Published in final edited form as:
Psychol Assess. 2001 March ; 13(1): 5–35.

Observation of Couple Conflicts: Clinical Assessment Applications, Stubborn Truths, and Shaky Foundations

Richard E. Heyman
State University of New York at Stony Brook

Abstract

The purpose of this review is to provide a balanced examination of the published research involving the observation of couples, with special attention toward the use of observation for clinical assessment. All published articles that (a) used an observational coding system and (b) relate to the validity of the coding system are summarized in a table. The psychometric properties of observational systems and the use of observation in clinical practice are discussed. Although advances have been made in understanding couple conflict through the use of observation, the review concludes with an appeal to the field to develop constructs in a psychometrically and theoretically sound manner.

For couples observation research, it was the best of times, it was the worst of times, it was the age of gained wisdom, it was the age of media foolishness, it was results deserving of belief, it was results deserving of incredulity, it was the Light of theory, it was the Darkness of data mining, it was the spring of hope for science, it was the winter of despair for science, we had everything before us, we had nothing before us, we were all going direct to Stockholm for the Nobel, we were all going direct the other way—in short, the period was so far like the present period that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only.[1]

Considering the paradoxes inherent in following the developments of couples observation research, making sense of nearly 200 studies and what they purportedly show is an incredibly daunting task. Charles Dickens wrote [most] of the first paragraph in 1859, over 100 years before the first couples observation study, and thus it speaks more to humans' perceptions of their endeavors in general than to this field's paradoxes in particular. In our period of family stress and divorce, there is intense media and public interest in understanding, treating, and preventing marital discord. Unfortunately, in our period of soundbites and factoids, the ability of the "noisiest authorities [to] insist on [their work] being received, for good or for evil, in the superlative degree of comparison only" (Dickens, 1859, p. 1) only serves to convince both the public and many professionals 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. One cannot read the nearly 200 studies and not be impressed at the advances of the field, nor can one turn on the television and not be dismayed at the distorted presentations of some of the very same studies.

The purpose of this review is to provide a balanced examination of the published research involving the observation of couples, with special attention toward the use of observation for clinical assessment.[2] By reviewing research that has high research and clinical utility, this

---

Correspondence concerning this article should be addressed to Richard E. Heyman, Department of Psychology, State University of New York at Stony Brook, Stony Brook, New York 11794-2500. Electronic mail may be sent torichard.heyman@sunysb.edu..

Preparation of this article was supported by National Institute of Mental Health Grant R01MH57779.

[1]Adapted from Charles Dickens (1859), *A Tale of Two Cities.*

[2]I should note from the outset that the criticisms that I make of "the field" often apply to my own work.

special issue in general and this article in particular, hopes to narrow the gap between researchers who study couples' communication and clinicians who treat distressed couples.

Observable Communication: The Foundation for Case Conceptualization, Assessment, and Treatment

All couples intervention involves, at the very least, an implicit model of what is important in relationships. Every clinician, with every client, must decide to intervene in some things and not others. This decision-making process typically is called *case formulation* or *case conceptualization* and is defined as "a general model... to understand problems and generate solutions to them, based on this understanding, in a coherent, systematic way" (Persons, 1989, p. xiii). Whether the case formulation process is formal and uses empirically supported models—such as behavioral couples treatment (BCT), e.g., Jacobson and Margolin, 1979; emotionally focused therapy for couples (EFT), Greenberg and Johnson, 1988; and insight-oriented couples therapy, Snyder and Wills, 1989; Snyder, Wills, & Grady-Fletcher, 1991—or is less formal, intervention is always founded on some approach to understanding relationship dysfunction and preventing or ameliorating it.

What is this understanding based on? Ideally, it would be based on research demonstrating the risk and protective factors for couple dysfunction. Yet, different theoretical schools have different ideas about what the etiological and maintaining factors are in relationship dysfunction, leading them to assess different constructs and to develop different interventions. Further, couples themselves have something to say about what they believe to be the important factors in their dysfunction. Unifying the chaos is this simple fact: Couple communication is the common pathway to relationship dysfunction across theories, therapies, therapists, and clients. All theories and therapies emphasize the role of communication (see Jacobson & Gurman, 1995), and both therapists (Geiss & O'Leary, 1981) and couples (e.g., Storaasli & Markman, 1990) rate communication as the top problem area.

However, every clinician knows two things about couples' complaints about communication. First, such presenting problems tell us everything and nothing at all. Communication is the common pathway to relationship dysfunction because it is the common pathway for getting what you want in relationships. Nearly all relationship-relevant conflicts, emotions, and neuroses are played out via observable communication—either verbally or nonverbally. A conceptualization of "the husband is "unhappy because he doesn't communicate well" is about as useful a conceptualization as "the patient died because his heart stopped beating." Second, couples' reports of their difficulties, although useful in understanding their own conceptualizations of their distress (e.g., Buehlman, Gottman, & Katz, 1992), may not provide the information necessary to construct useful case conceptualizations and treatment plans. Partners' reports are subject to attributional biases (see Bradbury & Fincham, 1990) and selective attention. A particularly common form, *sentiment override* (Weiss, 1980), involves distressed individuals attending almost exclusively to their partners' negative behaviors and interpreting even neutral or positive behavior through a negative filter (cf. Fincham, Gamier, Gano-Phillips, & Osborne, 1995). Therefore, outsiders' observation of communication can add very useful, nonredundant information to that gleaned from self-reports. This information is critical to both research models of marital distress (e.g., Gottman, 1994) and to clinicians' case conceptualizations (e.g., Gottman, 1999) and treatment plans (e.g., O'Leary, Heyman, & Jongsma, 1998).

However, for communication to be a useful concept for either researchers or clinicians, it must be operationalized and put to empirical test. For over 25 years, researchers have done just that and have produced an impressive body of "results deserving of belief, [and] results deserving of incredulity." Only after discerning which results deserve belief and which results deserve

incredulity will psychological approaches to relationship distress advance as a clinical science and not just as a therapeutic art.

Couples Observation Research: A Primer

In later sections I critically examine the methods and findings of the published observational literature. Before stakeholders can evaluate if couples observational work tells them something of importance, however, they must be conversant with the methods used to observe couples' communication and with the most robust findings derived from such work. (More detailed recent reviews can be found in Gottman, 1994, 1998, and Weiss and Heyman, 1990, 1997, among others).

Standard couples observational paradigm—The standard couples observational paradigm can be used in a typical intake assessment. In research contexts, investigators ask couples to discuss 1–2 conflict areas for 10–15 min each; in clinical contexts, about 5–7 min each is minimally sufficient (e.g., Gottman, 1999). Although in research contexts no one is in the room except for the partners, in clinical contexts the therapist usually instructs the couple to act as if she were not there, stays in the room, and takes notes about the communication process to provide feedback later (O'Leary et al., 1998).

The external validity of the standard couples observational paradigm has been established in several ways. First, Gottman (1979) compared home and laboratory observations of couples and found substantial similarities, with lab discussions overall being less negative. (Gottman and Krokoff, 1989, replicated lab-home behavioral similarities for husbands but not for wives.) Second, Foster, Caplan, and Howe (1997) had couples rate the typicality of their partners' behavior following a 15-min conflict videotaped at home. About half of the time, the partner was judged to be acting typically. When not acting typically, partners were far more likely to be judged as being more supportive and less undermining than usual. Thus, if anything, laboratory observations understate the differences between couples by reducing the variability of negativity. Third, spouses' self-consciousness and reactivity while being observed are relatively low (Christensen & Hazzard, 1983; Jacob, Tennenbaum, Seilhamer, Bargiel, & Sharon, 1994). Finally, Vincent, Friedman, Nugent, and Messerly (1979) demonstrated that even when couples are instructed by the researcher to "fake good" or "fake bad," observers can still reliably discern happy from unhappy couples. Unhappy couples leak negative affect even when they are trying to behave as if they were happy. To summarize, even if typical interaction samples researchers have collected are not quite as negative as they are at home, they still reveal detectable differences in affect, behavior, physiology, and interactional patterns and processes (Gottman, 1979, 1994, 1999).

Distressed couples' communication patterns: Stubborn facts—Across coding systems, countries, studies, spouses, and researchers, several "stubborn facts" (Notarius & Markman, 1989a) about observed couple processes have emerged: Distressed partners, compared with nondistressed partners (a) are more hostile, (b) start their conversations more hostilely and maintain it during the course of the conversation, (c) are more likely to reciprocate and escalate their partners' hostility, (d) are less likely to edit their behavior during conflict, resulting in longer negative reciprocity loops, (e) emit less positive behavior, (f) suffer more ill health effects from their conflicts, and (g) are more likely to show demand → withdrawal patterns.[3] Furthermore, both partners in distressed relationships characterized by husband-to-

[3]Research summarized in this article emanated from Australia, Canada, Germany, Holland, Spain, and the United States. "Healthy" couple behavior is undoubtedly culturally determined. Thus, the results should be interpreted, given the fact that the participants are by and large, European or European-descended middle class volunteers observed in the latter part of the 20th century.

wife aggression, compared with distressed/nonaggressive relationships, are more hostile and reciprocate hostility more.

Happy couples' communication patterns—Very little is known about the communication patterns of happy couples, other than that they differ from distressed couples in the ways listed above. As Robert Weiss has written,

> The literature often suggests an illogical conclusion: marital health is the absence of marital distress. This seems to suggest that in order to be well adjusted couples should not say and do what distressed couples say and do. Since aspirin cures headaches, we cannot conclude that the lack of aspirin causes headaches. Marital harmony is not just the absence of whatever it is that dissatisfied couples do. (Weiss & Heyman, 1997, p. 92)

Research on what communication facets result in happiness is expanding but is still in the early stages. The traditional conflict-resolution paradigm and its associated coding systems seem well suited to understanding what nondistressed couples *do not* do that perhaps protects them from distress but is poorly suited to understanding what they do do that promotes satisfaction. Whereas it is relatively easy to get unhappy couples to argue on command, behaviors that promote the various forms of love (e.g., Aron & Aron, 1986; Sternberg & Barnes, 1988), such as flirtation, support during acute crises, and spontaneous acts of caring, are much more challenging to witness in the laboratory (or even at home with sufficient frequency to allow for meaningful analyses). Thus far, published research has focused on social support provision by instructing partners to discuss individual problems and coding observed socially supportive behaviors (e.g., Pasch & Bradbury, 1998). However, a promising new approach by Roberts and Linney (in press) has partners write down their vulnerabilities and then discuss them with each other for 10 min. This paradigm, and its accompanying coding system, have excellent preliminary indications of content and construct validity.

As a result, the only stubborn truth to emerge regarding individuals in happy relationships is that they do not naturally paraphrase their partners' statements nor reflect their partners' implied feelings back to them during conflicts (Gottman, Coan, Carrere, & Swanson, 1998; Heyman, Weiss, & Eddy, 1995).[4] Whether this stubborn truth means that therapists should abandon teaching this unnatural behavior (Gottman et al., 1998) is being hotly debated (Gottman, Carrere, Swanson, & Coan, 2000; Stanley, Bradbury, & Markman, 2000). Both researchers and clinicians should be watchful about content validity issues here (i.e., would we expect active listening to be elicited in conflict, compared with nonconflict, situations?). Roberts and Linney (in press) successfully observed couples (especially happier couples) emit active listening behaviors during discussions of their vulnerabilities.[5] Although we have learned a lot about dysfunctional relationships by watching what unhappy couples do when they are fighting, we will probably only learn a lot about functional relationships by watching what happy couples do when they are not fighting.

Psychometric Properties of Couples Observation Constructs

There are many resources that cover the basic issues, technical issues, or both involved in observational coding of couples (e.g., Bakeman & Gottman, 1997; Gottman, 1979; Markman & Notarius, 1987; Notarius & Markman, 1989b). At the request of the journal editor, however, this review takes a very different tack by focusing on the published evidence for the

[4]Heyman et al.'s (1995) factor analysis of the MICS demonstrated that paraphrase/reflect was correlated nearly 0 with all four obtained factors. Although not presented in the published article, this was probably due to this code's extremely low frequency of occurrence.
[5]"Content validity is the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose" (Haynes et al., 1995, p. 238).

psychometric properties (i.e., reliability and validity) of couples observational measures. The evidence is presented in Table 1.

Three critical decisions shaped Table 1. First, I used a liberal criterion on what constitutes evidence of validity. Although few couples observation studies assess validity of the coding systems directly (i.e., establishing validity is the stated purpose of the study), most studies assess it indirectly (i.e., substantive studies that contrast distressed and nondistressed couples may be establishing the discriminative validity of the system) and were therefore considered worthy of inclusion.[6]

Second, I focused on constructs because the very purpose of psychometric evidence is to show that a device measures a construct with little error (i.e., reliability) and that the device measures the construct it is intended to measure and nothing more (i.e., validity).[7] One cannot establish the reliability and validity of a coding system overall but of specific constructs contained within the larger coding system.[8] Yet, constructs have been created in idiosyncratic ways to answer substantive questions without much attention being paid to establishing the psychometrics of the construct in question.

Third, I focused on replication. The hallmark of science is that results are replicable. Validity is not established by findings from one study but through the accumulation of evidence across studies. The over 25 years that this field has been in existence, producing nearly 300 studies, has given it ample opportunity to produce replicable findings. Therefore, Table 1 presents psychometrics only for studies containing at least one construct measured in the exact same way across two or more studies.

Reliability (Temporal Stability)

No studies have examined the test-retest reliability of observed couple behaviors in its conventional sense (two or more behavioral samples collected over a relatively brief period with the exact same procedures, to test the reproducibility of results). The closest research is a generalizability study of clinic couples by Wieder and Weiss (1980) that included a 1-week reassessment facet. (The other facets were couples, $n = 14$, and coders, $n = 4$). Couples' top two problems were videotaped in a laboratory setting. For most coding categories, 1 % or less of the variance was attributable to the reassessment facet alone, whereas substantial percentages of variance (36% of total positive behavior and 44% of total negative behavior) were attributable to differences across couples. However, the Reassessment × Couple facet accounted for large percentages of variance in observed behaviors (i.e., 46% of the total positive behavior and 26% of the total negative behavior). Thus, couples do behave differently on different occasions while discussing different topics.[9] Although this is to be expected (repeated

[6]To be included, the tested observational construct had to be conceptually related to the dependent variable in either a hypothesized or obvious, common sense manner. This is an especially difficult decision for studies testing discriminative relations; I included these studies only if they shed light on the validity of the observed construct (i.e., the observed construct would be expected to distinguish between the groups). Thus, a study demonstrating differences in hostility between couples with and without a spouse in chronic pain would not be included, whereas studies demonstrating differences in hostility between maritally distressed and nondistressed couples would be included.

[7]Throughout this article I will refer to the aggregation of codes as "constructs" rather than as response classes. As Foster and Cone (1995, p. 249) have noted: "Measures of hypothetical constructs and response classes are frequently difficult to distinguish ... because whether a hypothetical construct or a set of behaviors is being assessed lies in the interpretation of the assessor rather than in the measure per se." Because these aggregations are almost never based on functional analysis, they lack the necessary "empirical verification of functional homogeneity across their various topographies" (Barrett, Johnston, & Pennypacker, 1986, p. 170). Thus, it seems more appropriate to apply the more general term "construct."

