Psychological Assessment
2002, Vol. 14, No. 4, 423–438

1040-3590/02/$5.00 DOI: 10.1037//1040-3590.14.4.423

The Observation of Specific Affect in Marital Interactions: Psychometric Properties of a Coding System and a Rating System

Matthew D. Johnson
State University of New York at Binghamton

The Specific Affect Coding System (SPAFF; J. M. Gottman & L. J. Krokoff, 1989) has led to conclusions about which types of dyadic affect predict positive and negative outcomes in marriage, yet the lack of information about collinearity among the codes limits interpretation of SPAFF results. Psychometric properties of SPAFF were examined by assessing the interactions of 172 newlywed couples with SPAFF and with an affect rating system developed for this study. For husbands and wives, factor analysis indicated 4 distinct factors of affect, representing anger/contempt, sadness, anxiety, and humor/affection. Anger/contempt and humor/affection were associated with marital satisfaction, relationship beliefs, and appraisals of the interactions. Correlations were in the expected directions. The strengths, limitations, and implications of the data are discussed.

Expressions of positive emotion between spouses are associated with happier marriages and dyadic displays of negative emotion are associated with marital distress (for a review, see Heyman, 2001). These findings led to questions about which specific types of affect predict marital discord and happiness. To answer these questions, Gottman and Krokoff (1989) developed the Specific Affect Coding System (SPAFF). Other observational coding systems that assessed affect did so by dividing affect into positive, negative, and neutral categories (e.g., Couples Interaction Scoring System; Gottman, 1979; Kategoriensystem für Partnerschaftliche Interaktion; Hahlweg et al., 1984). However, it seemed likely that specific negative emotions, such as anger and sadness, might function differently from each other and therefore should be considered separately. For example, research has suggested that verbally aggressive and depressive behaviors in problem-solving interactions play different roles in the maintenance of depression (Biglan et al., 1985), marital discord (Schmaling & Jacobson, 1990), and dysfunctional problem solving (Nelson & Beach, 1990).

Although SPAFF has been used in marriage research many times (for reviews, see Gottman, 1994; Heyman, 2001), the psychometric properties of the SPAFF system have not been examined. A systematic analysis of the intercorrelations of the SPAFF system would aid in understanding results suggesting differential effects of individual SPAFF codes (e.g., Gottman, 1994). Katz and Gottman (1993) conducted a factor analysis on a subset of SPAFF codes as part of a larger study examining the relation of marital interaction behaviors and child outcomes. Five wife SPAFF codes (contempt, belligerence, anger, defensiveness, and domineering) and four husband SPAFF codes (contempt, belligerence, anger, and stonewalling) were factor analyzed together. Katz and Gottman reported that two factors emerged. They termed the first factor *Mutually Hostile*, which consisted of five codes with factor loadings greater than .50: wife contempt, wife belligerence, husband contempt, husband belligerence, and wife anger. The second factor was termed *Husband Angry and Withdrawn*, and it consisted of two behavioral codes with factor loadings greater than .50: husband anger and husband stonewalling. The Katz and Gottman study was neither intended as nor served as a complete explication of the intercorrelations of the SPAFF coding system, leaving the psychometric properties of the SPAFF coding system not fully examined.

Representative Findings Obtained With SPAFF

The first use of SPAFF was by Gottman and Krokoff (1989) and involved five negative affect codes (anger, disgust/contempt, whining, sadness, and fear/anxiety) and five positive affect codes (affection, humor, interest, anticipation, excitement/joy). Significant correlations were found between contempt (husband and wife) and concurrent distress (husband and wife; *r*s ranged from $-.48$ to $-.37$), and cross-sectional correlations were found between wife anger and marital satisfaction of husbands ($r = -.35$) and wives ($r = -.42$). However, when behaviors were used to predict change in marital satisfaction, the direction of these effects

These data were collected as part of a larger study of marital development during the early years of marriage and funded by National Institute of Mental Health Grant R29 MH48674 awarded to Thomas Bradbury. The observation of these data was supported by a dissertation grant from the American Psychological Association and a grant from the Fahs-Beck Fund for Research awarded to Matthew D. Johnson. Parts of this article were presented at the biennial meeting of the International Society for the Study of Personal Relationships, Brisbane, Queensland, Australia, June 2000.

Karen Barad, Dakota Bayard, Thomas Bradbury, Jessica Burry, Jennifer Christian-Herman, Kahni Clements, Catherine Cohan, Joanne Davila, Alessia Johns, Benjamin Karney, Jennifer Kitahara, Kirsten Krohn, Erika Lawrence, Nina Lee, Gregory Miller, Kelley Nay, Melody Nosce, Mo Ono, Ronald Rogge, Amita Shenoy, Kieran Sullivan, Linda Taniguchi, and Dan Yaeger helped with data collection and observation. Thomas Bradbury, Catherine Cohan, and Carolyn Pepper commented on earlier versions of this article. Their contributions are gratefully acknowledged.

Correspondence concerning this article should be addressed to Matthew D. Johnson, Department of Psychology, State University of New York at Binghamton, Binghamton, New York 13902-6000. E-mail: mjohnson@binghamton.edu

was either nullified or reversed, with wife anger and contempt predicting positive changes in wife marital satisfaction (*r*s = .26 and .27, respectively). This controversial finding may be a result of statistical problems resulting from the use of marital satisfaction difference scores as a criterion variable when behavior and Time 1 satisfaction are highly correlated and when Time 1 and Time 2 satisfaction are highly correlated (Bradbury, Cohan, & Karney, 1998; Woody & Costanzo, 1990; cf. Gottman & Krokoff, 1990).

In his book, Gottman (1994) reviewed the findings based on the SPAFF coding system. In the review, he listed 15 SPAFF codes that were used to predict concurrent marital quality, subsequent marital quality (2 years later), considerations of separation, actual separation, months separated, and divorce. Spouses were coded separately and entered into the analyses separately, but marital quality was considered only at the couple level. Of the husbands' SPAFF codes that represent affect (e.g., anger) rather than behavioral skills (e.g., domineering), husbands' tension predicted subsequent marital quality ($r = .37$). Surprisingly, contempt was associated with a lower rating on serious considerations of divorce ($r = -.29$). For wives, contempt was associated with lower follow-up marital satisfaction ($r = -.27$) and more considerations of divorce ($r = .34$). In addition, wives' humor was associated with concurrent marital satisfaction ($r = .51$) and wives' joy was associated with subsequent marital satisfaction ($r = .30$). Gottman (1994) also reported a few surprising findings that do not fit with the other data and appear to be anomalies (e.g., expressions of wife interest were associated with separation, $r = .27$, $p < .05$).

On the basis of the findings outlined in Gottman (1994), Gottman, Coan, Carrère, and Swanson (1998) hypothesized that "anger in marital interactions [would] not predict divorce" (p. 6), but that behaviors "called 'The Four Horsemen of the Apocalypse'—criticism, defensiveness, contempt, and 'stonewalling' (listener withdrawal)—reliably [would] predict divorce" (p. 6). After stating this hypothesis, Gottman et al. (1998) tested "anger as a destructive emotion versus the 'Four Horsemen' " (p. 6) as two competing theories of the effects of negative affectivity in marriage. To test the hypothesis, Gottman et al. (1998) used SPAFF to assess affect displayed by newlyweds having problem-solving discussions. The results indicated the ability of the following sets of behavior to predict divorce: combination of belligerence, defensiveness, and contempt (for husbands: $r_{effect} = .87$; for wives: $r_{effect} = .67$); anger alone (for husbands: $r_{effect} = .06$; for wives: $r_{effect} = .43$); combination of whining, anger, sadness, domineering, disgust, fear, and stonewalling (for husbands: $r_{effect} = .30$; for wives: $r_{effect} = .69$). However, methodological concerns have led some marital researchers to question these findings (e.g., Stanley, Bradbury, & Markman, 2000; cf. Gottman, Coan, Carrère, & Swanson, 2000). The results of Gottman and Krokoff (1989), Gottman (1994), and Gottman et al. (1998) led to the conclusion in the last of these articles that there is "no evidence in the study of newlyweds to support the model of anger as the destructive emotion in marriage" (pp. 16–17), but rather that contempt was the destructive emotion present in conflict interactions.[1]

The Problem of Determining Intercorrelations of Coded Data

The lack of information about the intercorrelations of the SPAFF codes limits the conclusions that may be drawn from studies using SPAFF. For example, if the codes of anger and contempt are assessing different constructs, this would lead to very different conclusions than if contempt was operationally and functionally being coded as a more severe form of anger. Furthermore, if variables are not orthogonal, the power of multiple regression or multivariate methods of data analysis is limited. Therefore, behavioral codes that overlap significantly cause problems of multicollinearity during data analysis. To get around this problem, researchers often analyze each variable separately; however, the increased number of tests may inflate the likelihood of Type I error. Ideally, the number of coded variables would be reduced into factors, but this poses a problem as well. Often, observational coding data cannot be reduced using factor analysis because in many coding systems (e.g., SPAFF), each behavioral code is assigned to the exclusion of the other codes. In other words, the assignment of Code A prevents Codes B and C from being assigned to a particular event. Yet one of the assumptions of factor analysis is that one observed variable does not preclude another, allowing variance to be shared and thus allowing the detection of possible latent variables (i.e., factors). Therefore, empirically deriving factors from coding systems that use mutually exclusive codes, such as SPAFF, is difficult and encumbers the accurate assessment of the intercorrelations of SPAFF codes.

Developing Affect Factors

To circumvent the problem of mutually exclusive codes violating an assumption of factor analysis while maintaining the integrity of SPAFF as it has been used previously, researchers developed the Behavioral Affective Rating System (BARS; M. D. Johnson, Johns, Kitahara, Ono, & Bradbury, 1998) to assess the same emotions assessed by the SPAFF coding system. A coding system (i.e., presence or absence of a behavior) and a rating system are different in that a coding system labels discrete behaviors in discrete units (e.g., anger characterized 5 s of speech), whereas a rating system assesses a set of behaviors during an interval of time and indicates the magnitude of each specific behavior (e.g., anger, sadness, anxiety) using a Likert-type scale. The advantage of a rating system over coding systems with mutually exclusive codes is that patterns of covariation can be more easily detected through the use of factor analysis.[2] The BARS system was developed to assess specific affect using ratings instead of codes, thus allowing the process of observing dyadic affect to be psychometrically examined and validated.

