	72	60	67	55	62	52	58	48	54	45

Note. ρ = true correlation between CPQ scale and other measure; Old = sample size required for .8 power using original scoring; new = sample size required for .8 power using revised scoring.

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

the 36 total correlations (3 samples \times 2 CPQs [male and female report] \times 3 subscales \times 2 DASs [male and female report]) were significant, whereas 28 were significant using the revised scoring. Twenty-seven of the 36 correlations were relatively stronger using the revised scoring compared with the original, whereas nine were relatively weaker. Nine of the stronger correlations were significantly larger, and two were larger at a trend level. No correlations were meaningfully larger when using the original subscales.

The ability of the CPQ to detect change in communication behavior over a course of couple therapy was examined using a series of MLMs. Because different numbers of items are included in each scale using the revised and original scoring systems, scale scores were generated using the mean of all items on a given scale. Additionally, demand/withdraw subscales were recoded to Female-Demand/Male-Withdraw and Male-Demand/Female-Withdraw in order for the specific type of behavior reported (one person demanding and the other withdrawing) to be the same regardless of reporter. Participants in the clinical trial sample were assigned to one of two couple therapies, Integrative Behavioral Couple Therapy (IBCT) or Traditional Behavioral Couple Therapy (TBCT; see Christensen et al., 2004, for a description of both therapies). Previous research has found that these two treatments produce significantly different amounts of change in positive and negative behaviors measured using observational coding methods during the active treatment phase (K. J. Baucom et al., 2015), so interactions involving type of treatment were included to test for possible treatment differences in all models. In the first analysis, each behavior was regressed onto an effect-coded variable for partner ($-.5 =$ male, $.5 =$ female), a dummy-coded variable indicating scoring system ($0 =$ revised, $1 =$ original), a dummy-coded variable indicating pretreatment (0) versus posttherapy (1), an effect-coded variable indicating type of treatment ($-.5 =$ IBCT, $.5 =$ TBCT), as well as the two-, three- and four-way interactions among these variables. There were no significant interactions involving scoring system for any behavior, indicating that the change in mean levels of behaviors is not significantly different using the original and revised scoring systems.

To further compare the ability of the original and revised scoring systems to detect significant changes in each type of behavior over the course of treatment, a series of MLMs was run in which each type of behavior was regressed onto the effect-coded variable for partner, the dummy-coded variable indicating pretreatment versus posttherapy, the effect-coded variable indicating type of treatment, and a two-way interaction between pre-/posttreatment and type of therapy. These models were run separately for each scoring system using sample sizes ranging from $n = 120$ to $n = 30$ using a bootstrapped resampling procedure that included 100 draws per model in which samples were selected using a stratified (type of treatment), clustered (partner) design with replacement. A sample size of $n = 30$ was selected as the lower limit of analyses because it is approximately equal to the average sample size ($n = 30.35$) of the 40 published outcome trials of behaviorally based couple therapy reported in Shadish and Baldwin (2005), Christensen et al. (2004), and Johnson, Hunsley, Greenberg, and Schindler (1999) combined. As reported in Table A6, significant changes emerged from pretreatment to posttherapy for Constructive Communication, Male-Demand/Female-Withdraw, and

Female-Demand/Male-Withdraw using the original and revised scoring systems for all sample sizes. A significant treatment by time effect emerged for Constructive Communication in sample sizes of $n \geq 40$ for the revised scoring and $n \geq 50$ for the original scoring. Similarly, the p value of the pretreatment to posttherapy effect for Male-Demand/Female-Withdraw was approaching the commonly accepted cutoff of $p < .05$ at $n = 30$ using the original scoring system but not using the revised scoring system ($p = .012$). Results for Female-Demand/Male-Withdraw were largely equivalent at each sample size across the two scoring systems. This collection of results suggests that both scoring systems are able to detect significant change in all three behaviors, and that the revised scoring system appears to be somewhat more sensitive to change over time at small sample sizes for Constructive Communication and Male-Demand/Female-Withdraw.