[8]Although most observational studies have used ultramicroanalytic coding systems (e.g., MICS, CISS), most researchers combine codes to perform analyses, thus forming constructs by combining theoretically similar codes. If an ultramicroanalytic code—which may be descriptive of a particular observable behavior and thus not qualify as a construct—meets the criteria set for Table 1 (has reliability reported for the code and is analyzed separately in two or more studies), it was included. Second generation microanalytic coding systems (SPAFF and RMICS) code at the construct level, although researchers often combine these codes to form higher-level constructs.

observation of couple conflict is not the same as repeated administration of an intelligence test), it does indicate that differences in spouse behavior across topics and/or time cannot be ignored.[10] It is possible that the results from any one observation are akin to those of a single item on a questionnaire. Psychologists have long known that scales produce far more reliable results than do single items. It is possible, although certainly not guaranteed, that multiple observations (across a representative number of topics/situations) are necessary to obtain reliable results. Further work is necessary to establish (a) if multiple observations are necessary; (b) what the optimal number of observations is; and (c) what the impact of such changes would be on substantive results.

Stability of couple interactions across time—Two studies have examined stability of laboratory assessments of top couple conflicts across several years. Gottman and Levenson (1999b), using the Specific Affect Coding System (SPAFF; Gottman, 1996), observed married couples twice over a 4-year period; Lord (1999), using the Rapid Marital Interaction Coding System (RMICS; Heyman & Vivian, 1993), observed couples three times over the first 5 years of marriage. Gottman and Levenson (1999b) found evidence of significant stability (with variance accounted for in the same range as those of Wieder & Weiss, 1980).[11] Lord (1999) found little stability in the coded behaviors of newly married couples. It is likely that Gottman and Levenson (1999b) and Wieder and Weiss (1980) found moderate degrees of stability across observations 1 week to 4 years apart because the couples had been married about 5 years, whereas Lord's (1999) couples were transitioning to marriage (observed at premarriage, 18 months, and 5 years). Obviously, it is difficult to make firm conclusions about two studies that obtain divergent results.

Length of time necessary to make reliable base-rate estimations—Stability is dependent on who is being studied, on how frequent the codes of interest are, and how long the observations are. Because the first two factors are often the independent and dependent variables of the study, the length of observation is typically an invariant, methodological decision made by the investigator. Although the 10–15 min observation is standard in the couples field, no published study has examined the adequacy (i.e., incremental validity) of this sample length. Heyman, Chaudhry, et al. (in press) recently used Waters' (1978) method of using Spearman-Brown coefficients to estimate the amount of time necessary to code behaviors reliably. Using three different samples (engaged, non-distressed community, and clinic), they found that 10-15 min of laboratory interaction was enough to witness enough behavior to make reliable (i.e., internally consistent) estimations of most RMICS code frequencies, as well as the frequency of negative reciprocity. Note that this does not contradict the stability findings, cited above, that found strong Situation × Couple interactions; rather, it indicates that 10–15 min is enough time to observe a single situation for most couples.

Reliability (Interrater Agreement)

Although interrater agreement is not reliability in the classical sense (i.e., estimation of true score via stability of results across observations), it does provide some evidence that the data derived from an observation are reliable across observers. This evidence is necessary to

[9]Given the significant interaction, the near-zero percentage of variance accounted for by the reassessment facet may be due to couples' changes from observation 1 to 2 canceling each other out. Thus, even though there's no overall difference between the first and second observations, this *does not* mean that behavior is stable across the observations.

[10]Christensen and Heavey (1990) and Heavey, Layne, & Christensen (1993) demonstrated a special instance of a test-retest by spouse interaction: whether the conflict topic is the husband's or wife's.

[11]Because Gottman & Levenson (1999b) study did not use a generalizability framework, the variances accounted for are estimated from the correlations presented. No direct comparison between the variances accounted for in this study and those reported by Wieder and Weiss (1980) is possible.

demonstrate that the obtained results have more to do with important differences in the couples' behavior and not with differences across coders.

Table 1 displays the interrater agreement statistics for all validity-related studies. Although the vast majority of studies in Table 1 provided some interrater agreement statistic, most did not warrant inclusion in Table 1 for one of two reasons. First, agreement is useful only when it is provided at the level that analyses are made. Many investigators provided agreement at the level of the coding system (e.g., the entire MICS), not the constructs being investigated (e.g., put-down code). Second, investigators some-times provided only the range of agreement found, making it difficult for the reader to ascertain if any particular construct was adequately measured. Although both points are fundamentals of research design, only about 20% of the published validity-related studies included reliability information for the constructs studied.

It can be argued that poor interrater agreement most likely adds error variance, not systematic variance, and thus poorly measured constructs that produce significant results negate worries about reliability. This argument is dangerous for two reasons. First, low reliability measures will, in the long run, hamper the advancement of science through (a) failures to replicate that are due to poor measurement rather than to substantive problems; (b) wasted time and money because of theorizing and attempting to replicate Type I errors that were due to a combination of low reliability and inflated Type I error levels (due to the failure to control for family-wise error); and (c) content and discriminant validity problems due to constructs' actual coding overlapping.[12] Second, our work garners tremendous attention from the general public and clinicians alike, and publicly disseminated "facts" on the basis of sloppy science are difficult to correct.

Validity

Validity is the extent to which a test measures what it is intended to measure. As Haynes, Richard, and Kubany (1995, pp. 239–241) note, "[V]alidity is a state, not a trait, of an obtained assessment instrument score... Statements such as ' ... has been shown to be a reliable and valid assessment instrument" do not reflect the conditional nature of validity and are usually unwarranted." To paraphrase Gordon Paul, validity is inferred from the cumulative results regarding *what* measure, administered *when*, is an accurate measure of *this* construct with *that* population and under *which* set of circumstances.[13]

Given the conditional nature of validity, the breadth of measures (i.e., coding systems), administered during a variety of life stages (e.g., premarriage, pre- and posttreatment, heterogeneously across years of marriage) and measured with a dizzying number of construct operationalizations with varied populations (e.g., alcohol dependent, partner abusive, distressed) under varied circumstances (i.e., lab vs home, specific way in which conversation was set up), neatly summarizing Table 1 is impossible. Instead, I offer a process for how clinicians and researchers can identify their specific needs and thus extract the relevant information from Table 1.[14] A flowchart (Figure 1) is provided to summarize this process.

---

[12]"The major problem resulting from the performance of a series of analytical comparisons on a set of data is the unpleasant fact that the more comparisons we conduct, the more type I errors we will make when the null hypothesis is true ... If we evaluated several comparisons in an experiment, each $\alpha = .05$, our probability of making a type I error would be .05 for each of the separate comparisons" (Keppel, 1991, p. 164). For example, Gottman and Krokoff (1989) present 20 correlations for husbands and wives at both initial assessment and follow-up (i.e., 80 correlations). If each gender's correlations at each time point were considered a family, $\alpha = .05$ would be maintained for the family-wise comparisons by dividing the alpha by the number of comparisons (i.e., .05/20 = .0025). Thus, a *p* value of .0025, not .05, would be required to be considered significant. This formula is known as the Bonferroni inequality. Note that because family-wise comparisons are being made, the alpha is .05 for each gender at each time point, not for the entire set of comparisons (which would require .05/80 = 0.000625). To obtain adequate power (.80) to detect a moderate effect size ($r = .3$), larger sample sizes would be necessary ($N = 139–174$, assuming one-tailed tests).

[13]"What treatment, by whom, is most effective for this individual with that specific problem and under which set of circumstances?" (Paul, 1967, p. 111)

As shown in Figure 1, clinicians typically do not have the inclination or resources to use coding formally. Thus, ultramicroanalytic (e.g., MICS, Couples Interaction Scoring System [CISS]) coding systems are overly laborious, whereas coding systems with established (e.g., SPAFF, Kategoriensystem fur Partnerschaftliche Interaktion [KPI]) or growing (e.g., RMICS) clinically relevant validity data bases may be useful for making informed decisions based on in-session conflict observations.

Before researchers can select a suitable coding system, a myriad of questions must be asked and answered. What is your general research question? What are you trying to find out via observation? Coding systems are tools and as such should be chosen to fit a particular need, as Bakeman and Gottman (1997, p. 15) have noted:

> We sometimes hear people ask: Do you have a coding scheme I can borrow? This seems to us a little like wearing someone else's underwear. [Using] a coding scheme is very much a theoretical act, one that should begin in the privacy of one's own study, and the coding scheme itself represents an hypothesis, even if it is rarely treated as such.

Once the research question is firmly in hand, more specific hypotheses can be made, and a search for an appropriate preexisting coding system can begin. Are you interested in supportive or conflictual behaviors? At what level of analysis (i.e., global or specific) are your questions (see Notarius & Markman, 1989a; Weiss, 1989)? For example, it is inefficient and expensive to use an ultramicroanalytic coding system (e.g., MICS), which comprises nearly 40 codes, if the hypothesis involves only positive, negative, and neutral behavior. Some constructs (e.g., global negativity, secure base attachment use) are more easily measured at a global level, whereas others (e.g., frequency of distress-maintaining attributions) are more easily measured at a micro level. What population are you studying? Coding systems that have established reliability and validity for one population may no longer be reliable/valid when applied to couples or coders from other populations (e.g., differing from the original study on racial/ethnic background, geography, psychopathology). Figure 1 provides suggestions for ways in which Table 1 can be personalized to see if a coding system exists with preexisting conditional reliability and validity for your needs.

Overall, culling the validity information in Table 1 provided disconcerting news concerning validity. The most pressing problem is that investigators have taken ultra microanalytic coding systems (e.g., MICS) and mixed and matched codes at will. Although some have described this as a strength of ultra microanalytic coding systems (e.g., Markman, Leber, Cordova, & St. Peters, 1995), it rarely is accompanied by the reliability and validity work necessary to establish the construct as adequately measured. Such inventive code combining—especially when accompanied by a failure to provide reliability information on the new construct— falls short of true construct building, and thus the validity results were censored in the creation of Table 1. This is not to say that such new constructs necessarily lack validity, but only that I imposed a lenient criterion that was not met (i.e., that because validity is inferred from an accumulation of evidence, *accumulation* required at least two studies to test the construct).

With this criterion in place, the most widely used coding system in the field, the MICS, retains little validity information. Because many of the stubborn facts about marriage were derived from MICS studies, the lack of retained validity information is almost certainly due not to a lack of construct validity but to a lack of agreement on how to construct the constructs. Two solutions to the MICS quandary are available. First, the MICS has demonstrated preliminary evidence of factorial validity by having three independent exploratory factor analytic studies

[14]To facilitate individual extraction of Table 1 results, we have made available (at http://www.psy.sunysb.edu/marital and at http://www.aabtcouples.org) Table 1 and associated files on how various studies combined codes to form their constructs.

—one large (*N* = 995, Heyman, Weiss, et al., 1995) and two small (*N* < 100, Jacob & Krahn, 1987; Kiecolt-Glaser et al., 1996)—produce similar solutions for combining MICS codes into categories. Investigators can create categories consistent with the large factor analysis. Second, Heyman and Vivian (1993) have created the RMICS based in large part on the Heyman, Weiss, et al. (1995) factor analysis. The RMICS and the other major second generation microanalytic coding system, the SPAFF (Gottman, 1996), code at the construct level—using about half as many codes as the MICS or CISS—thus presenting researchers with fewer quandaries about how to combine codes into categories.

Because the SPAFF was the first second-generation coding system developed, it has by far the best evidence of construct and criterion validity for its constructs. For example, SPAFF affection, anger, belligerence, contempt, domineering, humor, sad, and validation codes all have findings supportive of discriminative validity. The high intensity negative summary category (i.e., belligerence, defensiveness, contempt) has shown preliminary signs of predictive validity. Although these codes have been used in different configurations, it appears overall that these negative affects are risk factors for later divorce. However, predictive validity for individual risk factors for divorce does not imply validity for risk factors predicting individuals' divorces. Because researchers have confused the two, and because these studies have received so much attention in the mass media, I now discuss the predictive validity problem in more detail.

Predicting couple outcomes (predictive validity)—"The age of media foolishness" includes the uncritical mass dissemination of interesting findings, such as researchers' ability to predict who will divorce with near perfect accuracy (see Rogge & Bradbury, 1999, for a table listing the predictive claims of eight studies; see also more recent studies not included in the table: Carrere & Gottman, 1999; Gottman et al., 1998; Gottman & Levenson, 1999c). Once such findings enter the media echo chamber, they become established truths, impervious to later refutation, regardless of the soundness of the proof for the supposed truth (see Faludi, 1991).

Psychological studies are adept at identifying correlates (also called risk factors) of dysfunction. Because human behavior is multidetermined, individual prediction is extremely difficult. Yet discriminant function analyses and logistic regression can still be used to create weights and cut points to optimally predict individuals' likelihood of divorce. More accurately, however, we should say that the analyst asks the software to reconstruct, rather than predict, because the computer develops an equation to optimally reconstruct an already known group status. This is not a trivial, semantic distinction. Because of the problem known as overfilling (e.g., Tabachnick & Fidell, 1996), such statistical techniques are overly swayed by idiosyncrasies in a particular data set, causing the solution to be not generalizable and the estimations of the equation's predictive powers to be extremely overinflated.[15] Overfilling is most severe in the small sample (typically under 100 participants) studies typical in the couple observation field.

To establish the predictive validity of an equation, it must be crossvalidated in an independent sample. To demonstrate this point, Amy Slep and I (Heyman & Slep, in press) recently developed and crossvalidated a logistic regression equation to predict divorce in a large, nationally representative data set (Gelles & Straus, 1994). Like Gottman et al. (1998), we were able to predict divorce correctly for 90% of couples. However, in the crossvalidation, the accuracy fell by-one-third. Furthermore, in the cross-validation the equation was right only 29% of the time it predicted a couple was divorced. Similar results were found when developing

[15]Many of these studies, including Gottman et al. (1998), oversampled extreme groups, which compounds overfitting (Rogge & Bradbury, 1999).

and crossvalidating prediction equations for physical and emotional abuse in a data set of over 30,000 individuals in relationships (Heyman & Slep, 1999). In conclusion, prediction studies may inform us of heightened relative risk, but because of overfitting, should be ignored regarding supposed prediction (unless, of course, the equation, using the exact weights and cut-points from the development sample, crossvalidates in an independent sample).

Recommendations for Reducing Measurement and Inferential Errors

The following eight recommendations are for improving normal science within the couple observation field. As such, they are designed for incremental improvement of the field, not for fomenting scientific revolution (Kuhn, 1970).

Recommendation #1: Have a Theory

"All assessment systems, including behavioral observation, are based on inferences about some construct which they are assumed to measure" (Haynes, 1978, p. 177). Constructs and response classes are, by definition, theoretical entities. One cannot use science to discover stubborn facts without reliable and valid procedures and measures. One cannot establish procedures and measures as content and construct valid without theory about what they should be measuring. Although this seems rather obvious, a large number of studies in Table 1 seem to lack a theoretical structure for their hypotheses or for their use of observational systems, or both.