In the current study, the marital problem-solving conversations of 172 newlywed couples were observed with the SPAFF system and with the BARS system, which allowed the BARS data to be factor analyzed as a means of examining the intercorrelations of affective behaviors. In addition, assessing behavior with SPAFF and BARS allowed the two systems to be compared in terms of their ability to replicate behavioral associations with other constructs (e.g., marital satisfaction and dysfunctional relationship

[1] They also noted belligerence and defensiveness as "destructive patterns," but these are not discussed in this article as they are generally not considered emotions.

[2] Some coding systems permit multiple codes to be assigned to a single behavior, which would allow the intercorrelations of the codes to be examined.

beliefs). Thus, the purposes of the present study were to examine the psychometric properties of SPAFF and to assess the validity of BARS as a measure of important behaviors in marital interactions.

Method

Participants

All participants were recruited from marriage licenses filed in Los Angeles County between May 1993 and January 1994; couples were sent letters inviting them to participate in the study. Of the 3,606 letters that were sent, 637 couples (17.8%) expressed interest in participating, 41 letters were not deliverable (1.1%), and 2,928 letters (81.2%) went unanswered. The 17.8% response rate is similar to that of other studies recruiting married couples from public records (e.g., 18% by Kurdek, 1991). Couples who expressed interest in participating were screened for eligibility criteria in a telephone interview. To be eligible to participate, the husband and wife had to have been previously unmarried; in addition, they had to be over 18 years of age; have at least a 10th-grade education; be able to speak, read, and write English; have no children; and have no plans to leave the Los Angeles area within a year. To allow for the possibility that all couples could become parents during the course of the project, wives had to be less than 35 years of age. These eligibility criteria represented an attempt to develop a homogeneous sample of newlywed couples designed to explore several facets of marriage. The effects of using this sampling strategy and these criteria are reviewed in Karney et al. (1995). The first 172 couples who met the eligibility criteria and who kept their laboratory appointment comprised the sample. All couples were married less than 6 months when they began the study.

Wives averaged 26.0 ($SD = 3.4$) years of age with 16.2 ($SD = 3.4$) years of education, and their median annual income was between $11,000 and $20,000. Sixty-one percent were Caucasian, 15% were Asian American/Pacific Islander, 5% were African American, 16% were Latina/Chicana, 2% were Middle Eastern, and 1% self-identified as "other." Husbands averaged 27.6 ($SD = 3.9$) years of age with 15.6 ($SD = 2.2$) years of education, and their median annual income was between $21,000 and $30,000. Sixty-seven percent were Caucasian, 13% were Asian American/Pacific Islander, 4% were African American, 15% were Latino/Chicano, and 1% were Middle Eastern. The sample composition is comparable to other newlywed samples (e.g., M. D. Johnson & Bradbury, 1999) and to a 1990 census racial breakdown of Los Angeles County, which reported 57.0% Caucasian, 10.7% Asian American/Pacific Islander, 11.2% African Americans, 0.5% American Indians, and 20.6% "other."

Procedure

Eligible couples were scheduled for a 3-hr laboratory session in which spouses independently completed a set of questionnaires consisting of a consent form, demographic forms, measures of marital satisfaction, and a rating sheet of marital problems. Following a semistructured interview conducted separately with each spouse by an interviewer of the same gender (designed to assess characteristics of each spouse's current and past mental health, family of origin, chronic stresses, and perspective of the relationship history) and while the spouses were still separated, participants were asked to choose a problem to discuss from a list of marital problems (Geiss & O'Leary, 1981). The five most frequent problems selected by husbands were "tempers and moods," "communication," "in-laws, parents, relatives," "money management," and "time spent together." The five most frequent problems selected by wives were "tempers and moods," "money management," "household management," "communication," and "in-laws, parents, relatives." Each couple had two 10-min problem-solving discussions, with each spouse picking the topic of one of the discussions. Each spouse was asked to pick a topic because of the finding of differences in interaction behaviors exhibited when discussing husband topics and wife topics (Christensen & Heavey, 1990). Couples then were asked to engage in two videotaped 10-min problem-solving discussions. In each discussion, the couple was asked to work toward a resolution of an important marital complaint (e.g., money, in-laws, or sex). The experimenter asked the spouses to tell each other what topic they had chosen to discuss, and then the experimenter verified that both spouses were comfortable with both topics (or another topic was chosen). The order of the topics was determined previously by a coin toss, and the couples were then instructed to "discuss the topic for 10 minutes and try to work toward a mutually satisfying solution." If both spouses chose the same topic, that topic was assigned to the spouse whose topic was randomly chosen to be first, and the other spouse's second choice of topic was then used for the second discussion. Both spouses had the right to veto the topic chosen by the other spouse, but this option was rarely exercised.

Once the topics were chosen, the participants were given a questionnaire about their expectation of the upcoming discussion. The couples were instructed to begin each discussion once the door closed behind the experimenter. The spouses were seated in two chairs facing each other. Two cameras were set up on tripods behind and to the right of each spouse, and a split screen was used to allow the facial expressions of each partner to be observed. Sound was captured with a flat microphone that was mounted on the wall. Each discussion was interrupted at precisely 10 min with a knock on the door. Once the first discussion was finished, each spouse completed a questionnaire appraising the just-completed interaction and then, following a short break, answered a questionnaire about the upcoming (second) discussion. The same procedure was used for the second discussion. Problem-solving discussions, such as the ones used in this study, have been found to be representative of similar unobserved interactions (Foster, Caplan, & Howe, 1997).

The session concluded with the couple being debriefed, thanked, and paid $75. Spouses reported having either a neutral or positive reaction to the research, results that are consistent with prior research (Bradbury, 1994). Although none of the participants reported distress from the experience, when conflicts had clearly become heated, couples were informed that they should feel free to contact either of the experimenters or the principal investigator. None initiated such contact.

Behavioral Observation

Two behavioral observation systems were used to assess the affect displayed in the marital problem-solving interactions. Training for each system began with a series of meetings to review the system and discuss its objectives. The observers then began coding or rating sample interactions of couples who were not part of this study. Once the group had reached a consensus during these group observation tasks, all of the observers were assigned the same interaction and instructed to rate or code it outside of the group. The group would then meet to discuss the reliability of the ratings or coding of the interaction and to specifically review any points where consensus was not achieved. Once consistent intraclass correlation (ICC) coefficients of .70 or greater were achieved, observers were randomly paired with other observers to gain more experience. If ICC coefficients remained above .70, the observers were deemed ready to begin coding participant interactions. Once the observers began coding or rating on their own, weekly meetings were held to assess and maintain reliability among the observers.

Each observer watched an interaction once without assessing the behavior. Immediately following the first viewing, the observer focused on one randomly selected spouse while watching the tape a second time, and the other spouse while watching the tape a third time. The initial viewing of the interaction made tuning out the content of the conversation easier during the actual coding or rating because—it was hoped—the observer's urge to listen to what was being said and not how it was being said was reduced. None of the observers were trained in both of the systems described below to avoid carry-over effects within an observer. To determine interobserver

reliability of each system, two people independently observed 20% of the interactions. ICC coefficients of behavior were the inferential statistic used to determine interobserver reliability. ICC coefficients were chosen over a point-by-point estimate (e.g., kappa coefficient; Cohen, 1960, 1968) because they are an ideal estimate of reliability for the base rates of codes and sums of ratings (Hops, Davis, & Longoria, 1995; McGraw & Wong, 1996). Only observed behaviors with an ICC significantly different from zero were used in the following analyses.[3]

Affect coding. The problem-solving discussions were coded using SPAFF Version 1.0 (Gottman, 1994), with the modification that the codes of interest/curiosity and anticipation/surprise/excitement/enjoyment/joy were combined a priori. SPAFF is considered a "second generation" coding system, having been developed out of a coding system used to assess affect by observing facial expressions and modifying it for use in couple conflicts. SPAFF has been validated in several studies to date, and thus among marital interaction coding systems, "it has by far the best evidence of construct and criterion validity for its constructs" (Heyman, 2001, p. 25). Gottman (1994) reported SPAFF results that include other codes (e.g., belligerence, defensiveness, domineering, and stonewalling); however, these codes were not included in the original version of SPAFF (Gottman & Krokoff, 1989). The original version of SPAFF was chosen for this study because the preponderance of peer-reviewed findings based on SPAFF data have included the codes in this version and because these codes represent core constructs of affect as opposed to codes more accurately described as skill or skill-deficit codes (e.g., belligerence).[4] Trained coders were instructed to consider nonverbal cues, voice quality, and verbal content when coding affect. Originally, SPAFF was designed to separate the interaction into speaking turns so that sequential analysis could be used on the data. The present study, however, was not designed with the intent of using sequential analysis; rather, it was meant to examine the associations of base rates of specific affect. Therefore, to lower the possibility of multiple expressions of affect being present in a unit, which is often the case when coding with speaking turns, the 10-min interactions were divided into 5-s units. According to Heyman (2001), the SPAFF system has been used in 23 studies, and the unit of analysis has varied across these studies. The Gottman lab switched to online SPAFF coding that uses the amount of time each affect is displayed as the unit of analysis (Gottman, 1994), which closely approximates the method used in the present study. Each unit was coded as either one of five negative affects (anger, contempt, whining, sadness, or anxiety), one of three positive affects (humor, affection, or interest), or neutral affect. The behaviors of whining and interest were excluded from analyses because they were not reliably coded (see left side of Table 1). The number of 5-s units with each code was tallied to determine the base rates of each affect, and these base rates were used in the data analyses.

Affect rating. As a means of replicating and assessing the validity of SPAFF codes, a subset (for husband topic: $n = 156$; for wife topic: $n = 143$) of the problem-solving discussions were also rated using the BARS (M. D. Johnson et al., 1998). As with other rating systems, BARS requires an observer to quantify individual behaviors on a Likert-type scale, which is unlike the SPAFF system that requires observers to note the presence or absence of behaviors to the exclusion of other behaviors. As with SPAFF, trained raters were instructed to consider nonverbal cues, voice quality, and verbal content when rating affect. The 10-min interactions were divided into 30-s intervals. Raters stopped the tape after each 30-s unit to rate the interval for the 10 behavioral affects based solely on tone of voice, facial expression, and body movement. During each 30-s block of time, raters selected a number on a 5-point scale that represented the intensity (from *none* to *extreme*) of eight affective behaviors (affection, humor, anxiety, defensiveness, verbal aggression, scorn, frustration, and hurt). The behaviors that make up the BARS represent a replication of the SPAFF codes, excluding whining and interest. The codes of anger and contempt were separated into defensiveness, verbal aggression, scorn, and frustration by asking naive raters to describe the negative emotions they saw in a series of randomly selected interactions. These four behaviors were separated because in later versions of the SPAFF coding system, the number of codes is expanded to include behaviors that have more uncertain emotional underpinnings. Defensiveness, verbal aggression, scorn, and frustration were included separately to determine if one or more of the codes covaried with a SPAFF code other than anger and contempt. These four behaviors along with humor, affection, hurt (i.e., sadness), and anxiety were rated according to the rating scale outlined in the BARS manual (see the Appendix for details) and thus included in further analyses (see right side of Table 1).[5] The ratings for each affect were summed over the twenty 30-s segments to determine the base rates of affect as assessed by rating.