Discussion

The present study investigated the factor structure of the CPQ in four samples of heterosexual married couples. The primary aims of this study were to reexamine the optimal number of factors in the CPQ and to examine whether there was empirical justification for including additional, conceptually similar CPQ items in the scoring of its subscales. EFAs on the clinical trial sample and CFAs on the replication sample found that a three-factor solution provided an optimal fit for the data. Four items initially selected in EFAs were subsequently identified as driving misfit in CFAs and were dropped from the final subscales. The final scales were: Constructive Communication (nine items: 2, 6, 8, 23, 25, 27, plus reverse-scored Items 1, 24, and 26), Self-Demand/Partner-Withdraw (seven items: 3, 9, 11, 13, 17, 19, and 32), and Partner-Demand/Self-Withdraw (seven items: 4, 10, 12, 14, 18, 20, and 33).² Additional analyses showed that the revised scoring generally had significantly higher internal reliabilities than the original, had equal or larger correlations with relationship satisfaction, and was more sensitive to change over time at small sample sizes. Taken together, this collection of results strongly suggests that the revised scoring system offers substantial improvements over the original scoring system, and that the revised scoring system should be used in place of the original in future research and clinical practice. We consider the detailed results of each analysis in turn.

Overall, results of CFAs showed that factor loadings for each subscale were significant, in the same direction, and largely consistent across samples, although some subscales in the divorcing sample were sensitive to the specification of priors. Posterior predictive p values found that the model for 12 of the 18 scale-sample-sex combinations provided a less than perfect reproduction of the data. This result is not completely surprising, as the three-factor model in the clinical trial sample accounted for just 29.35% of the overall item variance in the CPQ. This low variance accounted for suggests that even though items generally loaded strongly, there is still a substantial amount of variance in the partners' responses to the items that

² Subscale values are computed by adding up all items within the subscale. Items 1, 24, and 26 on the Constructive Communication subscale should be reverse scored by subtracting each raw value from 10. Those interested may contact the first author for a free copy of the CPQ.

is unrelated to those subscales. This remaining variance may ultimately result in greater measurement error of demand/withdraw and constructive communication than is ideal, even though the revision results in a considerable improvement. Finally, tests of “weak” factorial invariance, or equivalency of loadings, for men and women on the IBCT sample were non-significant, indicating that item loadings on each scale are equivalent across sex.

It is important to note that the purpose of this study was to improve the scoring for an existing, widely used measure, not to test a measurement model *per se*. Thus, we were less conservative in our evaluation of the CFA model fit than one might be if developing a new scale. Conceptual considerations also weighed heavily in our evaluation of the revised subscales. The addition of four new items to the demand/withdraw scales helps to capture demand/withdraw behavior in a broader set of circumstances, assessing a fuller range of conceptually similar behaviors under the umbrella of demand/withdraw. The revised scoring may thus identify previously missed couples who simply engage in different types of demand/withdraw behavior. That is, couples may manifest the demand/withdraw pattern in different ways (e.g., pressuring instead of nagging), but the behaviors may serve the same function and may have developed from the same cycle of polarization hypothesized to contribute to the development, perpetuation, and worsening of this destructive behavior pattern (B. R. Baucom & Atkins, 2013).

The additional items may also better distinguish between couples at the higher end of the spectrum of demand/withdraw behaviors. The original demand/withdraw items (start/avoid discussion, nag/withdraw, criticize/defend) are relatively mild compared with some of the added items (pressure for action/resist, pressure to apologize/resist, threaten/give in, call names or attack character). Demand/withdraw behavior is thought to emerge over time, with behaviors becoming more extreme through a cycle of intermittent reinforcement (e.g., B. R. Baucom & Atkins, 2013). As such, there may be a rough sequence or hierarchy of behaviors in which couples early in the polarization process attempt to coerce their partner by nagging (one of the original items), for example, but move on to the more destructive behavior of threatening (one of the additional items) later in the polarization process. The revised demand/withdraw subscales may thus better distinguish levels of dysfunction among couples in the extreme range of demand/withdraw behavior, compared with the original scoring. However, the present study does not test whether individual items provide different information value at various points on the spectrum of couple functioning; such a question is better addressed by item response theory (IRT; e.g., Embretson & Reise, 2013), which would be a valuable direction for future research.