Recommendation #2: Make Construct Validity a Prime Concern

As Table 1 demonstrates, the vast majority of studies use idiosyncratic code combinations, making it nearly impossible to evaluate the construct validity of the coding systems. Just because a system has been used before does not mean that it is valid for the uses intended for a particular study. Furthermore, idiosyncratic combination of codes not only means that prior construct-validity information is no longer pertinent but also impedes the agglomeration of validity data.

Recommendation #3: Evaluate the Reliability of Constructs at the Level of Analysis

In classical measurement theory (e.g., Wiggins, 1973), validity is constrained by measurement error. Presenting data on the reliability of a coding system in its entirety obfuscates whether the constructs being tested are reliable. Reliability data should be presented for all constructs tested.

Recommendation #4: Move Toward Multimethod Assessment of Constructs

Patterson and his colleagues at Oregon Social Learning Center (OSLC) have, over the past 25 years, developed and tested multitrait, multimethod, multireporter models of child conduct disorder etiology and treatment (e.g., Bank, Dishion, Skinner, & Patterson, 1990; Bank & Patterson, 1992; Conger, Patterson, & Ge, 1995; Dishion, Li, Spracklen, Brown, & Haas, 1998; Eddy, Dishion, & Stoolmiller, 1998; Patterson, 1982, 1993; Patterson, Reid, & Dishion, 1992). But, as Patterson et al. (1992) describe it, they were coerced into shifting from being stubborn truth searchers to model builders:

> In the 1980s ... a group of... site visit[ors] from our funding agency asked 'Where are your theories?' and 'Where are your models?' Our answer was that we were behaviorists and that our strategy was to obtain data first and then develop a theory if one were justified. Their response was terse and to the point: ... 'If you want to collect data at all, you must first show us a model, (p. 1)

The resulting model, using structural equation modeling and other advanced statistical techniques, has empirically demonstrated how children develop into antisocial adults (i.e.,

basic training in coercion at a young age via parental ineffectiveness; reaction of the social environment via school failure and peer/adult rejection; adolescent association with deviant peers and refinement of antisocial skills; adult adjustment problems and antisocial behavior). Those interested should read Patterson et al.'s (1992) book, which describes their 10-stage method of model building (including construct validation) and the results of the empirical tests of the model.

Recommendation #5: Enhance Internal and Content Validity by Experimentally Controlling Discussion Topics

This area is rife with procedural and coding idiosyncrasies, which imply a problem with content validity. Content validity is too broad an area for thorough discussion here, and the interested reader is directed to Haynes et al.'s (1995) excellent overview. Although several content validity-related issues have been discussed (e.g., determining which codes to include in a construct, situational effects), three additional concerns are of note. First, more observational research must be conducted on nonconflict situations, such as expressing vulnerability or seeking/providing social support. Second, researchers have introduced unnecessary error variance by exerting too little experimental control in the selection of discussion topics. Third, researchers have paid too little attention to the gender of the complainant when choosing topics for discussion. I recommend that researchers (a) select the topics to be discussed; (b) narrow down broad topics such as communication through either a play-by-play interview (Gottman, 1996) or a specific questionnaire such as the Areas of Change Questionnaire (Weiss, Hops, & Patterson, 1973); (c) standardize (within and/or across studies) communication task instructions to couples and report them in published studies; and (d) experimentally control the gender of the complainant by either choosing two topics (e.g., the top female and male topics from a problem list) or by keeping the complainant's gender constant Clinicians should follow the same suggestions, except they should always watch (atleast) the male's and female's top topics.

Recommendation #6: Pay Attention to Validity of Cutpoints for Contrasted Groups

Researchers have paid too little attention to the distressed/nondistressed distinction in forming contrasted groups. Marital adjustment is not measured without error, and classifying an individual with a DAS score of 97 as distressed and one with a score of 98 as nondistressed is not empirically supportable. As an initial step, I recommend that we heed Jacobson and Truax's (1991) suggestion to use error bands in developing classification cutoffs. Furthermore, to provide construct validity, such cutoffs should be validated against a clinical diagnostic interview (see Heyman, Feldbau-Kohn, Ehrensaft, Langhinrichsen-Rohling, & O'Leary, in press).

Recommendation #7: Beware of Family-Wise Error

Researchers should either control for family-wise error or label their findings as exploratory.

Recommendation #8: Conduct Further External Validity Research

As a check on the external validity/generalizability of observed communication, partners should be asked to report on how representative the observed interaction was of similar conversations at home (see Foster, Caplan, & Howe, 1997, for a well-developed instrument.) In addition, further work comparing laboratory and home observations is necessary because (a) external validity results for wives' have been equivocal and (b) no published reports have examined the consistency of behavioral sequences across lab and home observations.

Clinical Assessment

Like any scientific field, couple observation research is not without its shortcomings, contradictions, and controversies, which tend to put off nonresearchers. This reality may dissuade many clinicians from including observation routinely in the pre- and posttreatment assessments, which would be a shame, considering all that observation has to offer clinicians pressed to use empirically supported treatments and to develop treatment plans consonant with such treatments. This final section describes how one can sensibly incorporate observation in clinical practice.

Use of Observation to Identify Problem Behaviors and Interactions

Although we cannot say whether marital distress causes high levels of hostility, or whether high frequency and intensity of hostility cause marital distress (or even whether some third factor, like incompatibility or neuroticism, causes both), high levels of hostility are the primary presenting problem for marital therapy (O'Leary, Vivian, & Malone, 1992). To develop adequate case conceptualizations and treatment plans, clinicians must be familiar with behaviors that are normative and those that are associated with distress. Several examples from one of the most complete theories in the field, Gottman's balance theory (e.g., Gottman, 1994), are instructive here.[16] First, Gottman and colleagues have reported that high intensity negative affect (i.e., belligerence, defensiveness, contempt), but not low intensity anger, is associated with high frequency husband-to-wife physical aggression (Jacobson et al., 1994) and is a risk factor for later divorce (Gottman et al., 1998). Second, discussions of distressed, but not nondistressed, couples start negatively and never recover (e.g., Carrere & Gottman, 1999; Gottman, 1979, 1994, 1999). Third, negative reciprocity, but especially husband's escalation from low to high intensity hostility, is a risk factor for later divorce. Gottman's (1999) latest book, subtitled "A scientifically based marital therapy," provides an expansive delineation of how he incorporates marital observation research findings into the assessment and treatment of couples.

While watching couples' conversations during assessment sessions, I ask myself the following questions: How does the conversation start? Does the level of anger escalate? What happens when it does? Do they enter repetitive negative loops? Do they indicate afterward that what occurred during the conversations is typical? Is their behavior stable between the two discussions? Do their behaviors differ when it is her topic versus his? Do they label the other person or the communication process as the problem? Because most forms of marital therapy include attempts at modifying couples' communication behaviors, being familiar with the basics of communication processes is very useful if one is to recognize communication faults and, importantly, being able to set appropriate treatment goals (e.g., teaching partners how to monitor and exit negative loops rather than admonishing them not to behave hostilely).

[16]Gottman's (1994, 1999) balance model borrows from physics and attempts to integrate psychophysiology, affect, behavior, cognition, couple typologies, and change over time. Gottman describes two levels of processing behavior: p-space and q-space. P-space is the overt behavioral level (conveniently represented by the ratio of positive to negative behaviors unfolding across time). Q-space is the subjective sense of well-being in the relationship. When the ratio of positive to negative behaviors dips below a threshold, the q-space variable flips from a positive to a negative state. Obviously, negative q-space cognitions, when held strongly enough, will begin to affect overt behavior. If this pattern continues, the behavioral interactions between the spouses will continue to deteriorate. When q-space remains negative for a long time, it can become p-space (e.g., fights over the selfish motivation of the partner, tracking the partner's selfishness); q-space then jumps to a higher level of abstraction (e.g., "He's a selfish person."). This process can iterate several times (eventually to "His selfish nature is making this relationship unsalvageable."). Left unchecked, spouses increasingly take steps toward divorce. Gottman (1999) has recently described the "Sound Marital House Theory," which presents many of these concepts in a far more user-friendly manner.

Incorporating Behavioral Observation Into Multimethod Assessment at Pre- and Posttreatment

In clinical practice, it is important (and becoming increasingly mandatory) to record a formal treatment plan that incorporates observable treatment goals (O'Leary et al., 1998). Observing communication during initial assessment and then during the course of treatment is necessary to assess the success in meeting the goals. (This, of course, will depend on both the goals and the therapists' functional analysis of what is promoting and maintaining the distress. Communications' importance, and thus the importance of assessing it, will vary across couples.) Although enlisting, observing, and coding couples for research is a difficult and expensive proposition, more informal observation costs nothing other than time in pre- and posttreatment assessment in clinical practice. For a well-informed clinician looking to create a solid treatment plan, I believe it to be time well spent.[17]

Pre- and posttreatment (and sometimes follow-up) observational assessment has been used in outcome studies to assess couples' learning of communication skills (e.g., Hahlweg, Schindler, Revenstorf, & Brengelmann, 1984; Gottman, 1979; Jacobson, 1977, 1978; Sayers, Baucom, Sher, Weiss, & Heyman, 1991). However, Jacobson (1985) has argued that, at most, these studies should be seen more as a manipulation check of the experiment (i.e., clinical intervention)—and one that is extremely susceptible to reactivity effects, where clients try to please the therapist by demonstrating the communication skills they have been taught— than as true treatment evaluation. He suggested that self-reports of marital satisfaction reign supreme. Although therapists often use multimethod assessment in clinical practice, its use in research is more complicated. However, the couples observation field is well behind the most sophisticated lab in the family observation field (e.g., Patterson, 1982; Patterson et al., 1992) in developing multi-construct, multimethod models and in testing interventions on the basis of these models.

Conclusion: Whither Couples Observational Research?

Ten years ago, Notarius and Markman (1989b) certified some of the findings reviewed here as having met Cook and Campbell's (1979, pp. 24–25) admonition that psychologists should discover "stubborn facts that speak for themselves .. . [and] are worthy of theoretical efforts." The hard work necessary to build a body of literature identifying these stubborn facts has resulted in the "best of times" for the field so far.

However, as I stated in the introduction, I believe that the problems of this field 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. Until one peruses the entire body of work summarized in Table 1, it is difficult to discern that our creativity and enthusiasm have gotten ahead of our science. I write this conclusion on the final day of the 20th century. May the 21st century find us iterating through the research cycle to tie up our loose ends by using modern model building technology (e.g., structural equation modeling, latent growth curve modeling, hierarchical linear modeling) to build solid constructs that are reliable, valid, replicable, and worthy of theoretical efforts. Some in the family observation field (e.g., Patterson et al., 1992) have been building such models for over 20 years and have demonstrated not only that it can be done but also that it is worth the effort. May the 21st century be truly the spring of hope for science in this area, not because our noisiest authorities proclaim it to be, but because we have owned up to our failings and striven to correct them.

[17]Clinicians interested in observing couples in clinical practice should probably familiarize themselves with the coding systems that have been used in the studies reviewed here. SPAFF is available, with a host of other measures, in book form (Gottman, 1996). Those interested in the MICS are referred to a factor-analysis-derived, second generation version, the RMICS (http://www.psy.sunysb.edu/marital). The Social Support Interaction Coding System (SSICS) can be found at http://www.psych.ucla.edu/resources/newed/ss.htm.

Acknowledgements

The entire field owes a debt of gratitude to Danielle A. Black (assisted by Daniela M. Costa and Katherine E. Arnold) for reading the body of published work and constructing the table summarizing it. Thanks also to Daniela Costa, Margaret Kopko, Ali McLaughlin, and Dianne Orlando for their help in locating and cataloging the published articles. Finally, thanks to Amy M. Smith Slep and Stephen N. Haynes for their insightful comments about an earlier draft.

References

This list includes all references from the unabridged table available on the World Wide Web at http://www.psy.sunysb.edu/marital or http://www.aabtcouples.org or from the author upon request.

Adam D, Gingras M. Short- and long-term effects of a marital enrichment program upon couple functioning. Journal of Sex and Marital Therapy 1982;8:97–118. [PubMed: 7120436]

Alberts JK. An analysis of couples' conversational complaints. Communication Monographs 1988;55:184–197.

Andrews JA, Foster SL, Capaldi D, Hops H. Adolescent and family predictors of physical aggression, communication, and satisfaction in young adult couples: A prospective analysis. Journal of Consulting and Clinical Psychology 2000;68:195–208. [PubMed: 10780119]

Aron, A., & Aron, E. (1986). *Love and the expansion of self: Understanding attraction and satisfaction.* Washington, DC: Hemisphere Publishing Corp.

Aron A, Norman CC, Aron EN, McKenna C, Heyman RE. Couples' shared participation in novel and arousing activities and experienced relationship quality. Journal of Personality and Social Psychology 2000;78:1–12.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis.* (2nd ed.). New York: Cambridge University Press.

Ball FL, Cowan P, Cowan CP. Who's got the power? Gender differences in partners' perceptions of influence during marital problem-solving discussions. Family Process 1995;34:303–321. [PubMed: 8582477]

Bank, L., Dishion, T. J., Skinner, M., & Patterson, G. R. (1990). Method variance in structural equation modeling: Living with glop. In G. R. Patterson (Ed.), *Aggression and depression in family interactions* (pp. 247–279). Hillsdale, NJ: Erlbaum.

Bank, L., & Patterson, G. R. (1992). The use of structural equation modeling in combining data from different types of assessment. In J. Rosen & P. McReynolds (Eds.), *Recent advances in psychological assessment* (Vol. 8, pp. 41–74). New York: Plenum.

Barrett, B. H., Johnston, J. M., & Pennypacker, H. S. (1986). Behavior: Its units, dimensions, and measurement. In R. O. Nelson & S. C. Hayes (Eds.), *Conceptual foundations of behavioral assessment* (pp. 156–200). New York: Guilford.

Basco MR, Birchler GR, Kalal B, Talbott R, Slater A. The Clinician Rating of Adult Communication (CRAC): A clinician's guide to the assessment of interpersonal communication skill. Journal of Clinical Psychology 1991;47:368–380. [PubMed: 2066405]

Basco MR, Prager KJ, Pita JM, Tamir LM, Stephens JJ. Communication and intimacy in the marriages of depressed patients. Journal of Family Psychology 1992;6:184–194.

Baucom DH. A comparison of behavioral contracting and problem-solving/communications training in behavioral marital therapy. Behavior Therapy 1982;13:162–174.

Baucom DH, Sayers SL, Sher TG. Supplementing behavioral marital therapy with cognitive restructuring and emotional expressiveness training: An outcome investigation. Journal of Consulting and Clinical Psychology 1990;58:636–645. [PubMed: 2254512]

Behrens BC, Sanders MR, Halford W. Behavioral marital therapy: An evaluation of treatment effects across high and low risk ' settings. Behavior Therapy 1990;21:423–433.