Table 1
Intraclass Correlation Coefficients Representing the Reliability of Behavioral Observations

	SPAFF coding		BARS rating	
Affect	Husbands' codes	Wives' codes	Husbands' ratings	Wives' ratings
Humor	.74**	.43**	.85**	.82**
Affection	.67**	.35**	.58**	.61**
Interest	.67**	.08	—	—
Anger	.48**	.85**	—	—
Contempt	.54**	.39**	—	—
Defensiveness	—	—	.41**	.72**
Aggression	—	—	.64**	.63**
Scorn	—	—	.79**	.49**
Frustration	—	—	.56**	.60**
Sadness/hurt	.67**	.55**	.69**	.50**
Anxiety	.55**	.74**	.71**	.51**

Note. For Specific Affect Coding System (SPAFF), $n = 61$. For Behavioral Affective Rating System (BARS), $n = 49$. Not all behaviors listed in the table were included in each of the two systems (indicated by dashes). ** $p < .01$.

Questionnaires

The marital problems discussed by the couples were selected using the Inventory of Marital Problems (IMP; Geiss & O'Leary, 1981), which asks spouses to indicate the extent to which they encounter difficulties in 19 common domains of marital disagreements (e.g., money, in-laws, sex) on an 11-point scale (1 = *not a problem*, 11 = *major problem*). Blank lines at the end of the inventory allow couples to add topics that are not otherwise listed. Following the individual interview, described previously, the researcher reviewed the IMP with each spouse individually to identify the most contentious issue, and the spouse was then encouraged to bring up that topic as the issue they wanted to discuss in the problem-solving discussion where they were to pick the topic. The manner in which the

[3] Zero is, arguably, a liberal value with which to compare the ICC coefficients; however, it seems that selecting a different and more conservative value would be even more arbitrary than the traditional value of zero. The reader is urged to carefully examine the relative strength of the ICC coefficients reported for each system.

[4] Although whining is included in Version 1.0 of SPAFF, it is also more accurately defined as a skill code and was subsequently dropped from the analyses because of the infrequency of the behavior.

[5] In addition, the BARS system includes ratings of spouses' engagement and disengagement in the interaction (see coding manual in the Appendix for details); because these behaviors are more representative of communication skill than affect, they were not included in this article.

inventory was used precluded the need for validity or reliability analyses of the questionnaire.

Convergent and discriminant validity. Measures of three traits were given to the couples to assess the convergent and discriminant validity of the BARS. The measures assessed marital satisfaction, beliefs that disagreements are destructive, and appraisals of the interaction. On the basis of previous findings and theoretical assumptions, these three constructs were assumed to be good tests of the convergent and discriminant validity of the BARS system.

Marital satisfaction has been theorized to be determined, in large part, by the behavior that spouses display toward one another (e.g., Jacobson & Margolin, 1979) and specifically by the affect they display during marital interactions (e.g., Gottman, 1994). A meta-analysis of longitudinal marital research indicated that there is evidence of an association between dyadic behavior and longitudinal marital satisfaction (Karney & Bradbury, 1995). Therefore, marital satisfaction was used as a criterion variable in this study.

The 15-item Marital Adjustment Test (MAT; Locke & Wallace, 1959) was used to assess marital satisfaction. This is a widely used measure consisting of one question about the spouse's global evaluation of the marriage; eight questions measuring the amount of agreement across different areas of possible conflict; and six questions measuring conflict resolution skills, marital cohesion, and communication. The range of the MAT score is 2 to 158, with higher numbers indicating greater satisfaction and 100 generally considered the threshold of satisfied and unsatisfied spouses. In the original validity study, 22 men and 26 women with documented marital problems were matched for gender and age with 48 spouses judged by observers to be happily married. Using a score of 100 as the cutoff between distressed and nondistressed couples, 17% of distressed couples scored at or above 100 and 4% of the well-adjusted group scored below 100 (Locke & Wallace, 1959). Previous research has indicated that the MAT may overestimate marital satisfaction (Fredman & Sherman, 1987). In this sample, the total score was significantly correlated with the first question of the MAT that queries the participant's global evaluation of the marriage (husbands: .75; wives: .80). In a different sample, the split-half reliability for this measure was found to be .90 (Locke & Wallace, 1959). This measure was chosen because of its widespread use, relative brevity (cf. Dyadic Adjustment Scale; Spanier, 1976), and relative sensitivity (cf. Quality Marriage Inventory; QMI; Norton, 1983). However, the use of the MAT in studies of marital behavior has been criticized because of items referring to dyadic behavior included on the scale (see Fincham & Bradbury, 1987; Huston, McHale, & Crouter, 1986); therefore, two other measures of marital satisfaction without behavioral items were administered at Time 1 and correlated with each other and with the MAT.

The first of the other two measures was the QMI (Norton, 1983), which has five items asking participants to rate on a 7-point scale the extent to which they agree with statements regarding their marriage (e.g., "We have a good marriage.") and one question on a 10-point scale asking about global satisfaction with the marriage. Scores on the QMI range from 6 to 45, with higher scores indicating increased satisfaction. The QMI was developed using factor analytic techniques on items with content "reflecting an essential 'goodness of the relationship gestalt'" (Norton, 1983, p. 143). In the original validation study, the QMI (a) correlated with similarity of attitudes, (b) had extremely low scores that were associated with couples indicating they would not be married in 5 years, and (c) was negatively associated with the number of times a participant seriously discussed ending the relationship with his or her spouse (Norton, 1983). The interitem reliability for the QMI in the present study was excellent, as indicated by a coefficient alpha of .93 for husbands and .91 for wives.

The second measure administered as an alternative to the MAT was based on the Semantic Differential (SMD; Osgood, Suci, & Tannenbaum, 1957), which quantifies evaluations by asking participants to rate their perception of the object—in this case their marriage—on 15 items. Each item has a 7-point scale between two opposite adjectives (e.g., good–bad, weak–strong; for a similar approach, see Huston & Vangelisti, 1991). SMD scores range from 15 to 105, with higher scores indicating higher levels of satisfaction. The measure was developed using thesaurus sampling and factor analysis. The SMD has been validated in personality and communications research (for a review of the theoretical basis, development, and validation of the SMD, see Osgood et al., 1957). This measure demonstrated excellent reliability in this sample, with alpha coefficients of .94 for husbands and .92 for wives.

These three measures were highly correlated for both husbands and wives, with effect sizes ranging from .72 to .89, $p < .01$. In addition, all three of the measures have been found to load on a single factor reflecting marital satisfaction (Karney, Bradbury, Fincham, & Sullivan, 1994). When the results of analyses involving marital satisfaction were examined with each questionnaire, the results were all in the same direction and had similar magnitude. For the sake of brevity, each participant's score on each measure of marital satisfaction was converted to a *z* score, and the mean of each individual's three *z* scores served as the indicator of marital satisfaction in all of the analyses involving marital satisfaction.

Relationship beliefs. Several theories of the development of marital distress link relationship beliefs to dyadic behaviors in relationships (e.g., Baucom & Epstein, 1990; M. D. Johnson, Rogge, Karney, & Bradbury, 2001; Karney & Bradbury, 1995). Thus, a measure of dysfunctional beliefs about relationships was included to assess the convergent validity of observed affect. The Disagreement Is Destructive subscale of the Relationship Belief Inventory (RBI; Eidelson & Epstein, 1982) was administered to both spouses to assess unrealistic beliefs about marriage. The full RBI consists of five 8-item subscales. It was developed by asking 20 marital therapists to identify relationship beliefs that seem to cause marital difficulties. The variances and item-total correlations of the resulting items were examined in a sample of 47 couples. The scale was reduced on the basis of these findings and administered to another sample of 100 couples. The item analysis in this study led to further reduction of the scale. Eidelson and Epstein (1982) found that all of the subscales were negatively correlated with marital satisfaction; however, Bradbury and Fincham (1993) found that two of the five scales were not significantly correlated with marital satisfaction. The subscale that was administered in the present study assesses the extent to which spouses believe that disagreement is destructive to a relationship (e.g., "If your partner expresses disagreement with your ideas, s/he probably does not think highly of you."). Only this subscale was used because of prior research indicating that it is highly correlated with the total score of the measure (husbands: $r = .87$, $p < .01$; wives: $r = .79$, $p < .01$) and that it has the highest internal consistency across three studies (Bradbury & Fincham, 1993). For each item, the spouse rates the degree to which the item is believed to be true (0 = *I strongly believe that the statement is false*, 5 = *I strongly believe that the statement is true*). Two of the items were reverse scored. The eight items were summed; thus the minimum score is 0 and the maximum is 40. In this sample, coefficient alpha for this scale was satisfactory and consistent with other studies using the RBI at .81 for husbands and .77 for wives.

Appraisals of marital interactions. The final criterion variable used to assess the convergent and discriminant validity of BARS was the spouses' appraisals of their own affect regarding their just-completed interactions. It was expected that observed affect would be associated with self-report affect; however, it was not expected to be perfect because of past findings suggesting that participants in dyadic interactions may have very different reports of those discussions (Christensen & Nies, 1980; Jacobson & Moore, 1981).

To determine the association of observed affect and self-reported affect in the problem-solving discussions, spouses' appraisals of the dyadic interactions were measured with the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) and with four global questions about the interaction. Participants were asked to circle a number that described how they "feel right now, at the present moment, as you think about the discussion you just completed." Following those instructions, 20

words describing different emotions (10 negative and 10 positive) were listed, with a 5-point rating scale (1 = *not at all*, 3 = *somewhat*, 5 = *very much*) next to each descriptor. The PANAS is a widely used measure, as indicated by the 1,469 citations of the original article (this figure is based on a search in the Social Sciences Citation Index). The scale was developed using factor analytic techniques to reduce 60 affective terms previously evaluated by Zevon and Tellegen (1982). The PANAS demonstrated excellent convergent and discriminant validity in both undergraduate and nonstudent samples (Watson et al., 1988). The reliability of the PANAS in the present study, as measured with coefficient alpha, was .87 for husbands and .87 for wives. For statistical analyses, the items of the PANAS that reflected negative affect were reverse scored and then a mean of all of the items was calculated; thus higher scores indicated more positive affect. If the measures of observed affect have convergent validity, the correlations of negative observed affect (e.g., anger) with the PANAS should be negative, and the correlations of positive observed affect (e.g., humor) with the PANAS should be positive.