Psychometric properties of the CPQ using the revised scales were substantially improved in all four samples compared with the original scales. ICCs using the revised scoring were significantly larger than when using the original scoring for 18 out of 24 subscale-sample-sex combinations. Such improvements in reliability result in substantially improved power to detect relationships with other variables, improving the CPQ’s utility in empirical studies. In addition, one of the main concerns with the CPQ as it was previously used was the substantial variability in internal reliability across samples (e.g., Christensen et al., 2006). The

revised scoring results in much greater consistency in internal reliability of the subscale scores across the four samples examined. This improved consistency suggests that the revised scoring is a good fit across the range of couple functioning, providing strong justification for its use in samples ranging from satisfied to severely distressed couples, and with couples who present for treatment in clinical trials, independent practice, community clinics, or other nonuniversity settings.

We also examined convergent validity of the subscales by comparing correlations of original and revised scales with relationship satisfaction, and by examining sensitivity to change from treatment. Twenty-seven of 36 correlations with relationship satisfaction were relatively larger when using the revised scoring compared with the original, though just nine of those were statistically significant. Both the revised and original scoring systems were able to detect significant changes in behavior over the course of treatment, and the magnitude of these changes in behavior was not significantly different across scoring systems. The revised scoring appears to be more sensitive to change at small sample sizes, but not at moderate to large sample sizes.

In sum, results of the current study demonstrate that although the original scoring for the CPQ is adequate, the revised scoring represents a substantial improvement overall, and we recommend its use in place of the original scoring in future research and clinical applications. Existing data can also be reanalyzed using the revised scales, as the items themselves remain unchanged. EFAs found three factors in the CPQ that were consistent for men and women and added additional items to each subscale. The factor solution was largely confirmed using CFAs on three diverse samples. Additionally, 18 of 24 internal reliabilities were significantly larger when using the revised scoring compared with the original, which translates into substantially improved power to detect relationships with other variables in the revised scoring. Lastly, the revised subscales demonstrate improved construct validity by, overall, having stronger associations with relationship satisfaction and being better able to predict change in therapy.

There are several important limitations to the current study. First, we examined CPQ data only from heterosexual married couples. At least one study using observational coding has found that same-sex couples engage in demand/withdraw behavior in ways highly similar to heterosexual couples (B. R. Baucom et al., 2010), and other studies have confirmed the utility of the CPQ in same-sex couples (e.g., Kurdek, 2004). However, we did not test the revised CPQ scoring with same-sex couples because of the unavailability of such data. Similarly, we examined only couples living within the United States. Finally, factor loadings in the divorcing sample were substantially more sensitive to priors than were those in the community or clinic samples. It is difficult to know whether this increased sensitivity was related to a restricted range of behavior present in divorcing couples, the smaller sample size of the divorcing sample, or some combination of the two. Despite this uncertainty, parameter estimates obtained for the divorcing sample using empirical priors were similar to those obtained for the other samples and demonstrate the acceptability of the revised scoring method for use in divorcing samples.

Conclusions and Future Directions

The findings of the current study improve the utility of the CPQ for both research and practice settings. The improved reliability of the revised scale scores directly translates into improved power for detecting meaningful relationships with other variables in empirical research. In applied settings, the revised scales allow for more accurate assessment of communication behavior in order to better inform treatment plans and to better assess treatment progress in couple therapy. Additionally, the current study found that the three-factor conceptualization of the CPQ can validly describe communication behavior across a range of couple functioning, from well-functioning couples in the community to couples in the process of getting divorced, and the loadings of items on subscales was found to be equivalent for men and women. Future research should examine the revised scales in same-sex couples and couples outside of the United States. IRT analysis on a large sample may also be fruitful for improving measurement of couple communication behavior by testing whether the items added to the demand/withdraw scales can better measure more extreme demand/withdraw behavior. These future directions could contribute to continued refinement of measurement of couple communication behavior and address some of the remaining issues with the CPQ. However, the CPQ has proven to be a highly useful self-report measure both in research and applied settings, and the current study both further confirms its utility across the range of couple functioning and enhances its utility in all examined contexts through improved reliability and conceptual clarity.