Beier EG, Sternberg DP. Subtle cues between newly weds. Journal of Communication 1977;27:92–97. [PubMed: 853141]

Berns SB, Jacobson NS, Gottman JM. Demand-withdraw interaction in couples with a violent husband. Journal of Consulting and Clinical Psychology 1999a;67:666–674. [PubMed: 10535233]

Berns SB, Jacobson NS, Gottman JM. Demand/withdraw interaction patterns between different types of batterers and their spouses. Journal of Marital and Family Therapy 1999b;25:337–348. [PubMed: 10405919]

Biglan A, Hops H, Sherman L, Friedman LS, Arthur J, Osteen V. Problem-solving interactions of depressed women and their husbands. Behavior Therapy 1985;16:431–451.

Billings A. Conflict resolution in distressed and nondistressed married couples. Journal of Consulting and Clinical Psychology 1979;47:368–376. [PubMed: 469085]

Birchler GR, Clopton PL, Adams NL. Marital conflict resolution: Factors influencing concordance between partners and trained coders. American Journal of Family Therapy 1984;12:15–28.

Birchler GR, Weiss RL, Vincent JP. Multimethod analysis of social reinforcement exchange between maritally distressed and nondistressed spouse and stranger dyads. Journal of Personality and Social Psychology 1975;31:349–360.

Boelens W, Emmelkamp P, MacGillavry D, Markvoort M. A clinical evaluation of marital treatment: Reciprocity counseling vs. system-theoretic counseling. Behavior Analysis Modification 1980;4:85–96.

Bowman ML. Coping efforts and marital satisfaction: Measuring marital coping and its correlates. Journal of Marriage and the Family 1990;52:463–474.

Bradbury TN. Unintended effects of marital research on marital relationships. Journal of Family Psychology 1994;8:187–201.

Bradbury TN, Beach SRH, Fincham FD, Nelson GM. Attributions and behavior in functional and dysfunctional marriages. Journal of Consulting and Clinical Psychology 1996;64:569–576. [PubMed: 8698951]

Bradbury TN, Campbell SM, Fincham FD. Longitudinal and behavioral analysis of masculinity and femininity in marriage. Journal of Personality and Social Psychology 1995;68:328–341. [PubMed: 7877096]

Bradbury TN, Fincham FD. Attributions in marriage: Review and critique. Psychological Bulletin 1990;107:3–33. [PubMed: 2404292]

Bradbury TN, Fincham FD. Attributions and behavior in marital interaction. Journal of Personality and Social Psychology 1992;63:613–628. [PubMed: 1447688]

Bradbury TN, Fincham FD. Assessing dysfunctional cognition in marriage: A reconsideration of the Relationship Belief Inventory. Psychological Assessment 1993;5:92–101.

Buehlman KT, Gottman JM, Katz LF. How a couple views their past predicts their future: Predicting divorce from an oral history interview. Journal of Family Psychology 1992;5:295–318.

Burger AL, Jacobson NS. The relationship between sex role characteristics, couple satisfaction, and couple problem-solving skills. American Journal of Family Therapy 1979;7:52–60.

Burman B, John RS, Margolin G. Effects of marital and parent-child relations on children's adjustment. Journal of Family Psychology 1987;1:91–108.

Burman B, John RS, Margolin G. Observed patterns of conflict in violent, nonviolent, and nondistressed couples. Behavioral Assessment 1992;14:15–37.

Burman B, Margolin G, John RS. America's angriest home videos: Behavioral contingencies observed in home reenactments of marital conflict. Journal of Consulting and Clinical Psychology 1993;61:40–50. [PubMed: 8450106]

Caceres J. Experimental study of couples' problem solving skills: Are we striving in the right direction? Sexual and Marital Therapy 1989;4:161–175.

Carels RA, Baucom DH. Support in marriage: Factors associated with on-line perceptions of support helpfulness. Journal of Family Psychology 1999;13:131–144.

Carrere S, Gottman JM. Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion. Family Process 1999;38:293–301. [PubMed: 10526767]

Carstensen LL, Gottman JM, Levenson RW. Emotional behavior in long-term marriage. Psychology and Aging 1995;10:140–149. [PubMed: 7779311]

Christensen A, Hazzard A. Reactive effects during naturalistic observation of families. Behavioral Assessment 1983;5:349–362.

Christensen A, Heavey CL. Gender and social structure in the demand/withdraw pattern of marital conflict. Journal of Personality and Social Psychology 1990;59:73–81. [PubMed: 2213491]

Cohan CL, Bradbury TN. Assessing responses to recurring problems in marriage evaluation of the Marital Coping Inventory. Psychological Assessment 1994;6:191–200.

Cohan CL, Bradbury TN. Negative life events, marital interaction, and the longitudinal course of newlywed marriage. Journal of Personality and Social Psychology 1997;73:114–128. [PubMed: 9216080]

Cohen RS, Christensen A. Further examination of demand characteristics in marital interaction. Journal of Consulting and Clinical Psychology 1980;48:121–123.

Conger RD, Patterson GR, Ge X. A mediational model for the impact of parents' stress on adolescent adjustment. Child Development 1995;66:80–97. [PubMed: 7497831]

Conger RD, Reuter MA, Elder GH. Couple resilience to economic pressure. Journal of Personality and Social Psychology 1999;76:54–71. [PubMed: 9972553]

Cook J, Tyson R, White J, Rushe R, Gottman J, Murray J. Mathematics of marital conflict: Qualitative dynamic mathematical modeling of marital interaction. Journal of Family Psychology 1995;9:110–130.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Cook WW, Medley DM. Proposed hostility and Pharisaic-virtue scales for the MMPI. Journal of Applied Psychology 1954;38:414–418.

Cordova JV, Jacobson NS, Gottman JM, Rushe R, Cox G. Negative reciprocity and communication in couples with a violent husband. Journal of Abnormal Psychology 1993;102:559–564. [PubMed: 8282924]

Cousins PC, Vincent JP. Supportive and aversive behavior following spousal complaints. Journal of Marriage and the Family 1983;45:679–682.

Cutrona, C. E., & Suhr, J. A. (1994). Social support communication in the context of marriage: An analysis of couples' supportive interactions. In B. R. Burleson & T. L. Albrecht (Eds.), *Communication of social support: Messages, interactions, relationships, and community* (pp. 113–135). Newbury Park, CA: Sage.

Davila J, Bradbury TN, Cohan CL, Tochluk S. Marital functioning and depressive symptoms: Evidence for a stress generation model. Journal of Personality and Social Psychology 1997;73:849–861. [PubMed: 9325596]

Deal JE, Hagan MS, Bass B, Hetherington EM, Clingempeel G. Marital interaction in dyadic and triadic contexts: Continuities and discontinuities. Family Process 1999;38:105–115. [PubMed: 10207713]

Dickens, C. (1859). *A tale of two cities.* New York: Vintage Books.

Dillard JP, Fitzpatrick MA. Compliance-gaining in marital interaction. Personality and Social Psychology Bulletin 1985;11:419–433.

Dishion, T. J., Li, F., Spracklen, K. M., Brown, G., & Haas, E. (1998). The measurement of parenting practices in research on adolescent problem behavior: A multimethod and multitrait analysis. In R. S. Ashery, E. B. Robertson, & K. L. Kumpfer (Eds.), *Research on drug SSabuse prevention through family interventions: N1DA Research Monograph, 177* (pp. 260–293). Rockville, MD: National Institute on Drug Abuse.

Eddy JM, Dishion TJ, Stoolmiller M. The analysis of intervention change in children and families: Methodological and conceptual issues embedded in intervention studies. Journal of Abnormal Child Psychology 1998;26:53–69. [PubMed: 9566546]

Eidelson RJ, Epstein N. Cognition and relationship maladjustment: Development of a measure of dysfunctional relationship beliefs. Journal of Consulting and Clinical Psychology 1982;50:715–720. [PubMed: 7142545]

Epstein N, Jackson E. An outcome study of short-term communication training with married couples. Journal of Consulting and Clinical Psychology 1978;46:207–212. [PubMed: 649796]

Escudero V, Rogers LE, Gutierrez E. Patterns of relational control and nonverbal affect in clinic and nonclinic couples. Journal of Social and Personal Relationships 1997;14:5–29.

Ewart CK, Taylor C, Kraemer HC, Agras W. Reducing blood pressure reactivity during interpersonal conflict: Effects of marital communication training. Behavior Therapy 1984;15:473–484.

Ewart CK, Taylor CB, Kraemer HC, Agras WS. High blood pressure and marital discord: Not being nasty matters more than being nice. Health Psychology 1991;10:155–163. [PubMed: 1879387]

Faludi, S. (1991). *Backlash: The undeclared war against American women.* New York: Crown Publishers.

Fals-Stewart W, Birchler GR. Marital interactions of drug-abusing patients and their partners: Comparisons with distressed couples and relationship to drug-using behavior. Psychology of Addictive Behaviors 1998;12:28–38.

Fehm-Wolfsdorf G, Groth T, Kaiser A, Hahlweg K. Cortisol response to marital conflict depend on marital interaction quality. International Journal of Behavioral Medicine 1999;6:207–227. [PubMed: 16250676]

Fichten CS, Wright J. Problem-solving skills in happy and distressed couples: Effects of videotape and verbal feedback. Journal of Clinical Psychology 1983;39:340–352. [PubMed: 6874967]

Filsinger EE. A machine-aided marital observation technique: The Dyadic Interaction Coding System. Journal of Marriage and the Family 1983;45:623–632.

Filsinger EE, Thoma SJ. Behavioral antecedents of relationship stability and adjustment: A five-year longitudinal study. Journal of Marriage and the Family 1988;50:785–795.

Fincham FD, Gamier PC, Gano-Phillips S, Osborne LN. Preinteraction expectations, marital satisfaction, and accessibility: A new look at sentiment override. Journal of Family Psychology 1995;9:3–14.

Fineberg BL, Lowman J. Affect and status dimensions of marital adjustment. Journal of Marriage and the Family 1975;37:155–159.

Fitzpatrick MA, Fallis S, Vance L. Multifunctional coding of conflict resolution strategies in marital dyads. Family Relations 1982;31:61–70.

Floyd FJ. Couples' cognitive/affective reactions to communication behaviors. Journal of Marriage and the Family 1988;50:523–532.

Floyd FJ, Markman HJ. Observational biases in spouse observation: Towards a cognitive/behavioral model of marriage. Journal of Consulting and Clinical Psychology 1983;51:450–457. [PubMed: 6863707]

Floyd FJ, O'Farrell TJ, Goldberg M. Comparison of marital observational measures: The Marital Interaction Coding System and the Communication Skills Test. Journal of Consulting and Clinical Psychology 1987;55:220–237.

Follingstad DR, Haynes SN. Naturalistic observation in assessment of behavioral marital therapy. Psychological Reports 1981;49:471–479. [PubMed: 7302063]

Foster DA, Caplan RD, Howe GW. Representativeness of observed couple interaction: Couples can tell, and it does make a difference. Psychological Assessment 1997;9:285–294.

Foster SL, Cone JD. Validity issues in clinical assessment. Psychological Assessment 1995;7:248–260.

Frankenstein W, Hay WM, Nathan PE. Effects of intoxication on alcoholics' marital communication and problem solving. Journal of Studies on Alcohol 1985;46:1–6. [PubMed: 3974232]

Frosch CA, Mangelsdorf SC, McHale JL. Correlates of marital behavior at six months postpartum. Developmental Psychology 1998;34:1438–1449. [PubMed: 9823523]

Geiss SK, O'Leary KD. Therapist ratings of frequency and severity of marital problems: Implications for research. Journal of Marital and Family Therapy 1981;7:515–520.

Gelles, R.J., & Straus, M.A. (1994). *Physical violence in American families* 1985 (3rd release) [computer file]. Durham, NH: University of New Hampshire Family Research laboratory [producer], 1988. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 1994.

Gill DS, Christensen A, Fincham FD. Predicting marital satisfaction from behavior: Do all roads really lead to Rome? Personal Relationships 1999;6:369–387.

Gingras M, Adams D, Chagnon GJ. Marital enrichment: The contribution of sixteen process variables to the effectiveness of a program. Journal of Sex and Marital Therapy 1983;9:121–136. [PubMed: 6655703]

Gotlib, I. H., & Kowalik, D. L. (1985). *Manual for coding dyadic interactions.* Unpublished manuscript, University of Western Ontario.

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

Gottman JM. Consistency of nonverbal affect and affect reciprocity in marital interaction. Journal of Consulting and Clinical Psychology 1980;48:711–717.

Gottman JM. The roles of conflict engagement, escalation, and avoidance in marital interaction: A longitudinal view of five types of couples. Journal of Consulting and Clinical Psychology 1993;61:6–15. [PubMed: 8450108]

Gottman, J. M. (1994). *What predicts divorce?* Hillsdale, NJ: Erlbaum.

Gottman, J. M. (1996). *What predicts divorce? The measures.* Hillsdale, NJ: Erlbaum.

Gottman JM. Psychology and the study of marital processes. Annual Review of Psychology 1998;49:169–197.

Gottman, J. M. (1999). *The marriage clinic: A scientifically-based marital therapy.* New York: Norton.

Gottman JM, Carrere S, Swanson C, Coan JA. Reply to "From basic research to interventions. Journal of Marriage and the Family 2000;62:265–273.

Gottman JM, Coan J, Carrere S, Swanson C. Predicting marital happiness and stability from newlywed interactions. Journal of Marriage and the Family 1998;60:5–22.

Gottman JM, Jacobson NS, Rushe RH, Shortt JW, Babcock J, Taillade JJ, Waltz J. The relationship between heart rate activity, emotionally aggressive behavior, and general violence in batterers. Journal of Family Psychology 1995;9:227–248.

Gottman JM, Krokoff LJ. Marital interaction and satisfaction: A longitudinal view. Journal of Consulting and Clinical Psychology 1989;57:47–52. [PubMed: 2487031]

Gottman JM, Levenson RW. Marital processes predictive of later dissolution: Behavior, physiology, and health. Journal of Personality and Social Psychology 1992;63:221–233. [PubMed: 1403613]

Gottman JM, Levenson RW. Dysfunctional marital conflict: Women are being unfairly blamed. Journal of Divorce and Remarriage 1999a;31(3–4):1–17.

Gottman JM, Levenson RW. How stable is marital interaction over time? Family Process 1999b;38:143–158. [PubMed: 10407716]

Gottman JM, Levenson RW. Rebound from marital conflict and divorce prediction. Family Process 1999c;38:287–292. [PubMed: 10526766]

Gottman JM, Levenson RW. What predicts change in marital interaction over time? A study of alternative models. Family Process 1999d;38:143–158. [PubMed: 10407716]

Gottman JM, Markman HJ, Notarius CI. The topography of marital conflict: A sequential analysis of verbal and nonverbal behavior. Journal of Marriage and the Family 1977;39:461–477.

Gottman JM, Swanson C, Murray J. The mathematics of marital conflict dynamic mathematical nonlinear modeling of newlywed marital interaction. Journal of Family Psychology 1999;13:3–19.

Gray-Little B, Baucom DH, Hamby SL. Marital power, marital adjustment, and therapy outcome. Journal of Family Psychology 1996;10:292–303.