The following four questions were developed for this project and when combined, were meant to provide a more global measure of the spouse's appraisal of the interaction: (a) "To what degree did you and your partner resolve the problem?" (measured on a 9-point scale, from *did not resolve the problem* to *resolved the problem*); (b) "How satisfied are you with the outcome of the discussion?" (9-point scale, from *not at all satisfied* to *very satisfied*); (c) "To what degree were you and your spouse able to work productively on the problem in the discussion?" (9-point scale, from *not at all able* to *very able*); (d) "What was the emotional tone of the discussion?" (9-point scale, from *negative or critical* to *positive or supportive*). The inverse of the reflection of the mean of these four items was used in the analyses to account for the severely negative skew of the distribution (as recommended by Tabachnick & Fidell, 2001). Coefficient alpha for these four items was .88 for husbands and .93 for wives.

Results

Descriptive Statistics and Preliminary Analyses

Tables 2 and 3 present the samplewide means of each specific affect assessed using SPAFF (Table 2) and BARS (Table 3). The results are presented separately for husband-selected and wife-selected topics. Of the reliably coded SPAFF behaviors, anger and anxiety had the highest means, and sadness and affection had the lowest means. Anxiety was the highest rated of the BARS behaviors, with several of the BARS behaviors having low ratings. Examining the means in relation to the standard deviations presented in Tables 2 and 3 suggests the presence of a floor effect, which is often found when assessing behaviors with low base rates. The correlation coefficients between the two topics indicate moderate and inconsistent associations between the behaviors displayed in the two topics.

Assessing the Similarity of BARS and SPAFF

Table 4 shows the associations between behavioral assessments with the BARS and SPAFF systems. These correlations indicated that similar behaviors assessed with SPAFF and BARS had higher correlations (i.e., convergent validity) than dissimilar behaviors (i.e., discriminant validity). Humor, affection, sadness (i.e., hurt), and anxiety demonstrated convergent validity across the two systems. In addition, the BARS ratings of aggression, frustration, defensiveness, and scorn were moderately to strongly correlated with the SPAFF codes of anger and contempt. Equally important were the negative and nonsignificant correlations between different specific affects across the two systems. For example, the lack of a significant correlation between humor assessed with BARS and anger assessed with SPAFF may be viewed as an indicator of discriminant validity. A notable exception was the significant correlation between humor assessed with BARS and anxiety assessed with SPAFF. In addition to the associations of specific affect codes and ratings, the association of two composite variables was assessed. The correlation of a BARS composite anger/contempt variable consisting of aggression, frustration, defensiveness, and scorn and a SPAFF composite anger/contempt variable consisting of anger and contempt codes was calculated. A strong association between the two composite anger/contempt variables was demonstrated for both husbands, $r = .73$, $p < .01$, and wives, $r = .66$, $p < .01$. These correlations and those described in Table 4 suggest that the BARS and the SPAFF systems have parallel categories that assess very similar and specific types of dyadic affect.

The correlations assessing discriminant validity were compared with the correlations assessing convergent validity to examine

Table 2
Means of Base Rates of Affect Coding in the Specific Affect Coding System and Correlations Between Husbands' and Wives' Topics

	Husbands' topics (base rate *M*)		Wives' topics (base rate *M*)		Husbands' topics (*SD*)		Wives' topics (*SD*)		*r* between topics	
Affect	HB	WB	HB	WB	HB	WB	HB	WB	HB	WB
Anger	4.4	8.2	5.0	8.3	9.6	12.8	10.4	14.4	.29**	.34**
Contempt	0.9	1.3	0.8	1.0	2.4	3.2	2.0	2.8	.46**	.09
Sadness	0.3	1.3	0.5	1.1	0.9	4.0	2.9	3.2	−.02	.25**
Anxiety	9.2	9.5	8.1	8.2	10.5	10.5	8.7	9.0	.49**	.36**
Humor	2.3	2.4	2.7	2.9	3.2	3.5	4.6	4.7	.43**	.35**
Affection	0.5	0.5	0.4	0.6	1.3	1.1	1.0	2.0	.16*	.05

Note. $N = 172$. The values in the table represent the sample mean of the number of 5-s units characterized by the specific affect on the left as demonstrated by the husband or wife. Behaviors not reliably coded are not included in this table. Correlations greater than .25 are considered significant using a Bonferroni correction ($\alpha_{ew} = .05$). HB = husbands' behavior; WB = wives' behavior.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Table 3
Means of Base Rates of Affect Rating in the Behavioral Affective Rating System and Correlations Between Husbands' and Wives' Topics

	Husbands' topics (base rate *M*)		Wives' topics (base rate *M*)		Husbands' topics (*SD*)		Wives' topics (*SD*)		*r* between topics	
Affect	HB	WB	HB	WB	HB	WB	HB	WB	HB	WB
Affection	2.9	3.3	2.7	3.2	3.7	3.9	3.6	3.9	.34**	.25**
Humor	6.4	6.7	6.1	7.6	6.3	6.3	5.9	7.2	.50**	.42**
Anxiety	13.2	13.2	12.1	12.4	8.9	9.6	9.0	8.9	.46**	.20*
Defensiveness	3.8	4.9	4.3	4.0	4.6	4.9	5.4	4.5	.05	.22*
Aggression	2.1	3.5	2.3	3.1	3.8	5.7	4.3	4.8	.29**	.49**
Scorn	3.1	3.3	3.6	3.1	6.9	5.3	6.1	5.6	.65**	.44**
Frustration	4.4	5.8	5.2	6.1	5.6	7.2	7.4	7.2	.17	.35**
Hurt	0.8	2.6	1.3	4.0	2.0	4.8	3.3	6.5	−.02	.38**

Note. For husbands' topics, $n = 156$; for wives' topics, $n = 143$. The values in the table represent the sample mean of the sum of the ratings for each specific affect in each interaction as demonstrated by each spouse. Correlations greater than .25 are considered significant using a Bonferroni correction ($\alpha_{ew} = .05$). HB = husbands' behavior; WB = wives' behavior.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

whether, as expected, discriminant associations were lower than convergent associations. The mean correlation coefficient of dissimilar SPAFF and BARS behaviors (e.g., humor assessed with BARS and anger assessed with SPAFF) was lower (husbands: $M = -.07$, $SD = .13$; wives: $M = -.06$, $SD = .12$) than both the mean correlation coefficient between SPAFF coding of and BARS rating of anger/contempt (husbands: $M = .48$, $SD = .13$; wives: $M = .44$, $SD = .14$), $z_{husbands} = 7.16$, $p < .01$, $z_{wives} = 6.48$, $p < .01$, and the mean correlation coefficient between SPAFF coding of and BARS rating of humor/affection (husbands: $M = .30$, $SD =$

Table 4
Correlations of BARS and SPAFF Behaviors for Both Husbands and Wives

	SPAFF behavior					
BARS behavior	Humor	Affection	Anger	Contempt	Sadness	Anxiety
Humor						
H	.57**	.11	−.09	−.12	.02	.42**
W	.60**	.06	−.09	−.05	−.17*	.32**
Affection						
H	.30**	.22**	−.24**	−.18*	.00	.13
W	.35**	.32**	−.26**	−.16*	−.04	.20**
Aggression						
H	−.19*	−.13	.66**	.46**	−.05	−.17*
W	−.17*	.10	.59**	.33**	−.02	−.15
Frustration						
H	−.22**	−.16*	.62**	.42**	−.05	−.11
W	−.19*	−.15	.55**	.41**	.02	−.20**
Defensiveness						
H	−.16*	−.09	.45**	.26**	.01	−.08
W	−.17*	−.06	.35**	.20**	−.04	−.14
Scorn						
H	−.07	−.12	.53**	.43**	−.05	−.12
W	−.09	−.09	.50**	.55**	.00	−.17*
Hurt						
H	−.13	−.10	.05	.09	.49**	−.10
W	−.19*	−.01	.04	.00	.56**	−.06
Anxiety						
H	.10	−.18*	−.22**	−.12	−.02	.38**
W	.11	−.07	−.23**	−.10	−.03	.33**

Note. Behaviors in the husband-selected topic and the wife-selected topic were collapsed for these analyses; hence, for all of the correlations in this table, $N = 172$. Correlations less than −.25 and correlations greater than .25 are considered significant using a Bonferroni correction ($\alpha_{ew} = .05$). BARS = Behavioral Affective Rating System; SPAFF = Specific Affect Coding System; H = husband affect; W = wife affect.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

.20; wives: $M = .33$, $SD = .22$), $z_{husbands} = 4.84$, $p < .01$, $z_{wives} = 5.15$, $p < .01$. These data suggest that the behaviors assessed with BARS are analogous to the behaviors assessed with SPAFF, thus supporting the possibility that the constructs underlying the BARS system are the same as those underlying the SPAFF system.

Intercorrelations of Affective Behaviors Within an Observational System

The intercorrelations of the behaviors assessed with SPAFF were examined to determine the associations between codes. However, when a SPAFF code is assigned to a behavior, it precludes the assignment of another code, thus dramatically lowering the power of the correlational method and increasing the likelihood of Type II error. For this reason, examining the correlations between SPAFF codes violates an assumption of correlation analyses. This is the same issue that precludes using factor analysis with SPAFF data. Nevertheless, understanding the intercorrelations of behaviors as an aid to the interpretation of SPAFF findings is important; therefore, the intercorrelations of SPAFF codes are presented in Table 5 with the aforementioned qualifier. The most notable of the overly conservative intercorrelations is the strong association of anger and contempt for both husbands and wives. Because of the Type II error inherent in the SPAFF intercorrelations, the intercorrelations of BARS were also examined. The effect sizes of the correlational findings involving BARS were expected to be larger than those involving SPAFF because units of behavior could be described with more than one specific affect and thus allowed to covary. The intercorrelations for BARS are displayed in Table 6. There were strong associations of frustration, defensiveness, verbal aggression, and scorn. These four emotions are negatively related to affection. In addition, humor and affection are correlated when assessed with both SPAFF and BARS.