References

- Balderrama-Durbin, C. M., Allen, E. S., & Rhoades, G. K. (2012). Demand and withdraw behaviors in couples with a history of infidelity. *Journal of Family Psychology, 26*, 11–17. <http://dx.doi.org/10.1037/a0026756>
- Baucom, B. R., & Atkins, D. C. (2013). Understanding marital distress: Polarization processes. In M. A. Fine & F. D. Fincham (Eds.), *Handbook of family theories: A content-based approach* (pp. 145–166). New York, NY: Routledge.
- Baucom, B. R., McFarland, P. T., & Christensen, A. (2010). Gender, topic, and time in observed demand-withdraw interaction in cross- and same-sex couples. *Journal of Family Psychology, 24*, 233–242. <http://dx.doi.org/10.1037/a0019717>
- Baucom, D. H., Epstein, N., Rankin, L. A., & Burnett, C. K. (1996). Assessing relationship standards: The Inventory of Specific Relationship Standards. *Journal of Family Psychology, 10*, 72–88. <http://dx.doi.org/10.1037/0893-3200.10.1.72>
- Baucom, K. J., Baucum, B. R., & Christensen, A. (2012). Do the naïve know best? The predictive power of naïve ratings of couple interactions. *Psychological Assessment, 24*, 983–994. <http://dx.doi.org/10.1037/a0028680>
- Baucum, K. J., Baucum, B. R., & Christensen, A. (2015). Changes in dyadic communication during and after integrative and traditional behavioral couple therapy. *Behaviour Research and Therapy, 65*, 18–28. <http://dx.doi.org/10.1016/j.brat.2014.12.004>
- Bodenmann, G., Kaiser, A., Hahlweg, K., & Fehm-Wolfsdorf, G. (1998). Communication patterns during marital conflict: A cross-cultural replication. *Personal Relationships, 5*, 343–356. <http://dx.doi.org/10.1111/j.1475-6811.1998.tb00176.x>
- Christensen, A. (1987). Detection of conflict patterns in couples. In K. Hahlweg & M. J. Goldstein (Eds.), *Understanding major mental disorder: The contribution of family interaction research* (pp. 250–265). New York, NY: Family Process Press.
- Christensen, A., Atkins, D. C., Berns, S., Wheeler, J., Baucom, D. H., & Simpson, L. E. (2004). Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples. *Journal of Consulting and Clinical Psychology, 72*, 176–191. <http://dx.doi.org/10.1037/0022-006X.72.2.176>
- Christensen, A., Eldridge, K., Catta-Preta, A. B., Lim, V. R., & Santagata, R. (2006). Cross-cultural consistency of the demand/withdraw interaction pattern in couples. *Journal of Marriage and Family, 68*, 1029–1044. <http://dx.doi.org/10.1111/j.1741-3737.2006.00311.x>
- Christensen, A., & Shenk, J. L. (1991). Communication, conflict, and psychological distance in nondistressed, clinic, and divorcing couples. *Journal of Consulting and Clinical Psychology, 59*, 458–463. <http://dx.doi.org/10.1037/0022-006X.59.3.458>
- Diedenhofen, B. (2016). *cocron: Statistical comparisons of two or more alpha coefficients*. R package version 1.0–1. Retrieved from <https://cran.r-project.org/web/packages/cocron/>
- Doss, B. D., Simpson, L. E., & Christensen, A. (2004). Why do couples seek marital therapy? *Professional Psychology: Research and Practice, 35*, 608–614. <http://dx.doi.org/10.1037/0735-7028.35.6.608>
- Eldridge, K. A., & Baucom, B. (2012). Demand-withdraw communication in couples. In P. Noller & G. C. Karantzas (Eds.), *The Wiley-Blackwell handbook of couples and family relationships* (pp. 144–158). West Sussex, UK: Wiley Blackwell.
- Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11*, 93–103. <http://dx.doi.org/10.1177/014662168701100107>
- Fincham, F. D., & Beach, S. R. (2002). Forgiveness in marriage: Implications for psychological aggression and constructive communication. *Personal Relationships, 9*, 239–251. <http://dx.doi.org/10.1111/1475-6811.00016>
- Gottman, J. M., & Levenson, R. W. (2000). The timing of divorce: Predicting when a couple will divorce over a 14-year period. *Journal of Marriage and the Family, 62*, 737–745. <http://dx.doi.org/10.1111/j.1741-3737.2000.00737.x>
- Harris, L. E. (1992). *Marital conflict and divorce: A cross-cultural study of conciliation court participants* (Unpublished doctoral dissertation). Los Angeles, CA: University of California, Los Angeles.
- Heavey, C., Gill, D. S., & Christensen, A. (1998). *The Couple Interaction Rating System* (Unpublished document). Los Angeles, CA: University of California, Los Angeles.
- Heavey, C. L., Larson, B. M., Zumtobel, D. C., & Christensen, A. (1996). The Communication Patterns Questionnaire: The reliability and validity of a constructive communication subscale. *Journal of Marriage and the Family, 58*, 796–800. <http://dx.doi.org/10.2307/353737>
- Holtzworth-Munroe, A., Smutzler, N., & Stuart, G. L. (1998). Demand and withdraw communication among couples experiencing husband violence. *Journal of Consulting and Clinical Psychology, 66*, 731–743. <http://dx.doi.org/10.1037/0022-006X.66.5.731>
- Johnson, S. M., Hunsley, J., Greenberg, L., & Schindler, D. (1999). Emotionally focused couples therapy: Status and challenges. *Clinical Psychology: Science and Practice, 6*, 67–79. <http://dx.doi.org/10.1093/clipsy.6.1.67>
- Karney, B. R., & Bradbury, T. N. (1995). The longitudinal course of marital quality and stability: A review of theory, method, and research. *Psychological Bulletin, 118*, 3–34. <http://dx.doi.org/10.1037/0033-2909.118.1.3>
- Kelly, A. B., Halford, W. K., & Young, R. M. (2002). Couple communication and female problem drinking: A behavioral observation study. *Psychology of Addictive Behaviors, 16*, 269–271. <http://dx.doi.org/10.1037/0893-164X.16.3.269>

- Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Press.
- Kurdek, L. A. (2004). Are gay and lesbian cohabiting couples really different from heterosexual married couples? *Journal of Marriage and Family*, 66, 880–900. <http://dx.doi.org/10.1111/j.0022-2445.2004.00060.x>
- Lee, I. A., & Preacher, K. J. (2013, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Retrieved from <http://quantpsy.org>
- Lee, S. Y., & Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*, 39, 653–686. http://dx.doi.org/10.1207/s15327906mbr3904_4
- Litzinger, S., & Gordon, K. C. (2005). Exploring relationships among communication, sexual satisfaction, and marital satisfaction. *Journal of Sex & Marital Therapy*, 31, 409–424. <http://dx.doi.org/10.1080/00926230591006719>
- McGinn, M. M., McFarland, P. T., & Christensen, A. (2009). Antecedents and consequences of demand/withdraw. *Journal of Family Psychology*, 23, 749–757. <http://dx.doi.org/10.1037/a0016185>
- Muthén, B. (2010). *Bayesian analysis in Mplus: A brief introduction* (Unpublished manuscript). Retrieved from www.statmodel.com/download/IntroBayesVersion, 203.
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17, 313–335. <http://dx.doi.org/10.1037/a0026802>
- Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.
- Noller, P., & White, A. (1990). The validity of the Communication Patterns Questionnaire. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2, 478–482. <http://dx.doi.org/10.1037/1040-3590.2.4.478>
- Rehman, U. S., Ginting, J., Karimiha, G., & Goodnight, J. A. (2010). Revisiting the relationship between depressive symptoms and marital communication using an experimental paradigm: The moderating effect of acute sad mood. *Behaviour Research and Therapy*, 48, 97–105. <http://dx.doi.org/10.1016/j.brat.2009.09.013>
- Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, 29, 304–321. <http://dx.doi.org/10.1177/0734282911406653>
- Schrodt, P., Witt, P. L., & Shimkowski, J. R. (2014). A meta-analytical review of the demand/withdraw pattern of interaction and its associations with individual, relational, and communicative outcomes. *Communication Monographs*, 81, 28–58. <http://dx.doi.org/10.1080/03637751.2013.813632>
- Shadish, W. R., & Baldwin, S. A. (2005). Effects of behavioral marital therapy: A meta-analysis of randomized controlled trials. *Journal of Consulting and Clinical Psychology*, 73, 6–14. <http://dx.doi.org/10.1037/0022-006X.73.1.6>
- Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family*, 38, 15–28. <http://dx.doi.org/10.2307/350547>
- Sullaway, M., & Christensen, A. (1983). Assessment of dysfunctional interaction patterns in couples. *Journal of Marriage and the Family*, 45, 653–660. <http://dx.doi.org/10.2307/351670>
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.
- Woodin, E. M. (2011). A two-dimensional approach to relationship conflict: Meta-analytic findings. *Journal of Family Psychology*, 25, 325–335. <http://dx.doi.org/10.1037/a0023791>

Received January 2, 2016

Revision received July 25, 2016

Accepted July 26, 2016 ■