Greenberg LS, Ford CL, Alden LS, Johnson SM. In-session change in emotionally focused therapy. Journal of Consulting and Clinical Psychology 1993;61:78–84. [PubMed: 8450111]

Greenberg, L. S., & Johnson, S. M. (1988). *Emotionally focused therapy for couples.* New York: Guilford Press.

Haber JR, Jacob T. Marital interactions of male versus female alcoholics. Family Process 1997;36:385–402. [PubMed: 9543660]

Haefner PT, Notarius CI, Pellegrini DS. Determinants of satisfaction with marital discussions: An exploration of husband-wife differences. Behavioral Assessment 1991;13:67–82.

Hahlweg K, Kaiser A, Christensen A, Fehm-Wolfsdorf G, Groth T. Self-report and observational assessment of couples' conflict: The concordance between the Communication Patterns Questionnaire and the KPI Observation System. Journal of Marriage and the Family 2000;62:61–67.

Hahlweg K, Markman HJ, Thurmaier F, Engl J, Eckert V. Prevention of marital distress results of a German prospective longitudinal study. Journal of Family Psychology 1998;12:543–556.

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 KPI. In K. Hahlweg & N. S. Jacobson (Eds.), *Marital interaction: Analysis and modification* (pp. 182–198). New York: Guilford Press.

Hahlweg K, Revenstorf D, Schindler L. Treatment of marital distress: Comparing formats and modalities. Journal of Behavior Research and Therapy 1982;4:57–74.

Hahlweg K, Revenstorf D, Schindler L. Effects of behavioral marital therapy on couples' communication and problem-solving skills. Journal of Consulting and Clinical Psychology 1984;52:553–566. [PubMed: 6470280]

Hahlweg, K., Schindler, L., Revenstorf, D., & Brengelmann, J. C. (1984). The Munich marital therapy study. In K. Hahlweg & N. S. Jacobson (Eds.), *Marital interaction: Analysis and modification* (pp. 3–26). New York: Guilford.

Halford WK, Sanders MR, Behrens BC. A comparison of the generalization of behavioral marital therapy and enhanced behavioral marital therapy. Journal of Consulting and Clinical Psychology 1993;61:51–60. [PubMed: 8450107]

Hammen C, Adrian C, Gordon D, Burge D, Jaenicke C, Hiroto D. Children of depressed mothers: Maternal strain and symptom predictors of dysfunction. Journal of Abnormal Psychology 1987;88:190–198. [PubMed: 3680756]

Hardman RK, Hoopes MH, Harper JM. Verbal interaction styles of two marital combinations: Based on a systems approach to sibling positions. American Journal of Family Therapy 1987;15:131–144.

Harrell, J., & Guerney, B. (1976). Training married couples in conflict negotiation skills. In D. H. Olson (Ed.), *Treating relationships* (pp. 151–165). Iowa: Graphic Pub. Co.

Hautzinger M, Linden M, Hoffman N. Distressed couples with and without a depressed partner: An analysis of their verbal interaction. Behavior Therapy and Experimental Psychiatry 1982;13:307–314.

Hawkins JL, Weisberg C, Ray DL. Marital communication style and social class. Journal of Marriage and the Family 1977;39:479–490.

Haynes, S. N. (1978). *Principles of behavioral assessment.* New York: Gardner Press.

Haynes SN, Chavez RE, Samuel V. Assessment of marital communication and distress. Behavioral Assessment 1984;6:315–321.

Haynes SN, Floyd FJ, Lemsky C, Rogers E, Winemiller D, Heilman N, Werle M, Murphy T, Cardone L. The Marital Satisfaction Questionnaire for older persons. Psychological Assessment 1992;4:473–482.

Haynes SN, Follingstad DR, Sullivan JC. Assessment of marital satisfaction and interaction. Journal of Consulting and Clinical Psychology 1979;47:789–791.

Haynes SN, Jensen BJ, Wise E, Sherman H. The marital intake interview: A multimethod criterion validity assessment. Journal of Consulting and Clinical Psychology 1981;49:379–387. [PubMed: 7276327]

Haynes SN, Richard DCS, Kubany ES. Content validity in psychological assessment: A functional approach to concepts and methods. Psychological Assessment 1995;7:238–247.

Heavey CL, Christensen A, Malamuth NM. The longitudinal impact of demand and withdrawal during marital conflict. Journal of Consulting and Clinical Psychology 1995;63:797–801. [PubMed: 7593873]

Heavey CL, Larson BM, Zumtobel DC. The Communication Patterns Questionnaire: The reliability and validity of a constructive communication subscale. Journal of Marriage and the Family 1996;58:796–800.

Heavey CL, Layne C, Christensen A. Gender and conflict structure in marital interaction: A replication and extension. Journal of Consulting and Clinical Psychology 1993;61:16–27. [PubMed: 8450102]

Heyman RE, Brown PD, Feldbau-Kohn SR, O'Leary KD. Couples' communication variables as predictors of dropout and treatment response in wife abuse treatment programs. Behavior Therapy 1999;30:165–190.

Heyman, R. E., Chaudhry, B. R., Treboux, D., Crowell, J., Lord, C., Vivian, D., & Waters, E. B. (in press). How much observational data is enough? An empirical test using marital interaction coding. *Behavior Therapy*.

Heyman RE, Eddy JM, Weiss RL, Vivian D. Factor analysis of the Marital Interaction Coding System. Journal of Family Psychology 1995;9:209–215.

Heyman, R. E., Feldbau-Kohn, S. R., Ehrensaft, M. K., Langhinrichsen-Rohling, J., & O'Leary, K. D. (in press). Can questionnaire reports correctly classify relationship distress and partner physical abuse? *Journal of Family Psychology*.

Heyman, R. E., & Slep, A. M. S. (1999). *Development of algorithms for estimating family violence rates in Air Force communities.* Unpublished manuscript, State University of New York at Stony Brook.

Heyman, R. E., & Slep, A. M. S. (in press). The hazards of predicting divorce without crossvalidation. *Journal of Marriage and the Family.*

Heyman, R. E., & Vivian, D. (1993). *RMICS: Rapid Marital Interaction Coding System: Training manual for coders.* Unpublished manuscript, State University of New York, Stony Brook, NY. (Available at http://www.psy.sunysb.edu/marital)

Heyman RE, Weiss RL, Eddy JM. Marital Interaction Coding System: Revision and empirical evaluation. Behavioral Research and Therapy 1995;33:737–746.

Holtzworth-Munroe A, Smutzler N, Stuart GL. Demand and withdraw communication among couples experiencing husband violence. Journal of Consulting and Clinical Psychology 1998;66:731–743. [PubMed: 9803691]

Holtzworth-Munroe A, Stuart GL, Sandin E, Smutlzer N, McLaughlin W. Comparing the social support behaviors of violent and nonviolent husbands during discussions of wife personal problems. Personal Relationships 1997;4:395–407.

Hooley JM. Expressed emotion and depression: Interactions between patients and high-versus low-expressed-emotion spouses. Journal of Abnormal Psychology 1986;95:237–246. [PubMed: 3745645]

Hooley JM, Hahlweg K. Marital satisfaction and marital communication in German and English couples. Behavioral Assessment 1989;11:119–133.

Iverson A, Baucom DH. Behavioral marital therapy outcomes: Alternate interpretations of the data. Behavior Therapy 1990;21:129–138.

Jacob T, Krahn GL. The classification of behavioral observation codes in studies of family interaction. Journal of Marriage and the Family 1987;49:677–687.

Jacob T, Krahn GL. Marital interactions of alcoholic couples: Comparison with depressed and nondistressed couples. Journal of Consulting and Clinical Psychology 1988;56:73–79. [PubMed: 3346452]

Jacob T, Leonard KE. Alcoholic-spouse interaction as a function of alcoholism subtype and alcohol consumption. Journal of Abnormal Psychology 1988;97:231–237. [PubMed: 3385076]

Jacob T, Leonard KE. Sequential analysis of marital interactions involving alcoholic, depressed, and nondistressed men. Journal of Abnormal Psychology 1992;101:647–656. [PubMed: 1430603]

Jacob T, Seilhamer RA, Rushe RH. Alcoholism and family interaction: An experimental paradigm. American Journal of Drug and Alcohol Abuse 1989;15:73–91. [PubMed: 2923113]

Jacob T, Tennenbaum D, Seilhamer RA, Bargiel K, Sharon T. Reactivity effects during naturalistic observation of distressed and nondistressed families. Journal of Family Psychology 1994;8:354–363.

Jacobson NS. Problem solving and contingency contracting in the treatment of marital discord. Journal of Consulting and Clinical Psychology 1977;45:92–100. [PubMed: 845311]

Jacobson NS. Specific and nonspecific factors in the effectiveness of a behavioral approach to the treatment of marital discord. Journal of Consulting and Clinical Psychology 1978;46:442–452. [PubMed: 670488]

Jacobson NS. The role of observational measures in behavioral therapy outcome research. Behavioral Assessment 1985;7:297–308.

Jacobson NS, Anderson EA. The effects of behavior rehearsal and feedback on the acquisition of problem-solving skills in distressed and nondistressed couples. Behaviour Research and Therapy 1980;18:25–36. [PubMed: 7369985]

Jacobson, N. S., & Gottman, J. M. (1998). *When men batter women: New insights into ending abusive relationships.* New York: Simon & Schuster.

Jacobson NS, Gottman JM, Gortner E, Berns S, Shortt JW. Psychological factors in the longitudinal course of battering: when do the couples split up? When does the abuse decrease? Violence and Victims 1996;11:371–392. [PubMed: 9210278]

Jacobson NS, Gottman JM, Waltz J, Rushe R, Babcock J, Holtzworth-Monroe A. Affect, verbal content, and psychophysiology in the arguments of couples with a violent husband. Journal of Consulting and Clinical Psychology 1994;62:982–988. [PubMed: 7806730]

Jacobson, N. S., & Gurman, A. S. (Eds.). (1995). *Clinical handbook of couples therapy.* New York: Guilford.

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

Jacobson NS, Schmaling KB, Holtzworth-Munroe A, Katt JL, Wood LF, Follette VM. Research-structured vs. clinically flexible versions of social learning-based marital therapy. Behaviour Research and Therapy 1989;27:173–180. [PubMed: 2930443]

Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology 1991;59:12–19. [PubMed: 2002127]

Johnson MD, Bradbury TN. Marital satisfaction and topographical assessment of marital interaction: A longitudinal analysis of newlywed couples. Personal Relationships 1999;6:19–40.

Johnson SL, Jacob T. Marital interactions of depressed men and women. Journal of Consulting and Clinical Psychology 1997;65:15–23. [PubMed: 9103730]

Johnson SL, Jacob T. Sequential interactions in the marital communication of depressed men and women. Journal of Consulting and Clinical Psychology 2000;68:4–12. [PubMed: 10710835]

Julien D, Markman H, Lindahl K. A comparison of global and a microanalytic coding system: Implications for future trends in studying interactions. Behavioral Assessment 1989;11:81–100.

Kaiser A, Hahlweg K, Fehm-Wolfsdorf G, Groth T. The efficacy of a compact psychoeducational group training program for married couples. Journal of Consulting and Clinical Psychology 1998;66:753–760. [PubMed: 9803693]

Karney BR, Bradbury TN. Neuroticism, marital interaction, and the trajectory of marital satisfaction. Journal of Personality and Social Psychology 1997;72:1075–1092. [PubMed: 9150586]

Katz LF, Gottman JM. Patterns of marital conflict predict children's internalizing and externalizing behaviors. Developmental Psychology 1993;29:940–950.

Keppel, G. (1991). Design and analysis: A researcher's handbook (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Kiecolt-Glaser JK, Malarkey WB, Chee M, Newton T, Cacioppo JT, Mao H, Glaser R. Negative behavior during marital conflict is associated with immunological down-regulation. Psychosomatic Medicine 1993;55:395–409. [PubMed: 8265740]

Kiecolt-Glaser JK, Newton T, Cacioppo JY, MacCallum RC, Glaser R, Malarkey WB. Marital conflict and endocrine function: Are men really more physiologically affected than women? Journal of Consulting and Clinical Psychology 1996;64:324–332. [PubMed: 8871417]

Kimmel D, VanderVeen F. Factors of marital adjustment in Locke's Marital Adjustment Test. Journal of Marriage and the Family 1974;36:57–63.

Klinetob NA, Smith DA. Demand-withdraw communication in marital interaction: Tests of interpersonal contingency and gender role hypotheses. Journal of Marriage and the Family 1996;58:945–957.

Knox, D. (1971). Marriage happiness: A behavioral approach to counseling. Champaign, IL: Research Press.

Kobak RR, Hazan C. Attachment in marriage: Effects of security and accuracy of working models. Journal of Personality and Social Psychology 1991;60:861–869. [PubMed: 1865324]

Koren H, Carlton K, Shaw H. Marital conflict: Relations among behaviors, outcomes, and distress. Journal of Consulting and Clinical Psychology 1980;48:460–468. [PubMed: 7400431]

Kowalik DL, Gotlib IH. Depression and marital interaction: Concordance between intent and perception of communication. Journal of Abnormal Psychology 1987; 96:127–134. [PubMed: 3584661]

Krokoff LJ. Recruiting representative samples for marital interaction research. Journal of Social and Personal Relationships 1987;4:317–328.

Krokoff LJ, Gottman JM, Hass SD. Validation of a global Rapid Couples Interaction Scoring System. Behavioral Assessment 1989;11:65–79.

Krokoff LJ, Gottman JM, Roy AK. Blue-collar and white-collar marital interaction and communication orientation. Journal of Social and Personal Relationships 1988;5:201–221.

Krueger DL, Smith P. Decision-making patterns of couples: A sequential analysis. Journal of Communication 1982;32:121–133.

Kuhn, T. S. (1970). The structure of scientific revolutions (2nd ed.). Chicago: The University of Chicago Press.

Leonard KE, Jacob T. Sequential interactions among episodic and steady alcoholics and their wives. Psychology of Addictive Behaviors 1997;11:18–25.

Leonard KE, Roberts LJ. The effects of alcohol on the marital interactions of aggressive and nonaggressive husbands and their wives. Journal of Abnormal Psychology 1998;107:602–615. [PubMed: 9830248]

Levenson RW, Gottman JM. Marital interaction: Physiological linkage and affective exchange. Journal of Personality and Social Psychology 1983;45:587–597. [PubMed: 6620126]

Liberman RP, Levine J, Wheeler E, Sanders N, Wallace CJ. Marital therapy in groups: A comparative evaluation of behavioral and interactional formats. Acta Psychiatrica Scandinavica 1976;266:1–34.

Lindahl KM, Clements M, Markman H. Predicting marital and parent functioning in dyads and triads: A longitudinal investigation of marital processes. Journal of Family Psychology 1997;11:139–151.

Linden N, Hautzinger M, Hoffmann N. Discriminant analysis of depressive interactions. Behavior Modification 1983;7:403–422. [PubMed: 6679304]

Locke HJ, Wallace KM. Short marital adjustment and prediction tests: Their reliability and validity. Marriage and Family Living 1959;21:251–255.

Lord, C. C. (1999). *Stability and change on interactional behavior in early marriage.* Unpublished doctoral dissertation, State University of New York, Stony Brook, NY.