Factor Analysis of Affect Ratings

Factor analysis was used to determine whether correlated behaviors assessed with BARS would load on the same dimension. On the basis of the intercorrelations of the behaviors, it was expected that four factors would emerge. The first factor was expected to comprise frustration, defensiveness, verbal aggression, and scorn. The remaining factors were expected to consist of humor and affection on one factor, hurt on another factor, and anxiety on the final factor. The first decision point with the analysis was the selection of a method of extraction: principal-components analysis (PCA) or factor analysis. The fact that the primary purpose of the analysis was an empirical summary of the data set suggested the use of PCA (Tabachnick & Fidell, 2001); however, factor analysis was chosen over PCA because it is considered to be a more accurate estimate of factor loadings when "only a small number of items load on each dimension and the items have relatively low communality" (Floyd & Widaman, 1995, p. 291). A minimum eigenvalue of 1.0 was the preliminary criterion for selecting the number of factors. This indicated the presence of four factors for husbands and three for wives (eigenvalues for husbands: 2.72, 1.15, 1.10, 1.01, 0.64, 0.59, .046, and 0.33; for wives: 2.72, 1.25, 1.06, 0.91, 0.63, 0.59, 0.44, and 0.39). A visual examination of the scree plots using the Cattell–Nelson–Gorsuch (Cattell, 1966; Gorsuch, 1983) scree test indicated that the last significant drop came after the fourth factor for both husbands and wives, suggesting the presence of four factors for wives and corroborating the presence of four factors for husbands. A review of the eigenvalues indicated the possibility of one factor; however, if there is doubt about the number of factors, it is suggested that one err "on the side of too many factors, provided the common factors do not degenerate" (Gorsuch, 1974, p. 160). Therefore, four factors were extracted using the principal factors technique. The squared multiple correlations of the factors ranged from .48 to .79 for husbands and from .31 to .79 for wives, which indicated moderate-to-high internal consistency of the factors. The communality values of the behaviors ranged from low to moderate, reflecting the heterogeneity of different behaviors assessed by the BARS (see "Communalities" or h^2 columns in Table 7). In data sets with a greater number of variables, those with low communality values would be excluded from the factor loading rotation; however, with so few variables, even those with low communality values were included. This led to the extraction of some factors with a single loaded variable.

Table 5
Intercorrelations of Reliable Codes in the Specific Affect Coding System

Affect	1	2	3	4	5	6
1. Humor	—	.22**	−.20**	−.17*	−.12	−.08
2. Affection	.26**	—	−.12	−.06	.04	−.09
3. Anger	−.11	−.19*	—	.53**	.12	−.17*
4. Contempt	−.14	−.14	.67**	—	.08	−.08
5. Sadness	−.09	−.10	.03	−.02	—	−.11
6. Anxiety	.02	−.16*	−.04	−.07	.03	—

Note. $N = 172$. Correlations for wives are above the diagonal; correlations for husbands are below the diagonal. For this analysis, behavioral codes were collapsed across the two discussions, one with a topic selected by husbands and the other with a topic selected by wives. Correlations less than −.25 and correlations greater than .25 are considered significant using a Bonferroni correction ($\alpha_{ew} = .05$).
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Table 6
Intercorrelations of Ratings in the Behavioral Affective Rating System

Affect	1	2	3	4	5	6	7	8
1. Humor	—	.35**	−.17*	−.13	−.11	−.04	−.13	.12
2. Affection	.35**	—	−.26**	−.22**	−.15*	−.16*	−.04	.10
3. Frustration	−.19*	−.27**	—	.43**	.56**	.57**	.16*	−.20**
4. Defensive	−.11	−.31**	.50**	—	.43**	.43**	.07	−.04
5. Aggression	−.25**	−.22**	.62**	.49**	—	.54**	.12	−.12
6. Scorn	−.12	−.23**	.38**	.29**	.44**	—	.00	−.18*
7. Hurt	−.04	.05	.19*	.15	.02	−.07	—	.04
8. Anxiety	.18*	−.05	−.11	−.06	−.23**	−.16*	.02	—

Note. $N = 172$. Correlations for wives are above the diagonal; correlations for husbands are below the diagonal. For this analysis, behavioral codes were collapsed across the two discussions, one with a topic selected by husbands and the other with a topic selected by wives. Correlations less than −.25 and correlations greater than .25 are considered significant using a Bonferroni correction ($\alpha_{ew} = .05$).
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Oblique rotation of the factors indicated that the strongest correlation among the factors for husbands was .31 and for wives was −.40. These correlations are sufficiently weak so as to suggest that orthogonal rotation is appropriate (Tabachnick & Fidell, 2001). The loadings of the behaviors on the factors, the communalities, and the variance explained by each factor are displayed in Table 7. The behaviors that loaded on the factors suggest that there were three negative affect factors and one positive affect factor. The four factors may be described as Anger/Contempt, Humor/Affection, Anxiety, and Hurt. The behavioral ratings paralleling the SPAFF codes of anger and contempt (i.e., defensiveness, frustration, verbal aggression, and scorn) all loaded on the same factor, labeled *Anger/Contempt.* The second factor consisted of the two positive behaviors that were rated: affection and humor. The third factor that was extracted was labeled *Anxiety,* but it is important to note that husbands' humor also loaded on this factor. This result possibly suggests that husbands use humor to mask their anxiety to a greater degree than wives. The final factor to be extracted for both husbands and wives was Hurt (i.e., Sadness in SPAFF).

Convergent and Discriminant Validity

To determine the convergent and discriminant validity of the BARS ratings, the associations between affect and other marital constructs were examined. The sums of the BARS ratings and the base rates of the SPAFF codes were used as predictor variables and correlated with the following criterion variables: belief that disagreement is destructive, marital satisfaction, and postinteraction appraisals. In both SPAFF and BARS, the positive (i.e., humor/affection) behaviors were summed together and the anger/con-

Table 7
Rotated Factor Loadings, Communalities (h^2), and Variance Explained by Each Factor of Affect Ratings in the Behavioral Affective Rating System at Time 1

	Husbands' behavior					Wives' behavior				
	Factor					Factor				
Behavior	1	2	3	4	h^2	1	2	3	4	h^2
Humor	−.05	.47	.58	−.10	.57	−.03	.57	.13	−.18	.37
Affection	−.27	.80	−.06	.08	.72	−.18	.60	.05	.03	.40
Frustration	.73	−.13	−.12	.23	.62	.71	−.18	−.23	.20	.63
Defensive	.62	−.19	.03	.17	.45	.58	−.20	.09	.03	.39
Aggression	.79	−.04	−.32	−.02	.72	.70	−.06	−.09	.15	.53
Scorn	.51	−.10	−.14	−.17	.32	.76	−.01	−.20	−.10	.63
Hurt	.06	.03	.02	.63	.40	.07	−.07	.05	.50	.40
Anxiety	−.15	−.09	.39	.05	.18	−.11	.10	.49	.05	.26
	Variance explained by each factor									
Sums of squared loadings	1.89	0.93	0.63	0.53		1.98	0.78	0.36	0.36	

Note. On the basis of the loadings, the interpretive labels for the factors are as follows: Factor 1 = Anger/Contempt, Factor 2 = Humor/Affection, Factor 3 = Anxiety, and Factor 4 = Hurt. Factor loadings were extracted with the principal factors technique and rotated using the varimax procedure.

Table 8
Correlation of Behaviors Displayed and Other Constructs in Marriage

Method of observation	Husbands' topics				Wives' topics			
	Relationship beliefs		Marital satisfaction		Relationship beliefs		Marital satisfaction	
	H	W	H	W	H	W	H	W
Anger/contempt								
SPAFF								
Husband	−.16*		−.01	$-.04_{a}$	.02		−.10	−.13
Wife		.11	−.11	$-.29_{b}$**		.13	−.14	$-.29_{a}$**
Anger								
Husband	−.17*		.01	$-.02_{a}$	.03		−.11	−.14
Wife		.11	−.13	$-.31_{b}$**		.12	−.14	$-.30_{a}$**
Contempt								
Husband	−.04		−.10	−.11	−.04		−.05	$-.05_{b}$
Wife		.06	.01	$-.07_{a}$		.12	−.07	−.13
BARS								
Husband	−.06		−.05	−.10	.09		−.19*	−.10
Wife		.05	−.05	−.23**		.07	−.07	−.21*
Humor/affection								
SPAFF								
Husband	$-.25_{c}$**		.13	.22**	−.16*		.16*	.26**
Wife		−.19*	.11	.16*		−.11	.16*	.24**
BARS								
Husband	$-.02_{d}$		.12	.09	−.09		.16	.21*
Wife		−.05	.18*	.13		−.16*	.18*	.20*
Sadness								
SPAFF								
Husband	.07		−.17*	$-.22_{e}$**	.06		$.06_{e,f}$	.01
Wife		.06	−.04	$.03_{f}$		$.22_{e}$**	$-.14_{f,g}$	−.18*
BARS								
Husband	.01		−.03	−.09	.01		$.12_{e}$	−.02
Wife		.06	−.14	−.15		$-.04_{f}$	$-.17_{g}$*	−.19*
Anxiety								
SPAFF								
Husband	.15		−.04	−.05	.10		−.14	−.13
Wife		.07	.01	.05		−.11	.01	.08
BARS								
Husband	−.01		.02	.02	.12		−.10	−.05
Wife		−.01	.04	.05		−.06	−.05	.11

Note. $n = 172$ for analyses involving the Specific Affect Coding System (SPAFF); $n = 156$ for analyses involving the Behavioral Affective Rating System (BARS) in husbands' topics; $n = 143$ for analyses involving BARS in wives' topics. Different subscripts within a behavior (e.g., anger/contempt) and within a column indicate the correlations are significantly different, $\alpha = .05$ ($z > 1.96$). The lack of a subscript indicates none of the correlation coefficients were significantly different within that group. Correlations less than −.25 and correlations greater than .25 are considered significant using a Bonferroni correction ($\alpha_{ew} = .05$). H = husband; W = wife.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

tempt behaviors (i.e., for SPAFF, anger and contempt; for BARS, defensiveness, aggression, scorn, and frustration) were summed together. However, anger and contempt, as assessed with SPAFF, were examined separately to determine whether they were uniquely associated with other variables, despite being correlated with each other. Sadness and anxiety did not require the summation of individual codes or ratings because only one observed affect loaded on each factor.[6]

Belief that disagreement is destructive. The correlations between affect and the belief that disagreement is destructive are in Columns 2, 3, 6, and 7 of Table 8. Humor/affection was negatively

[6] To maintain consistency across spouses, husbands' humor was not combined with husbands' anxiety, despite the fact that they loaded on the same factor.

associated with this belief for both husbands and wives in husbands' topics when assessed using SPAFF. The results for humor/affection were inconsistent in wives' topics and when assessed with BARS. In addition, wife sadness, assessed with SPAFF in wife-selected topics, was associated with higher levels of this belief. Lastly, there was an unexpected finding indicating that husbands' anger in the husbands' topics was associated with lower levels of the dysfunctional relationship beliefs. This unexpected finding was weak and was not replicated with wives' behavior or in wives' topics. Overall, it appears that affect and the belief that disagreement is destructive are modestly associated, but only when SPAFF is used to assess affect.