Margolin G. A multilevel approach to the assessment of communication positiveness in distressed marital couples. American Journal of Family Therapy 1978a;6:81–89.

Margolin G. Relationships among marital assessment procedures: A correlational study. Journal of Consulting and Clinical Psychology 1978b;46:1556–1558.

Margolin G, Burman B, John RS. Home observations of married couples reenacting naturalistic conflicts. Behavioral Assessment 1989;11:101–118.

Margolin G, Hattem D, John RS, Yost K. Perceptual agreement between spouses and outside observers when coding themselves and a stranger dyad. Behavioral Assessment 1985;7:235–247.

Margolin G, John RS, Gleberman L. Affective responses to conflictual discussion in violent and nonviolent couples. Journal of Consulting and Clinical Psychology 1988;56:24–33. [PubMed: 3346445]

Margolin G, Wampold BE. Sequential analysis of conflict and accord in distressed and nondistressed marital partners. Journal of Consulting and Clinical Psychology 1981;49:554–567. [PubMed: 7264037]

Margolin G, Weiss RL. Communication training and assessment: A case of behavioral marital enrichment. Behavior Therapy 1978a;9:508–520.

Margolin G, Weiss RL. Comparative evaluation of therapeutic components associated with behavioral marital treatments. Journal of Consulting and Clinical Psychology 1978b;46:1476–1486. [PubMed: 730900]

Markman, H. J., Duncan, S. W., Storaasli, R. D., & Howes, P. (1987). The prediction and prevention of marital distress: A longitudinal investigation. In K. Halweg & M. Goldstein (Eds.), *Understanding major mental disorder: The contribution of family interaction research* (pp. 266–289). New York: Family Process.

Markman HJ, Floyd FJ, Stanley SM, Storaasli RD. Prevention of marital distress: A longitudinal investigation. Journal of Consulting and Clinical Psychology 1988;56:210–217. [PubMed: 3372828]

Markman HJ, Leber BD, Cordova AD, St Peters M. Behavioral observation and family psychology: Strange bedfellows or happy marriage?: Comment on Alexander et al. (1995). Journal of Family Psychology 1995;9:371–379.

Markman, H. J., & Notarius, C. I. (1987). Coding marital and family interaction: Current status. In T. Jacob (Ed.), *Family interaction and psychopathology: Theories, methods, and findings* (pp. 329–390). New York: Plenum Press.

Markman HJ, Renick MJ, Floyd FJ, Stanley SM, Clements M. Preventing marital distress through communication and conflict management training: A 4- and 5-year follow-up. Journal of Consulting and Clinical Psychology 1993;61:70–77. [PubMed: 8450110]

Markman HJ, Silvern L, Clements M, Kraft-Hanak S. Men and women dealing with conflict in heterosexual relationships. Journal of Social Issues 1993;49:107–125.

Martin R. Relational cognition complexity and relational communication in personal relationships. Communication Monographs 1992;59:150–163.

Matthews LS, Wickrama KAS, Conger RD. Predicting marital instability from spouse and observer reports of marital interaction. Journal of Marriage and the Family 1996;58:641–655.

McKnight DL, Nelson-Gray RO, Gullick E. Interactional patterns of bipolar patients and their spouses. Journal of Psychopathology and Behavioral Assessment 1989;11:269–289.

Melby JN, Conger RD, Ge X, Warner TD. The use of structural equation modeling in assessing the quality of marital observations. Journal of Family Psychology 1995;9:280–293.

Mendolia M, Beach SRH, Tesser A. The relationship between marital interaction behaviors and affective reactions to one's own and one's spouse's self-evaluation needs. Personal Relationships 1996;3:279–292.

Miller GE, Bradbury TN. Refining the association between attributions and behavior in marital interaction. Journal of Family Psychology 1995;9:196–208.

Miller GE, Dopp JM, Myers HF, Stevens SY, Fahey JL. Psychosocial predictors of natural killer cell mobilization during marital conflict. Health Psychology 1999;18:262–271. [PubMed: 10357507]

Murphy CM, O'Farrell TJ. Couple communication patterns of maritally aggressive and nonaggressive male alcoholics. Journal of Studies on Alcohol 1997;58:83–90. [PubMed: 8979216]

Nelson GM, Beach SRH. Sequential interaction in depression: Effects of depressive behavior on spousal aggression. Behavior Therapy 1990;21:167–182.

Newton DA, Burgoon JK. The use and consequences of verbal influence strategies during interpersonal disagreements. Human Communication Research 1990;16:477–518.

Newton TL, Kiecolt-Glaser JK, Glaser R, Malarkey WB. Conflict and withdrawal during marital interaction: The roles of hostility and defensiveness. Personality & Social Psychology Bulletin 1995;21:512–524.

Noel NE, McCrady BS, Stout RL, Fisher-Nelson H. Gender differences in marital functioning of male and female alcoholics. Family Dynamics Addiction Quarterly 1991;1:31–38.

Noller P. Channel consistency and inconsistency in the communications of married couples. Journal of Personality and Social Psychology 1982;1982:432–741.

Notarius CI, Benson PR, Sloane D, Vanzetti NA, Hornyak LM. Exploring the interface between perception and behavior: An analysis of marital interaction in distressed and nondistressed couples. Behavioral Assessment 1989;11:39–64.

Notarius CI, Johnson JS. Emotional expression in husbands and wives. Journal of Marriage and the Family 1982;44:483–489.

Notarius CI, Markman HJ. Coding marital interaction: A sampling and discussion of current issues. Behavioral Assessment 1989a;11:1–11.

Notarius, C. I., & Markman, H. J. (Eds.). (1989b). Coding marital interaction [Special issue]. *Behavioral Assessment, 11*.

O'Farrell TJ, Birchler GR. Marital relationships of alcoholic, conflicted, and nonconflicted couples. Journal of Marital and Family Therapy 1987;13:259–274.

O'Farrell TJ, Cutter HSG, Floyd FJ. Evaluating behavioral marital therapy for male alcoholics: Effects on marital adjustment and communication from before to after treatment. Behavior Therapy 1985;16:147–167.

O'Leary, K. D., Heyman, R. E., & Jongsma, A. E. (1998). *The couples therapy treatment planner*. New York: Wiley.

O'Leary KD, Vivian D, Malone J. Assessment of physical aggression in marriage: The need for a multimodal method. Behavioral Assessment 1992;14:5–14.

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

Paley B, Cox MJ, Burchinal MR, Payne CC. Attachment and marital functioning: Comparison of spouses with continuous-secure, earned-secure, dismissing, and preoccupied attachment stances. Journal of Family Psychology 1999;13:580–597.

Pasch LA, Bradbury TN. Social support, conflict, and the development of marital dysfunction. Journal of Consulting and Clinical Psychology 1998;66:219–230. [PubMed: 9583325]

Pasch LA, Bradbury TN, Davila J. Gender, negative affectivity, and observed social support behavior in marital interaction. Personal Relationships 1997;4:361–378.

Pasupathi M, Carstensen LL, Levenson RW, Gottman JM. Responsive listening in long-married couples: A psycholinguistic perspective. Journal of Nonverbal Behavior 1999;23:173–193.

Patterson, G. R. (1982). *Coercive family process.* Eugene, OR: Castalia.

Patterson GR. Orderly change in a stable world: The antisocial trait as a chimera. Journal of Consulting and Clinical Psychology 1993;61:911–919. [PubMed: 8113492]

Patterson, G. R., & Hops, H. (1972). Coercion, a game for two: Intervention techniques for marital conflict. In R. E. Ulrich & P. Mountjoy (Eds.), *The experimental analysis of social behavior* (pp. 424–440). New York: Appleton-Century-Crofts.

Patterson GR, Hops H, Weiss RL. Interpersonal skills training for couples in early stages of conflict. Journal of Marriage and the Family 1975;37:295–302.

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys.* Eugene, OR: Castalia.

Paul GL. Strategy of outcome research in psychotherapy. Journal of Consulting Psychology 1967;31:109–118. [PubMed: 5342732]

Persons, J. B. (1989). *Cognitive therapy in practice: A case formulation approach.* New York: Norton.

Raush, H. L., Barry, W. A., Hertel, R. K., & Swain, M. E. (1974). *Communication, conflict and marriage.* San Francisco: Jossey-Bass.

Revenstorf, D., Hahlweg, K., Schindler, L., & Vogel, B. (1984). Interaction analysis of marital conflict. In K. Hahlweg & N. S. Jacobson (Eds.), *Marital interaction: Analysis and modification* (pp. 159–181). New York: Guilford Press.

Revenstorf D, Vogel B, Wegener C, Hahlweg K, Schindler L. Escalation phenomena in interaction sequences: An empirical comparison of distressed and non-distressed couples. Behavior Analysis and Modification 1980;4:97–115.

Roberts LJ, Krokoff LJ. A time-series analysis of withdrawal, hostility, and displeasure in satisfied and dissatisfied marriages. Journal of Marriage and the Family 1990;52:95–105.

Roberts, L. J., & Linney, K. D. (in press). Observing intimacy process behavior: Vulnerability and partner responsiveness in marital interaction. *Journal of Personality and Social Psychology.*

Roberts MCF, Floyd FJ, O'Farrell TJ, Cutter HSG. Marital interactions and the duration of alcoholic husbands' sobriety. Journal of Drug and Alcohol Abuse 1985;11:303–313.

Robey, E. B., Canary, D. J., & Burggraf, C. S. (1998). Conversational maintenance behaviors of husbands and wives: An observational analysis. In D. J. Canary & K. Dindia (Ed.), *Sex differences and similarities in communication: Critical essays and empirical investigations of sex and gender in interaction,* (pp. 373–392). Hillsdale, NJ: Erlbaum.

Robinson EA, Price MG. Pleasurable behavior in marital interaction: An observational study. Journal of Consulting and Clinical Psychology 1980;48:117–118.

Rogers, L. E., Castleton, A., & Lloyd, S. A. (1996). Relational control and physical aggression in satisfying marital relationships. In D. D. Cahn & S. A. Lloyd (Eds.), *Family violence from a communication perspective* (pp. 218–239). Thousand Oaks, CA: Sage.

Rogge RD, Bradbury TN. Til violence does us part: the differing roles of communication and aggression in predicting adverse marital outcomes. Journal of Consulting and Clinical Psychology 1999;67:340–351. [PubMed: 10369054]

Royce WS, Weiss RL. Behavioral cues in the judgment of marital satisfaction: A linear regression analysis. Journal of Consulting and Clinical Psychology 1975;43:816–824.

Ruscher SM, Gotlib IH. Marital interaction patterns of couples with and without a depressed partner. Behavior Therapy 1988;19:455–470.

Sagrestano LM, Christensen A, Heavey CL. Social influence techniques during marital conflict. Personal Relationships 1998;5:75–89.

Sagrestano LM, Heavey CL, Christensen A. Perceived power and physical violence in marital conflict. Journal of Social Issues 1999;55:65–79.

Sanders MR, Halford WK, Behrens BC. Parental divorce and premarital couple communication. Journal of Family Psychology 1999;13:60–74.

Sayers SL, Baucom DH. Role of femininity and masculinity in distressed couples' communication. Journal of Personality and Social Psychology 1991;61:641–647. [PubMed: 1960655]

Sayers SL, Baucom DH, Sher TG, Weiss RL, Heyman RE. Constructive engagement, behavioral marital therapy, and changes in marital satisfaction. Behavioral Assessment 1991;13:25–49.

Schaap, C. (1984). A comparison of the interaction of distressed and nondistressed married couples in a laboratory situation: Literature survey, methodological issues, and an empirical investigation. In K. Hahlweg, & N. S. Jacobson (Ed.), *Marital interaction: Analysis and modification* (pp. 133–158). New York: Guilford.

Schaap C, Jansen-Nawas C. Marital interaction, affect and conflict resolution. Sexual and Marital Therapy 1987;2:35–51.

Schafer J, Birchler GR, Fals-Stewart W. Cognitive, affective, and marital functioning of recovering male polysubstance abusers. Neuropsychology 1994;8:100–109.

Schmaling KB, Jacobson NS. Marital interaction and depression. Journal of Abnormal Psychology 1990;99:229–236. [PubMed: 2212272]

Schmaling KB, Waboldt F, Telford L, Newman KB, Hops H, Eddy JM. Interaction of asthmatics and their spouses: A preliminary study of individual differences. Journal of Clinical Psychology in Medical Settings 1996;3:211–218.

Schumm WR, Paff-Bergen LA, Hatch RC, Obiorah FC, Copeland JM, Meens LD, Bugaighis MA. Concurrent and discriminant validity of the Kansas Marital Satisfaction Scale. Journal of Marriage and the Family 1986;48:381–387.

Sher TG, Baucom DH. Marital communication: Differences among maritally distressed, depressed, and nondistressed-nondepressed couples. Journal of Family Psychology 1993;7:148–153.

Sher TG, Baucom DH, Larus JM. Communication patterns and response to treatment among depressed and nondepressed maritally distressed couples. Journal of Family Psychology 1990;4:63–79.

Shoham V, Rohrbaugh MJ, Stickle TR, Jacob T. Demand-withdraw couple interaction moderates retention in cognitive-behavioral versus family-systems treatments for alcoholism. Journal of Family Psychology 1998;12:557–577.

Siegal JM. The Multidimensional Anger Inventory. Journal of Personality and Social Psychology 1986;51:191–200. [PubMed: 3735067]

Smith DA, Vivian D, O'Leary KD. Longitudinal prediction of marital discord from premarital expressions of affect. Journal of Consulting and Clinical Psychology 1990;58:790–798. [PubMed: 2292628]

Snyder, D. K. (1981). *Marital Satisfaction Inventory manual.* Los Angeles: Western Psychological Services.

Snyder DK, Mangrum LF, Wills RM. Predicting couples' response to marital therapy: A comparison of short- and long-term predictors. Journal of Consulting and Clinical Psychology 1993;61:61–69. [PubMed: 8450109]

Snyder DK, Trull TJ, Wills RM. Convergent validity of observational and self-report measures of marital interaction. Journal of Sex and Marital Therapy 1987;13:224–268. [PubMed: 3669081]

Snyder DK, Wills RM. Behavioral versus insight-oriented marital therapy: Effects on individual and interspousal functioning. Journal of Consulting and Clinical Psychology 1989;57:39–46. [PubMed: 2925973]

Snyder DK, Wills RM, Grady-Fletcher A. Long-term effectiveness of behavioral versus insight-oriented marital therapy: A 4-year follow-up study. Journal of Consulting and Clinical Psychology 1991;59:138–141. [PubMed: 2002129]

Spanier GB. Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. Journal of Marriage and the Family 1976;38:15–28.

Stanley SM, Bradbury TN, Markman HJ. Structural flaws in the bridge from basic research on marriage to interventions for couples: Illustrations from Gottman, Coan, Carrere, and Swanson (1998). Journal of Marriage and the Family 2000;62:256–264.

Stein SJ, Girodo M, Dotzenroth S. The interrelationships and reliability of a multilevel behavior-based assessment package for distressed couples. Journal of Behavioral Assessment 1982;4:343–360.