Marital satisfaction. The correlations between affect and marital satisfaction are in Columns 4, 5, 8, and 9 of Table 8. As expected, there was an association between affective behavior and marital satisfaction. Replicating past findings suggesting an association between negativity and distress (see Weiss & Heyman, 1990), the results suggested that wives' anger/contempt is negatively correlated with wives' marital satisfaction. This finding was present in both systems of observational assessment and in both types of interactions. For husbands, the association between anger/contempt and marital satisfaction was only significant when assessed with BARS and when discussing the wives' topics. When SPAFF assessments of anger and contempt were analyzed separately, the correlations between wives' anger and wives' marital satisfaction were statistically identical to the correlations between the anger/contempt composite variable and marital satisfaction. However, the correlations were not significant for any other associations of anger and marital satisfaction. In addition, none of the associations between contempt and marital satisfaction were significant. In summary, it appears that for women, BARS anger/contempt, SPAFF anger/contempt, and SPAFF anger all predict marital satisfaction in a similar manner (the largest difference in corresponding *r*s was $z = 1.11$, *ns*), whereas for husbands, none of the associations between anger/contempt and marital satisfaction were significant when controlling for experimentwise Type I error.

The association between humor/affection and marital satisfaction also replicated past findings (see Weiss & Heyman, 1990), with small-to-medium correlation coefficients between the two constructs. Husbands' sadness displayed in topics the husband chose was modestly associated with lower marital satisfaction. A similar finding was true for wives' sadness in wives' topics, but it was statistically significant ($\alpha = .05$) only for wives' marital satisfaction. There were no significant associations of anxiety and marital satisfaction.

Postinteraction appraisals. To examine the interactions from the participants' perspective, each spouse appraised their problem-solving interactions. The correlations of affect displayed in the interaction and appraisals of the interaction are presented in Table 9. Generally, for both SPAFF and BARS, husbands' and wives' negative appraisals were associated with anger/contempt in husbands' and wives' topics. Of the replicated correlations, a few did not meet significance, but in no case did a nonsignificant finding replicate across methods of assessing appraisals. However, husbands' anger/contempt displayed in wives' topics was not significantly associated with negative affective (i.e., assessed with the PANAS) appraisals made by the wife. This nonsignificant finding replicated across both methods of observing affect. Although beyond this exception, it appeared that anger/contempt is consistently associated with negative appraisals of interactions across topics, spouses, methods of observation, and type of appraisal. In addition, there were no significant differences between the correlations measuring the association between SPAFF and appraisals and the correlations measuring the association between BARS and appraisals. When the composite SPAFF code of anger/contempt was separated into its components and correlated with appraisals, there were more significant correlations involving anger than contempt. In husband-selected topics, wives' anger, as assessed with SPAFF, was more strongly correlated with husbands' and wives' global appraisals than was wives' contempt: for husbands' appraisals, $z = -3.38$, $p < .01$, for wives' appraisals, $z = -2.86$, $p < .01$.

The composite code of humor/affection was correlated in the expected direction with spouses' appraisals of their interactions. As with previous associations, this correlation was replicated 32 times with affect assessed with either SPAFF or BARS, topics selected by either the husband or wife, and appraisals rated by either the husband or wife. Some of the correlations did not reach significance; however, nonsignificance did not replicate across methods of assessing affect or methods of assessing appraisals.

In contrast to the results for anger/contempt and humor/affection, anxiety and sadness did not appear to be associated with appraisals of interactions. There were a few correlations that met the criteria for significance, but generally these did not replicate across conditions. The exception was that husbands' displays of sadness, assessed with both SPAFF and BARS, were associated with negative appraisals of affect by the wife in topics selected by the wife.

Incremental Association of Criterion Variables

To determine if assessing behaviors with SPAFF or BARS led to different effect sizes in the association between affective behavior and the criterion variables (i.e., relationship beliefs, marital satisfaction, and appraisals), the correlation coefficients of each method of behavioral observation were compared. The subscripts in Tables 8 and 9 indicate correlation coefficients that are significantly different ($\alpha = .05$). There was only one difference between SPAFF and BARS in coefficients measuring the same associations. Husbands' displays of humor/affection in husbands' topics was associated more strongly with dysfunctional relationship beliefs when the affect was observed with SPAFF than when the affect was observed with BARS, $z = -2.11$, $p < .05$ (see Table 8). Otherwise, there were no significant differences between the associations of the two systems.

Discussion

The purpose of this study was to assess the validity of the BARS system as an alternative to the SPAFF system and to examine the psychometric properties of both systems. To that end, a series of statistical analyses were performed. Correlations between SPAFF coding and BARS ratings of the same behaviors were positive and significant, indicating convergent validity. The correlations of different behaviors were generally not significant, indicating discriminant validity. These results replicated across spouse and topic. These intersystem correlations are evidence that BARS is a valid alternate to the SPAFF system.

Table 9
Correlation of Behaviors Displayed in Interactions and Appraisals of the Same Interactions

Method of observation	Husbands' topics				Wives' topics			
	PANAS		Global		PANAS		Global	
	H	W	H	W	H	W	H	W
Anger/contempt								
SPAFF								
Husband	−.19*	−.19*	−.22**	−.31**	−.25**	−.13	−.21**	−.31**
Wife	−.15	−.20**	$-.28_a$**	−.39**	−.29**	−.19*	−.31**	−.37**
Anger								
Husband	−.15*	−.20*	−.20**	−.29**	−.26**	−.13	−.21**	−.31**
Wife	−.15*	−.20**	$-.30_a$**	−.39**	−.28**	−.17*	−.30**	−.36**
Contempt								
Husband	−.24**	−.05	−.17*	−.20**	$-.13_b$	−.09	−.13	−.23**
Wife	−.04	−.08	$-.05_b$	−.19*	−.23**	−.20**	−.24**	−.27**
BARS								
Husband	−.21**	−.25**	$-.33_a$**	−.34**	−.27**	−.12	−.32**	−.28**
Wife	−.18*	−.15	$-.31_a$**	−.33**	$-.36_a$**	−.18*	−.30**	−.31**
Humor/affection								
SPAFF								
Husband	.23**	.14	.27**	.30**	.18*	.17*	.28**	.33**
Wife	.23**	.24**	.33**	.36**	.21**	.28**	.33**	.40**
BARS								
Husband	.10**	.22**	.21**	.19*	.18*	.19*	.28**	.31**
Wife	.16	.21*	.19*	.19*	.12	.14	.19*	.27**
Sadness								
SPAFF								
Husband	−.05	−.07	−.01	−.06	−.11	−.21**	−.08	−.16
Wife	.03	−.18*	−.01	−.04	.05	−.05	−.01	−.07
BARS								
Husband	.01	.09	.04	.01	−.11	−.25**	.02	−.16
Wife	.15	.03	.04	−.05	.00	−.10	.06	.03
Anxiety								
SPAFF								
Husband	−.08	.01	.03	−.12	−.06	−.04	−.14	−.16*
Wife	−.07	−.03	.01	−.04	−.07	−.09	−.18	−.05
BARS								
Husband	−.04	−.10	−.12	−.05	−.17*	−.01	−.05	−.12
Wife	−.06	−.10	−.10	−.03	.04	−.04	−.02	.01

Note. $n = 172$ for analyses involving the Specific Affect Coding System (SPAFF); $n = 156$ for analyses involving the Behavioral Affective Rating System (BARS) in husbands' topics; $n = 143$ for analyses involving BARS in wives' topics. Different subscripts within a behavior (e.g., anger/contempt) and within a column indicate the correlations are significantly different, $\alpha = .05$ ($z > 1.96$). The lack of a subscript indicates none of the correlation coefficients were significantly different within that group. Correlations less than −.25 and correlations greater than .25 are considered significant using a Bonferroni correction ($\alpha_{ew} = .05$). PANAS = sum of the Positive and Negative Affect Scale, used to describe affect during the just-completed interaction; Global = sum of the ratings of general questions about the just-completed interaction; H = husband; W = wife.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Assessing the psychometric properties of the two systems was an iterative process that began by examining the intercorrelations of the specific affects assessed within an observational system. The strength of the intercorrelations of the SPAFF codes was artificially suppressed because of the nature of the system that allows only one code to be assigned to a 5-s block of time. Therefore, the intercorrelations of the BARS system were also examined. For both systems, specific affects of anger and contempt were positively associated, as were the affects of humor and affection. These correlations replicated across spouse and topic. To further examine whether these associations may be indicative of latent constructs, a factor analysis of the BARS ratings was conducted. The system of assigning codes in the SPAFF system violated assumptions of factor analysis; thus SPAFF was not analyzed with factor analytic methods. The factor analysis extracted four factors for husbands and wives that may be described as anger/contempt, humor/affec-

tion, sadness (i.e., hurt), and anxiety. The results from the intercorrelations of the specific affects and the factor analysis were consistent.

Having examined the psychometric properties of the two observational systems, attention was turned to the convergent and discriminant validity of the two systems. Anger/contempt was associated with lower marital satisfaction. Humor/affection was associated with higher marital satisfaction. These results generally replicated across observational systems, gender of the spouse exhibiting the behavior, and gender of the spouse who picked the topic. The findings involving sadness and anxiety did not replicate across systems, spouses, or topics in a consistent way. In addition to examining the associations between affect and marital satisfaction, the belief that disagreement is destructive and appraisals of the interactions were analyzed for their association with affect. There was modest and inconsistent support for the collinearity of the belief that disagreement is destructive and affect. The weak association of these two variables may indicate either that beliefs and behavioral mechanisms operating in close relationships are largely orthogonal or that one moderates the other in relation to marital satisfaction, a fact that has received some empirical support with attributions (M. D. Johnson et al., 2001). The other cognitive variable that was analyzed in relation to affective behavior was the spouse's appraisal of the interaction. Appraisals are fundamentally different from relationship beliefs because they are related to the specific task that was behaviorally coded. Therefore, it was expected that affect in the interaction would be related to the appraisals of the interaction to a greater degree. The affective factors of anger/contempt and humor/affection consistently have significant correlations with appraisals, and these are the same factors that had the significant associations with marital satisfaction. Thus, displays of anger/contempt and humor/affection appear to have a robust effect on marital satisfaction and spouses' appraisals of the interaction. However, the relationship between anxiety and sadness appears to be orthogonal to appraisals of the interaction. The lack of association of appraisals with anxiety and sadness may be indicative of the complex nature of anxiety and sadness. Anxiety may represent anxiety about the topic or the process of discussing the topic, but it may also represent situational anxiety related to the laboratory setting. Similarly, sadness may represent sadness that is shared with the spouse (e.g., sadness about the death of a loved one) or may represent sadness that is caused by the spouse (e.g., sadness about a spouse's criticism). Therefore, the lack of association between these two affective factors and appraisals of the interaction should not be interpreted as a preclusion of their importance. The interaction of observed sadness or anxiety and appraisals might predict marital satisfaction. In addition, the effects of sadness and anxiety may be more predictive in samples with a spouse who is depressed or anxious. Therefore, it is important to examine these relationships in a clinical sample before reaching conclusions involving sadness and anxiety. Further exploration of the relation of affect, appraisals, and distress is needed before this issue can be resolved; however, it is beyond the scope of this article.

Relation of Data to Past Findings

These findings both support and contradict previous SPAFF findings. The emergence of three separate negative affects suggests that Gottman and Krokoff (1989) were right to expand the assessment of affect in dyadic interactions by not lumping all negative expressions of affect together. Prior to the development of the SPAFF coding system, affect was generally considered to be positive, neutral, or negative. The data in this study suggest that a trichotomous categorization of affect may lessen the power of affect to predict marital outcome. The tendency of researchers to consider affect as either good, bad, or neutral was reflected in the early versions of behavioral marital interventions (e.g., Jacobson & Margolin, 1979); however, more recent advances in psychotherapy for couples acknowledge the complex role of emotions (e.g., Jacobson & Christensen, 1996; S. M. Johnson & Greenberg, 1995). These advances in treatment have developed without the aid of a firm understanding of which dyadic emotional processes contribute to the decline of marital satisfaction and which processes maintain satisfaction. Therefore, the results of this study may be considered a call for improved theories of the role of affect in marriage and better testing of current theories.

Although the data presented in this study suggest that affect displayed in interactions is more complicated than positive, negative, and neutral, these findings also suggest that, at present, there might be limits to the level of complexity in affect that can be assessed with current observational methods. Specifically, it appears that distinguishing the constructs of anger and contempt is an especially difficult task. On the basis of anecdotal and clinical data, it seems evident that anger and contempt are distinct, and conceptually it seems likely that they may have a differential impact on marital quality. However, the data presented here lead to questions about the ability of SPAFF and BARS to distinguish the two types of affect in a valid and reliable manner. The argument made by Gottman et al. (1998) that their data demonstrated that anger and contempt in newlywed marriages operate as distinct mechanisms and that they are associated with different outcomes is not supported by the psychometric properties of the SPAFF system found in the present study. Their conclusions are conceptually appealing, but without supporting evidence regarding the psychometric properties of the SPAFF system, it is unclear that anger and contempt may be differentiated with observational assessment. The data in this study suggest that for newlywed couples, anger and contempt both load on the same factor and are both associated with concurrent marital discord.

There are at least two different explanations for the discrepancy in the findings of this study and the Gottman et al. (1998; see also Gottman & Krokoff, 1989) findings. It may be that observers were not reliable across the laboratories and that the Gottman laboratory was able to differentiate the constructs of anger and contempt. Unfortunately, without psychometric data on the validity of the individual codes in the SPAFF system from the Gottman laboratory, it is difficult to determine if this is the case. It may also be that methodological flaws in the Gottman et al. (1998) study, such as the use of extreme groups derived from their sample, resulted in accentuating differences between anger and contempt that would otherwise not be significant (Stanley et al., 2000; cf. Gottman et al., 2000). Nevertheless, the data presented here do not suggest that there are no differences in anger and contempt, but rather that, as Heyman (2001) observed, "the problems of [behavioral coding] have convinced us that we know far more than we do, that our theories have received more support than they have, and that our methodology is more robust than it is" (p. 28).

Limitations

There are several aspects of this study that limit interpretation of the results. First, the data reported in this study are correlational and subject to the limitations inherent in correlational designs. Second, the data are limited by the level of interrater reliability for both the SPAFF coding and the BARS rating. Third, it should be noted that the SPAFF coders on this project attempted to closely adhere to the original rules governing the SPAFF coding system, yet there are bound to be differences in how two different laboratories code data. No interlab reliability estimates were obtained; therefore, it is possible that coders in another lab would have different findings based on the same behavioral codes. Fourth, the sample used in this study consisted of newlyweds, which is similar to the sample used by Gottman et al. (1998) but limits the ability of these findings to generalize to more established marriages. Fifth, all of the analyses, including those assessing convergent and discriminant validity, were cross-sectional, thus limiting the assessment of truly "predictive" validity. Sixth, no data were collected regarding the time and expense of the BARS versus the SPAFF observational system; therefore, no conclusions can be drawn about the relative cost of each system. Seventh, two of the factors derived in the factor analysis had only a single significant loading on the factor. Therefore, those factors only assessed the variance associated with those specific variables, which may render them unstable and difficult to replicate (Floyd & Widaman, 1995; Gorsuch, 1974). Eighth, it may be possible that only one or two factors should have been extracted from the data. However, this would be consistent with neither the use of SPAFF in the marital literature nor the Kaiser–Guttman criterion and Cattell–Nelson–Gorsuch scree test that indicated the presence of four factors. In addition, the greater number of factors extracted allows more of an opportunity for anger and contempt to load on separate factors. The use of more rather than fewer factors is also consistent with data on the structure of emotional expressivity (Trierweiler, Eid, & Lischetzke, 2002). Nevertheless, the characteristics of this sample are not ideal for factor analysis; thus the factors of sadness and anxiety should be interpreted tentatively until these findings are replicated. Ninth, it is possible that observers may have an implicit expectation that two affects (e.g., anger and contempt) will be associated, leading them to mark one behavior near their assessment of the other behavior. This would artificially increase the likelihood of finding covariation between anger and contempt. Lastly, the sample required the participants to volunteer for the study after receiving an invitation. Thus, to the extent that the sample is self-selected, there are limitations of the generalizability of the results. For example, depressed or anxious spouses may not have volunteered for the study, possibly suppressing the amount of sadness and anxiety that would be observed in marital interactions in the general population.

Strengths

Despite its correlational design, this study used several methodological strategies that allow relatively strong conclusions to be reached. The strengths of the study are in line with the eight recommendations for observational research of marital interactions suggested by Heyman (2001). First, the hypotheses and theoretical framework followed an established line of research. Second, construct validity was a primary concern of the study, which led to the following four design features: (a) The codes that were used in previous SPAFF studies were used in this study, enabling the agglomeration of validity data across studies; (b) the BARS rating system was developed and rated independently of the SPAFF coding team (i.e., nobody was a member of both teams), with both teams blind to the hypotheses being tested; (c) by using a rating system in the factor analysis instead of a coding system, the violation of a key assumption of factor analysis was avoided; and (d) the associations between affect and other relationship traits (e.g., marital satisfaction) were examined to establish convergent and discriminant validity. Third, the reliability of the coding was done at the level of analysis. Fourth, a multitrait–multimethod (MTMM; Campbell & Fiske, 1959) approach was incorporated into the design by having a behavioral data set observed with two separate groups of observers using two separate systems (i.e., coding and rating). This design is particularly helpful in understanding the validity of the observed behavior and is rare because of the high cost. The examination of several specific types of affect using both ratings and coding and the comparison of the two systems resulted in a more precise estimation of the validity of the assessment. The MTMM design was also used in the assessment of the criterion variables (e.g., marital satisfaction), again allowing greater insight into the validity of the assessment. Fifth, the internal and content validity of the study was enhanced by experimentally controlling the topics chosen for discussion and by having one topic selected by the wife and one selected by the husband. Sixth, extreme groups were not used, but rather all of the couples with complete data were included in the sample. Seventh, analyses were replicated across topics, coding systems, and spouses. Thus, the different results for each combination may be compared. Eighth, asking spouses to report their appraisals of their just-completed interaction assessed external validity. These eight points represent an attempt to address Heyman's (2001) recent "recommendations for reducing measurement and inferential errors" (p. 26).

Conclusion

In this study, I introduced the BARS as a rating system for the behavioral assessment of specific affect that could be substituted for the SPAFF system and examined the psychometric properties of the SPAFF system. Regarding the first goal of the study, the BARS system demonstrated that it is a valid alternative to the SPAFF system. This was demonstrated through assessments of convergent and discriminant validity. Regarding the second goal of the study, the data suggest that specific affect in marital interactions may be combined into four categories: anger/contempt, humor/affection, sadness, and anxiety. Of these four categories, anger/contempt and humor/affection appear to be the two with direct associations with marital satisfaction. Anger/contempt is negatively correlated with marital satisfaction and humor/affection is positively correlated with satisfaction. This finding was true when affect was assessed with both SPAFF and BARS. In conclusion, combining behavioral observations of affect into these four combinations appears to be a valid strategy for data aggregation when using either a rating or a coding system for observed dyadic affect in newlyweds.

References

Baucom, D. H., & Epstein, N. (1990). *Cognitive–behavioral marital therapy.* New York: Brunner/Mazel.

Biglan, A., Hops, H., Sherman, L., Friedman, L. S., Arthur, J., & Osteen, V. (1985). Problem solving interactions of depressed women and their spouses. *Behavior Therapy, 16,* 431–451.

Bradbury, T. N. (1994). Unintended effects of marital research on marital relationships. *Journal of Family Psychology, 8,* 187–201.

Bradbury, T. N., Cohan, C. L., & Karney, B. R. (1998). Optimizing longitudinal research for understanding and preventing marital dysfunction. In T. N. Bradbury (Ed.), *The developmental course of marital dysfunction* (pp. 279–311). New York: Cambridge University Press.

Bradbury, T. N., & Fincham, F. D. (1993). Assessing dysfunctional cognition in marriage: A reconsideration of the Relationship Belief Inventory. *Psychological Assessment, 5,* 92–101.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245–276.