Sternberg, R. J., & Barnes, M. L. (Eds.) (1988). *The psychology of love.* New Haven, CT: Yale University Press.

Storaasli RD, Markman HJ. Relationship problems in the early stages of marriage: A longitudinal investigation. Journal of Family Psychology 1990;4:80–98.

Straus MA. Measuring intra family conflict and violence: The Conflict Tactics (CT) Scales. Journal of Marriage and the Family 1979;41:75–88.

Straus, M. A., & Gelles, R. J. (1990). *Physical violence in American Families.* New Brunswick, NJ: Transaction Publishers.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics.* New York: Harper Collins.

Terman, L. M. (1938). *Psychological factors in marital happiness.* New York: McGraw-Hill.

Thomas EJ, Walter CL, O'Flaherty K. A verbal problem checklist for use in assessing family verbal behavior. Behavior Therapy 1974;5:235–246.

Thomsen DG, Gilbert DG. Factors characterizing marital conflict states and traits: Physiological, affective, behavioral, and neurotic variable contributions to marital conflict and satisfaction. Personality and Individual Differences 1998;25:833–855.

Ting-Toomey S. An analysis of verbal communication patterns in high and low marital adjustment groups. Human Communication Research 1983;9:306–319.

Turgeon L, Julien N, Dion E. Temporal linkages between wives' pursuit and husbands' withdrawal during marital conflict. Family Process 1998;37:323–334. [PubMed: 9879002]

Vincent JP, Friedman LC, Nugent J, Messerly L. Demand characteristics in observations of marital interaction. Journal of Consulting and Clinical Psychology 1979;47:557–566. [PubMed: 528723]

Vincent JP, Weiss RL, Birchler GR. A behavioral analysis of problem solving in distressed and nondistressed married and stranger dyads. Behavior Therapy 1975;6:475–487.

Walker JG, Johnson S, Manion I, Cloutier P. Emotionally Focused Marital Intervention for couples with chronically ill children. Journal of Consulting and Clinical Psychology 1996;64:1029–1036. [PubMed: 8916632]

Walsh VL, Baucom DH, Tyler S, Sayers SL. Impact of message valence, focus, expressive style, and gender on communication patterns among maritally distressed couples. Journal of Family Psychology 1993;7:163–175.

Waters E. The reliability and stability of individual differences in infant-mother attachment. Child Development 1978;49:483–494.

Webster-Stratton C, Hammond M. Marital conflict management skills, parenting style, and early-onset conduct problems: Processes and pathways. Journal of Child Psychology and Psychiatry and Allied Disciplines 1999;40:917–927.

Weiss, R. L. (1980). Strategic behavioral marital therapy: Toward a model for assessment and intervention. In J. P. Vincent (Ed.), *Advances in family intervention, assessment and theory* (Vol. 1, pp. 229–271). Greenwich, CT: JAI Press.

Weiss RL. The circle of voyeurs: Observing the observers of marital and family interactions. Behavioral Assessment 1989;11:135–147.

Weiss RL, Cerreto MC. The Marital Status Inventory: Development of a measure of dissolution potential. American Journal of Family Therapy 1980;8:80–86.

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.

Weiss, R. L., & Heyman, R. E. (1997). Couple interaction. In W. K. Halford & H. J. Markman (Eds.), *Clinical handbook of marriage and couples intervention* (pp. 13–41). New York: Wiley.

Weiss, R. L., Hops, H., & Patterson, G. R. (1973). A framework for conceptualizing marital conflict: A technology for altering it, some data for evaluating it. In L. D. Handy & E. L. Mash (Eds.), *Behavior change: Methodology concepts and practice* (pp. 309–342). Champaign, IL: Research Press.

Weiss RL, Tolman AO. The Marital Interaction Coding System-Global (MICS-G): A global companion to the MICS. Behavioral Assessment 1990;12:271–294.

Wieder GB, Weiss RL. Generalizability theory and the coding of marital interactions. Journal of Consulting and Clinical Psychology 1980;48:469–477. [PubMed: 7400432]

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment.* Malabar, FL: Robert E. Krieger Publishing Company.

Williamson RN, Fitzpatrick MA. Two approaches to marital interaction: Relational control patterns in marital types. Communication Monographs 1985;52:236–252.

Wilson GL, Bornstein PH, Wilson LJ. Treatment of relationship dysfunction: An empirical evaluation of group and conjoint behavioral marital therapy. Journal of Consulting and Clinical Psychology 1988;56:929–931. [PubMed: 3204205]

Witkin SL, Rose SD. Group training and communication skills for couples: A preliminary report. International Journal of Family Counseling 1978;6:45–56.

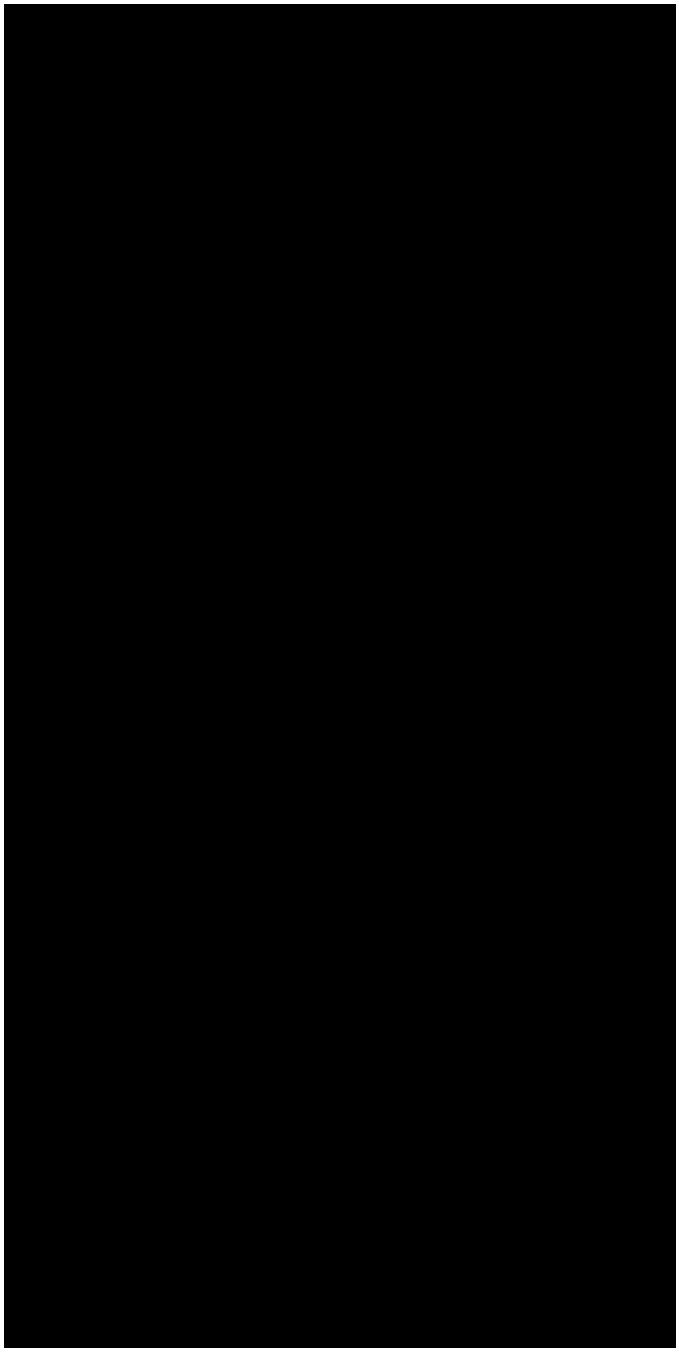

Figure 1.
Flowchart for deriving individual study-relevant reliability and validity data from Table 1. SPAFF = Specific Affect Coding System; RMICS = Rapid Marital Interaction Coding System; KPI = Kategorien system für Partnerschaftliche Interaktion.

Table 1

Description of Study Methods and Psychometrics for Published, Validity Related Couples Observation Studies (Abridged)

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
Adam & Gingras (1982)	Type: T 38 married couples (19 T, 19 WLCG): C[1,2]	Setting: Lab Situation: 12-min. 1[b,c]	MICS	Reliability[h] Validity (Tx sensitivity): Full details on WWW. (See table note.)
Andrews, Foster, Capaldi, & Hops (2000)	Type: L 716 adolescents and their parents at Tl: C-letters sent to parents of 9–12th graders at local schools T3: 314 couples at 6-year follow-up (153 married, 161 dating for at least 1 month)	Setting: Lab Situation: Two 10-min. 1[b] 153 couples did interactions	LIFE	Reliability: (ICC): Participant Aversive (.81), Partner Aversive[l] (.83), Participant Facilitative (.81), Partner Facilitative (.82). Validity (concurrent, postdictive): Full details on WWW
Aron, Norman, Aron, McKenna, & Heyman (2000)	Type: C 35 married couples: C[1,3]	Setting: Lab Situation: (1) 5-min. 6[b] (2) 5-min. 7* Topic: Planning vacation; home improvements. *N* = 26 couples coded	RMICS	Reliability (kappa): Acceptance (.64), Hostility (.66). Validity[k]
Basco, Birchler, Kalal, Talbott, & Slater (1991)	Type: A 36 randomly selected videotapes 10 ND and 24 D couples: C[1], Cl[1]	Setting: Lab Situation: 10-min. 1[b,c]	CRAC, MICS-II	Reliability CRAC (internal consistency): Involvement (.44), Communication Skills (.94), Abusiveness (.89), Problem-Solving Skills (.93), Attribution of Blame (.85), Reliability CRAC (test-retest): Reliability (ICC): Involvement (.86), Communication Skill (.93), Abusiveness (.93), Problem-Solving Skill (.93), Attribution of Blame (.61). Reliability MICS-III[f] Validity CRAC and MICS-III (convergent): Full details on WWW.
Basco, Prager, Pita, Tamir, & Stephens (1992)	Type: C 34 married couples: (17 depressed*, 17 control): C[3,7], Cl[2]; *At least 1 spouse (13 wives and 4 husbands) met *DSM–III* criteria for MDD and Hamilton Rating Scale for Depression ≥ 14	Setting: Lab Situation: 8-min. 1[b]	CRAC	Reliability[f] Validity (discriminative): Full details on WWW
Baucom (1982)	Type: T 72 D couples (randomly assigned to BCT and WLCG): C[1,2]	Setting: Lab Situation: (1) 5–6 min. 1[c] (2) 5–6 min. 5[b]	MICS	Reliability[f] Validity (Tx sensitivity): Full details on WWW
Baucom, Sayers, & Sher (1990)	Type: T 60 D couples: Cl[1] Randomly assigned to BCT, CR + BCT, BCT + EET, CR + BCT + EET, WLCG	Setting: Lab Situation: Two 7-min. 1[b]	MICS-III	Reliability[i] Validity (Tx sensitivity): Full details on WWW
Behrens, Sanders, & Halford (1990)	Type: T 4 D couples (assigned to BCT + Cognitive BCT): Cl[1]	Setting: Lab and home. Situation: Lab: unspecified 1[b] Home: Two unspecified 1[b]	KPI (adapted), Home Audiotaped	Reliability[f] Validity (Tx sensitivity) Full details on WWW
Berns, Jacobson, & Gottman (1999a)[7]	Type: C 91 couples (47 PV, 28 D/NV, 16 ND/NV): C[1,2,4,8] PV: At least 6 moderate acts, 2 severe acts, or 1	Setting: Lab Situation: 15-min. 1[b]	CRS (adapted), SPAFF	Reliability SPAFF[g] Reliability CRS *(M* ICC): Demand (.82), Withdraw (.81), Pos. Communication (.84) Neg. Communication (.85).

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
	life threatening act in the past year			Validity CRS (concurrent, discriminative): Full details on WWW. Validity SPAFF and CRS (convergent): Full details on WWW
Berns, Jacobson, & Gottman (1999b)[7]	Type: C 47 PV couples: C[1,2,4,8]. See (1999a) for PV inclusion criteria	Setting: Lab Situation: 15-min. 1[b]	CRS (adapted)	Reliability[f] Validity (discriminative): Full details on WWW
Biglan, Hops, Sherman, Friedman, Arthur, & Osteen (1985)	Type: C 52 couples (27 depressed* W, 25 nondepressed): C[1], Cl[1,2]	Setting: Lab Situation: Two 10-min. 1[b,c]	LIFE	Reliability (*rs* between observers): Depressive (H = .46, W = .66), Aversive (H = .96; W = .94), Facilitative (H = .87, W = .92), Propose Solution (H = .73, W = .77), Self-Disclosure (H = .65, W = .49), Elicit Response (H = .88, W = .81), Other (H = .61, W = .56). Validity (discriminative): Full details on WWW
Birchler, Clopton, & Adams (1984)	Type: C 28 ND and 28 D couples: C[1], Cl[1]	Setting: Lab Situation: 10-min. 1[b,c]	MICS	Reliability[g] Validity (discriminative): Full details on WWW
Birchler, Weiss, & Vincent (1975)[13]	Type: C 12 ND and 12 D couples: C[1]	Setting: Lab Situation: (1) 4-min. 3[c] (2) 10-min. 5[b]	MICS	Reliability[f] Validity (discriminative): Full details on WWW
Boelens, Emmelkamp, MacGillavry, & Markvoort (1980)	Type: T 21 couples living together (8 reciprocity counseling, 8 systemic-theoretic, 5 no T): Cl[1]	Setting: Lab Situation: 20-min. 5[b]	MICS	Reliability[i] Validity[j]
Bradbury & Fincham (1992)[15]	Type: C Study 1: 47 married couples (27 "mildly dissatisfied to mildly satisfied": C[4] Study 2: 40 married couples (29 nonclinic, 11 clinic): C[1], Cl[1]	Setting: Lab Situation: 15-min. 1[b]	Study 1: Developed for study Study 2: VTCS (adapted)	Reliability (developed for study)[g]; Validity (developed for study)[k] Reliability (VTCS)[f] Validity VTCS (concurrent, discriminative): Full details on WWW
Bradbury & Fincham (1993)[15]	Type: C 43 married couples (32 nonclinic, 11 clinic): C[1], Cl[1]	Setting: Lab Situation: 15-min. 1[b]	VTCS	Reliability[f] Validity (concurrent, discriminative): Full details on WWW
Bradbury, Beach, Fincham, & Nelson (1996)	Type: C 52 married couples (19 ND, 13 D/ Nondepressed*, 20 D/W Depressed): C[1] *Hs' BDI below 14, Ws' BDI above 14, Ws' met DSM–III criteria for MDD	Setting: Lab Situation 10-min 1[c]	KPI Audiotaped	Reliability[f] Validity (concurrent, discriminative): Full details on WWW
Bradbury, Campbell, & Fincham (1995)[15]	Type: L 40 married couples (29 nonclinic, 11 clinic): C[1], Cl[1]; 32 couples at 12 mo. follow-up (24 nonclinic, 8 clinic)	Setting: Lab Situation: 15-min. 1[b]	VTCS	Reliability[f] Validity (concurrent): Full details on WWW
Burger & Jacobson (1979)	Type: C 44 couples either married or living together: C[1,7]	Setting: Lab Situation: (1) Two 10-min. 1[b,*] (2) 10-min. 5[b]	MICS (adapted)	Reliability[g] Validity[k]
Burman, Margolin, & John (1993)[3]	Type: C 65 couples: (17 PV, 15 VA, 18 WI, and 15 ND): C[1,2,4]. PV: one severe act during the past year or 6–10 acts needed depending on severity. VA: score of greater than 45 on CTS items d–h, or	Setting: Home Situation: Two 4[c,d] or 1[c] Lengths not specified	SPAFF (adapted)	Reliability (kappa)[f] Validity[k]