Christensen, A., & Heavey, C. L. (1990). Gender and social structure in the demand/withdraw pattern of marital conflict. *Journal of Personality and Social Psychology, 59,* 73–82.

Christensen, A., & Nies, D. C. (1980). The spouse observation checklist: Empirical analysis and critique. *American Journal of Family Therapy, 8,* 69–79.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70,* 213–220.

Eidelson, R. J., & Epstein, N. (1982). Cognition and relationship maladjustment: Development of a measure of dysfunctional relationship beliefs. *Journal of Consulting and Clinical Psychology, 50,* 715–720.

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

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7,* 286–299.

Foster, D. A., Caplan, R. D., & Howe, G. W. (1997). Representativeness of observed couple interaction: Couples can tell, and it does make a difference. *Psychological Assessment, 9,* 285–294.

Fredman, N., & Sherman, R. (1987). *Handbook of measurements for marriage and family therapy.* New York: Brunner/Mazel.

Geiss, S. K., & O'Leary, K. D. (1981). Therapist's ratings of frequency and severity of marital problems: Implications for research. *Journal of Marital and Family Therapy, 7,* 515–520.

Gorsuch, R. L. (1974). *Factor analysis.* Philadelphia: Saunders.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gottman, J. M. (1979). *Marital interactions: Experimental investigations.* New York: Academic Press.

Gottman, J. M. (1994). *What predicts divorce: The relationship between marital processes and marital outcomes.* Hillsdale, NJ: Erlbaum.

Gottman, J. M., Coan, J., Carrère, S., & Swanson, C. (1998). Predicting marital happiness and stability from newlywed interactions. *Journal of Marriage and the Family, 60,* 5–22.

Gottman, J. M., Coan, J., Carrère, S., & Swanson, C. (2000). Reply to "From Basic Research to Interventions." *Journal of Marriage and the Family, 62,* 265–273.

Gottman, J. M., & Krokoff, L. J. (1989). Marital interaction and satisfaction: A longitudinal view. *Journal of Consulting and Clinical Psychology, 57,* 47–52.

Gottman, J. M., & Krokoff, L. J. (1990). Complex statistics are not always clearer than simple statistics: A reply to Woody and Costanzo. *Journal of Consulting and Clinical Psychology, 58,* 502–505.

Hahlweg, K., Reisner, L., Kohli, G., Vollmer, M., Schindler, L., & Revenstorf, D. (1984). Development and validity of a new system to analyze interpersonal communication: Kategoriensystem für Partnerschaftliche Interaktion. In K. Hahlweg & N. S. Jacobson (Eds.), *Marital interaction: Analysis and modification* (pp. 182–198). New York: Guilford Press.

Heyman, R. E. (2001). Observation of couple conflicts: Clinical assessment applications, stubborn truths, and shaky foundations. *Psychological Assessment, 13,* 5–35.

Hops, H., Davis, B., & Longoria, N. (1995). Methodological issues in direct observation: Illustrations with the Living in Familial Environments (LIFE) coding system. *Journal of Clinical Child Psychology, 24,* 193–203.

Huston, T. L., McHale, S. M., & Crouter, A. C. (1986). When the honeymoon's over: Changes in the marriage relationship over the first year. In R. Gilmour & S. Duck (Eds.), *The emerging field of personal relationships* (pp. 109–132). Hillsdale, NJ: Erlbaum.

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

Jacobson, N. S., & Christensen, A. (1996). *Integrative couple therapy: Promoting acceptance and change.* New York: Norton.

Jacobson, N. S., & Margolin, G. (1979). *Marital therapy: Strategies based on social learning and behavior exchange principles.* New York: Brunner/Mazel.

Jacobson, N. S., & Moore, D. (1981). Spouses as observers of the events in their relationship. *Journal of Consulting and Clinical Psychology, 49,* 269–277.

Johnson, M. D., & Bradbury, T. N. (1999). Marital satisfaction and topographical assessment of marital interaction: A longitudinal analysis of newlywed couples. *Personal Relationships, 6,* 19–40.

Johnson, M. D., Johns, A., Kitahara, J., Ono, M., & Bradbury, T. N. (1998). *Behavioral Affective Rating System* (*BARS*) *manual.* Los Angeles: University of California, Los Angeles.

Johnson, M. D., Rogge, R., Karney, B. R., & Bradbury, T. N. (2001). The role of attributions in the longitudinal association between marital behavior and marital quality. In V. Manusov & J. H. Harvey (Eds.), *Attribution, communication behavior, and close relationships* (pp. 173–192). New York: Cambridge University Press.

Johnson, S. M., & Greenberg, L. S. (1995). The emotionally focused approach to problems in adult attachment. In N. S. Jacobson & A. S. Gurman (Eds.), *Clinical handbook of couple therapy* (pp. 121–141). New York: Guilford Press.

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.

Karney, B. R., Bradbury, T. N., Fincham, F. D., & Sullivan, K. T. (1994). The role of negative affectivity in the association between attributions and marital satisfaction. *Journal of Personality and Social Psychology, 66,* 413–424.

Karney, B. R., Davila, J., Cohan, C. L., Sullivan, K. T., Johnson, M. D., & Bradbury, T. N. (1995). An empirical investigation of sampling strategies in marital research. *Journal of Marriage and the Family, 57,* 909–920.

Katz, L. F., & Gottman, J. M. (1993). Patterns of marital conflict predict children's internalizing and externalizing behaviors. *Developmental Psychology, 29,* 940–950.

Kurdek, L. A. (1991). Predictors of increases in marital distress in newlywed couples: A 3-year prospective longitudinal study. *Developmental Psychology, 27,* 627–636.

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

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30–46.

Nelson, G. M., & Beach, S. R. H. (1990). Sequential interaction in depression: Effects of depressive behavior on spousal aggression. *Behavior Therapy, 21,* 167–182.

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

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana: University of Illinois Press.

Schmaling, K. B., & Jacobson, N. S. (1990). Marital interaction and depression. *Journal of Abnormal Psychology, 99,* 229–236.

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.

Stanley, S. M., Bradbury, T. N., & Markman, H. J. (2000). Structural flaws in the bridge from basic research on marriage to interventions for couples. *Journal of Marriage and the Family, 62,* 256–264.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.

Trierweiler, L. I., Eid, M., & Lischetzke, T. (2002). The structure of emotional expressivity: Each emotion counts. *Journal of Personality and Social Psychology, 82,* 1023–1040.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070.

Weiss, R. L., & Heyman, R. E. (1990). Observation of marital interaction. In F. D. Fincham & T. N. Bradbury (Eds.), *The psychology of marriage: Basic issues and applications* (pp. 87–117). New York: Guilford Press.

Woody, E. Z., & Costanzo, P. R. (1990). Does marital agony precede marital ecstasy? A comment on Gottman and Krokoff's "Marital Interaction and Satisfaction: A Longitudinal View." *Journal of Consulting and Clinical Psychology, 58,* 499–501.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology, 43,* 111–122.

Appendix

Behavioral Affective Rating Scale: Definitions and Examples

The Behavioral Affective Rating Scale (BARS) allows one to rate the affect in couples' interactions on a scale from 0 to 4 solely on the basis of the couples' body language, facial expressions, and tone of voice. The actual content of couples' interactions is not taken into consideration at all. A 0 is the absence of the affect, a 1 is mild, a 2 is medium, a 3 is strong, and a 4 is extreme. The following list includes examples for each of the ratings for all the affects.

It should be noted that during some periods of the interactions, none of the affects will be displayed. It is expected that the absence of these affects will be the rating most often used. The majority of the couples' affect will fall in the range of 0 to 2.

It is also important to recognize that some of the behavioral affects need to occur only briefly during the 30-s interval to receive high ratings. This is because some behavioral affects are primarily mercurial in nature. An asterisk (*) identifies these affects. The remaining affects need to occur in longer duration to receive higher values.

*Affection: genuine care, support, warmth, and tenderness. Scores: 0 = absence; 1 = genuine smiles; 2 = warm laughter; 3 = flirting, little love taps; 4 = holding hands, hugging, kissing.

*Humor: genuine, honest smile or laughter in a positive and agreeable situation, with no ill intention shared by the couple. Scores: 0 = absence; 1 = laughing smile; 2 = genuine laughter; 3 = goofiness; 4 = uncontrollable laughter.

Anxiety: nervousness, tenseness, and discomfort. Scores: 0 = absence; 1 = anxious tone of voice, shifting; 2 = nervous giggle, extended fidgeting; 3 = stuttering; 4 = sweating, panicky, skittish.

Engaging: showing positive involvement and focusing on the conversation. Scores: 0 = absence; 1 = steady, active eye contact, nodding; 2 = steady, active eye contact, nodding, affirmative vocal cues; 3 = steady, active eye contact, leaning, verbal cues, nodding; 4 = steady, active eye contact, body contact, leaning, verbal cues.

Disengaging: displaying a total disinterest in the conversation and not listening. Scores: 0 = absence; 1 = extended break of eye contact; 2 = over-talk; 3 = closed body position, no eye contact; 4 = totally unresponsive.

Defensive: self-justification. Scores: 0 = absence; 1 = shaking head, inward, defensive hand motions; 2 = more adamant head shaking and inward hand motions; 3 = aroused body posture, interrupting in spurts; 4 = very animated, prolonged defensive motions.

Aggressive: attacking, accusing, forcefully communicating. Scores: 0 = absence; 1 = forceful tone of voice, pointing; 2 = more aggressive tone of voice, outward hand motions; 3 = prolonged forcefulness in the tone of voice and body movements; 4 = in face, yelling.

Scorn: insulting, condescending, contemptuous, and sarcastic. Scores: 0 = absence; 1 = rolling eyes, light sarcastic tone of voice; 2 = contemptuous voice, more sarcasm; 3 = very condescending voice, withering looks; 4 = dismissive body posture, extremely sarcastic.

Frustration: flustered, upset, loss of patience and tense. Scores: 0 = absence; 1 = sighing, tense body posture; 2 = more sighing, holding head at an angle; 3 = clenching teeth, slight stuttering; 4 = so flustered unable to talk, red in face.

*Hurt: genuine emotional pain, sadness, and wounded. Scores: 0 = absence; 1 = hurt look, passively looking down; 2 = more expressions of sadness; 3 = shaky voice, watery eyes; 4 = crying.

Received August 21, 2000
Revision received June 18, 2002
Accepted June 24, 2002 ■