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
	DAS score of less than 97, and VA score greater than WI score. WI: score of greater than 45 on CTS items e and f, or DAS score of less than 97, and WI score greater than VA score. ND: no PV during history of relationship, DAS score of greater than 97, and score of less than 45 on WI and VA score			
Caceres (1989)	Type: C 10 ND and 10 D couples (5 separated, 5 not separated): Cl[1]	Setting: Lab Situation: Approximately 14–15-min 1[b,c]	KPI	Reliability (kappa): Reported kappas for each code, but did not specify which code corresponded with which kappa (i.e., .85, .83, .75, .90, .83, .80, .73, .85). Validity (discriminative): Full details on WWW
Carrere & Gottman (1999)[12]	Type: L 124 newlywed couples: C . 17 divorced, 107 married at 3–6-year follow-up	Setting: Lab Situation: 15-min. 1[b,*] *Several topics.	SPAFF	Reliability (alpha)[f,g] Validity[k]
Carstensen, Gottman, & Levenson (1995)	Type: C 156 couples: (35 middle aged/ND, 47 middle aged/D, 43 older/ND, and 31 older/D): C[1,2,3,4]	Setting: Lab Situation: (1) 15-min. 2 (2) 15-min. 1[b] (3) 15-min. 6[e]	SPAFF	Reliability[h] Validity (discriminative): Full details on WWW
Christensen & Heavey (1990)	Type: C 31 married couples (18 had son with ADHD; 13 had son without ADHD): C[1,*], Cl[2,4] *local schools	Setting: Lab Situation: Two 6-min 1[b,d]	CRS	Reliability[f,g] Validity (concurrent): Full details on WWW
Cohan & Bradbury (1994)[2]	Type:L 60 newlywed couples: C[1]. 53 couples at 6-mo. follow-up	Setting: Lab Situation: 15-min. 1[c]	SPAFF Audiotaped	Reliability (*r*s between coders, for H and W, respectively): Humor (.83, .92), Affection (.55, .56), anger (.79, .88), Contempt (.81, .99), Whining (.69, .81), and Sadness (.95, .61). Validity (convergent): Full details on WWW
Cohan & Bradbury (1997)[2]	Type: L 60 newlywed couples: C[1]. 57 couples at 18-mo. follow-up	Setting: Lab Situation: 15-min. 1[b,c]	SPAFF, VTCS Audiotaped	Reliability SPAFF (see Cohan & Bradbury, 1994). Validity SPAFF (concurrent): Full details on WWW Reliability VTCS[f] Validity VTCS (concurrent): Full details on WWW
Cohen & Christensen (1980)	Type: T 12 married couples assigned to BCT: C[9]	Setting: Lab Situation: Four 7-min. 7: Couples communicated as (1) & (4) they typically would at home, (2) at their best, and (3) at their worst.	MICS (Adapted) Audiotaped	Reliability[f] Validity[k]
Conger, Reuter, & Elder (1999)	Type: L 383 married couples: C-letters and phone calls to parents of seventh graders. 373 couples at 3-year follow-up	Setting: Home Situation: 1 per year for 3 years 25-min. 7* *topics included history/status of the relationship, future, areas of agreement/ disagreement	IFIRS	Reliability[g] Validity[k]
Cook et al. (1995)[11]	Type: L	Setting: Lab Situation: 15-min. 1[a,b]	RCISS, MICS-III	Reliability: RCISS[f,g]; MICS[f,g]

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
	79 married couples: C[1]. 72 couples at 4-year follow-up.			Validity RCISS and MICS (concurrent). Full details on WWW
Cordova, Jacobson, Gottman, Rushe, & Cox (1993)	Type: C 57 couples: (13 ND, 29 D/PV, and 15 D/NV): C[1,2]. PV: At least 6 moderate acts, 2 severe acts, or 1 life-threatening act in the past year	Setting: Lab Situation: Two 15-min. 1[b,c]	MICS-III	Reliability (kappa): H Aversive (.79), H Facilitative (.64), H Neutral (.63), W Aversive (.75), W Facilitative (.67), and W Neutral (.60). Validity[k]
Cousins & Vincent (1983)	Type: C 42 married ND couples: C[9]	Setting: Lab Situation: Two 5-min. 7[c,*] *Topic: An upsetting event outside of the marriage	MICS	Reliability (kappa): Vague Complaint (.62), Focused Expression (.70), State Label (.77), Approve/Caring, (.59), Legitimize/Empathize (.82), Neg. Behavior (.61) Validity[k]
Davila, Bradbury, Cohan, & Tochluk (1997)[2]	Type: L 172 newlywed couples: C[10] 154 couples at 1-year follow-up	Setting: Lab Situation: Two 10-min. 5[b]	SSICS	Reliability (Pearson *rs*): Pos. Instrumental Helper (.75), Pos. Emotional Helper (.86), Pos. Other Helper (.86), Neg. Helper (.80), Pos. Helpee (.79), Neg. Helpee (.75). Validity (concurrent): Full details on WWW
Escudero, Rogers, & Gutierrez (1997)	Type: C 12 ND and 18 D couples: Cl[1], C[4]	Setting: Lab Situation: No time limit 1[b] Note: Participants were interrupted 20 min. into the discussion	Spanish translations of RCCCS, CISS	Reliability RCCS (kappa): Support (.79), Nonsupport (.76), Extension (.67), Answer (.88), Instruction (.42), Order (.75), Disconfirmation (.67), Topic Change (.20), Initiation-termination (1.0). Reliability CISS (kappa): Pos. Affect (.84), Neg. Affect (.71), Neu. Affect (.64). Validity CISS (discriminative): Full details on WWW Validity RCCCS[k]
Ewart, Taylor, Kraemer, & Agras (1984)	Type: T 20 hypertensive patients and their partners (randomly assigned to Communication Training or WLCG): Cl[2]	Setting: Lab Situation: (1) 10-min. 5[b] (2) 10-min. 1[a,b]	MICS (adapted) Audiotaped	Reliability (median % agreement): Pos. Comments (78%), Neg. Comments (80%), Neu. Comments (82%). Validity (Tx sensitivity): Full details on WWW
Ewart, Taylor, Kraemer, & Agras (1991)	Type: C 43 hypertensive patients and their partners: Cl[2]	Setting: Lab Situation: 10-min. 5[a,b]	MICS (adapted) Audiotaped	Reliability (median kappa): Hostile (.57), Supportive (.68), Neu. Codes (.66). Validity (concurrent): Full details on WWW
Fals-Stewart & Birchler (1998)	Type: C 34 married couples (17 H with drug abuse, 17 D/no drug abuse): Cl[1,2]	Setting: Lab Situation: 10-min. 1[b,c] Only 16 of the 34 tapes were coded (6 drug abuse, 10 D/no drug abuse)	CRAC	Reliability[g] Validity (discriminative, concurrent): Full details on WWW
Fehm-Wolfsdorf, Groth, Kaiser, & Hahlweg (1999)	Type: C 80 couples (65 married, 15 living together for at least 3 years): C[1]	Setting: Lab Situation: 15-min. 1[b]	KPI	Reliability[g] Validity[k]
Fichten & Wright (1983)	Study 1 — Type: C 28 couples (18 D, 10 ND): C[4], Cl[1]. Study 2— Type: C 48 D couples: C[4], Cl[1]. (18 D couples from Study 1)	Study 1 — Setting: Lab Situation: (1) 10-min. 3[c,*] (2) 10-min. 1[a,c] Topic: neu. issue Study 2 — Setting: Lab Situation: Two 10-min. 1[a,c.]	MICS (adapted)	Reliability[g] Validity[k]

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
Fitzpatrick, Fallis, & Vance (1982)	Type: C 43 couples living together for at least 6 mo.: C[9]	Setting: Home Situation: Two 10–15 min. 1[b]	VTCS (adapted)	Reliability[f] Validity[k]
Floyd (1988)[10]	Type: C 40 ND premarital couples (27 engaged couples, 13 dating couples): C[4]	Setting: Lab Situation: (1) 15-min 5[b,*] (2) 15-min 1[b,c]; *only one partner could talk at a time	CST	Reliability[f,g] Validity (convergent): Full details on WWW
Floyd, O'Farrell, & Goldberg (1987)	Type: C 47 couples (22 with H with alcohol dependence [AD] and 25 with non-AD H): C[1], Cl[2]	Setting: Lab Situation: 10-min 1[b]	CST MICS	Reliability CST[f,g]; Reliability MICS[g] Validity CST (concurrent, discriminative): Full details on WWW Validity MICS[j]
Follingstad & Haynes (1981)	Type: T (within-subjects design) 7 married couples assigned to BCT: C[1]	Setting: Home Situation: Three 30-min. 3[e]	MICS (Adapted)	Reliability[g] Validity (Tx sensitivity): Full details on WWW
Frankenstein, Hay, & Nathan (1985)	Type: C 2 W with AD, 6 H with AD and their spouses: C[1]	Setting: Lab Situation: 10-min 1[b]	MICS	Reliability[h] Validity[j]
Gill, Christensen, & Fincham (1999)	Type: L 40 married couples (29 ND, 11D): C[1], Cl[1]. 30 couples at 1-year follow-up	Setting: Lab Situation: 15-min 1[b]	RCRS, VTCS	Reliability RCRS (alpha): H Negativity (r = .90), W Negativity (r = .88), H Avoidance (r = .80), W Avoidance (r = .73). Validity RCRS[k] Reliability VTCS[f] Validity VTCS (concurrent): Full details on WWW
Gingras, Adam, & Chagnon (1983)	Type: T 19 married couples assigned to BCT: C[1,2]	Setting: Lab Situation: 12-min. 1[c]	MICS (adapted)	Reliability[i] Validity (postdictive): Full details on WWW
Gottman (1979, pp. 237–248)	Type: C 17 married, moderately adjusted couples (10 lab, 7 home): C[1]. Moderate adjustment: MRI scores 85–102	Setting: Lab and home Situation: Group 1 : 1[b] in lab Group 2: 1[b] at home Lengths not specified	CISS (Nonverbal only); Group 2 Audiotaped	Reliability (alpha): Voice Pos. (.93); Voice Neu. (.997); Voice Neg. (.988) Validity (Content: Generalizability of interaction): Full details on WWW
Gottman (1979, pp. 278–288)	Type: T (Phase 3 only) 27 married couples (assigned to BCT): C[1]	Setting: Lab Situation: No time limit 1[b]	CISS	Reliability (alpha): Problem Feeling (.99), Mindreading (.97), Problem Solving (.98), Communication Talk (.96), Agreement (.99), Disagreement (.96), Summarizing Other (.89), Summarizing Self (.81), Question (.99), Pos. Affect (.87), Neu. Affect (.99), Neg. Affect (.91). Validity (Tx sensitivity): Full details on WWW
Gottman (1980)	Type: C 19 ND & 19 D couples: C[1], Cl[1]. Marital satisfaction based on marital relationship Inventory; D: <85, ND: >102.	Setting: Lab Situation: (1) 15-min. 6[b,c] (2) Length not specified 5[b]	CISS	Reliability (kappa)[f], Reliability (alpha): Pos. Affect (.90), Neu. Affect (.96), Neg. Affect (.99). Validity (content): Full details on WWW
Gottman (1993)[11]	Type: L 79 married couples: C[1]. 72 couples at 4-year follow-up	Setting: Lab Situation: 15-min. 1[a,b]	MICS-III, SPAFF, RCISS	Reliability MICS[f], SPAFF[f], RCISS[f] Validity CISS (predictive): Full details on WWW
Gottman, Coan, Carrere, & Swanson (1998)[12]	Type: L 130 newlyweds: C[1]. 130 couples at 3–6 yr. follow-up. Only analyzed 17 divorced, 20 "stable happily	Setting: Lab Situation: One interaction per year/6 years: 15-min. 1[b]	SPAFF	Reliability[g] Validity[k]

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
	married," 20 "stable unhappily married" at T2			
Gottman et al. (1995)[7]	Type: L 61 PV couples: C[1,2,4,8]. Couples were grouped based on the Hs' heart rates during the first third of the interaction. Type 1 group comprised Hs who reduced their heart rate (n = 12). Type 2 group comprised Hs who increased their hear rates (n = 49). PV: At least 6 moderate acts, 2 severe, or 1 life-threatening act in the past year	Setting: Lab Situation: 15-min. 1[b] ½ of couples completed the same interaction at 2-year follow-up.	SPAFF	Reliability[f] Validity (discriminative): Full details on WWW
Gottman & Krokoff (1989)	Type: L Sample 1: 25 married couples: C[4,8]. 22 married couples at 3-year follow-up. Sample of Krokoff, Gottman, & Roy, 1988. Sample 2: 30 married couples: C[1]. 19 married couples at 3-year follow-up. Sample of Levenson & Gottman, 1983	Setting: Lab and home Situation: (1) Lab, 15-min. 1[b](2) Home, 15-min. 1[b] (Audiotaped Sample 1 only)	SPAFF, CISS, MICS-III	Reliability CISS[f], SPAFF[f], MICS[f] Validity CISS (concurrent): Full details on WWW Validity MICS (concurrent, predictive): Full details on WWW
Gottman & Levenson (1992)[11]	Type: L 79 married couples: C[1]. 72 couples at 4-year follow-up	Setting: Lab Situation: 15-min. 1[a,b]	MICS-III, SPAFF, RCISS	Reliability MICS[f]; SPAFF[f]; RCISS[f] Validity (discriminative, predictive): Full details on WWW
Gottman & Levenson (1999d)[11]	Type: L 79 married couples: C[1]. 42 couples at 4-year follow-up	Setting: Lab Situation: First interaction: 15-min. 1[a,b] 4-year follow-up: (1) 15-min 7[b,*]; (2) 15-min 1[a,b] *Topic: events of last 4 years	SPAFF	Reliability[i] Validity SPAFF (postdictive): Full details on WWW
Gottman & Levenson (1999b)[11]	Type: L 79 married couples: C[1]. 42 couples at 4-year follow-up	Setting: Lab Situation: See Gottman & Levenson (1999a)	SPAFF	Reliability[h] Validity (predictive): Full details on WWW
Gottman & Levenson (1999c)	Type: L 85 married couples: C[1]. 69 couples at 4-year follow-up	Setting: Lab Situation: (1) 15-min. 1[a,b] (2) 15-min. 6[b,c]	SPAFF	Reliability[f] Validity (predictive): Full details on WWW
Gottman, Markman, & Notarius (1977)	Type: C 14 ND and 14 D couples: C[1], Cl[1]	Setting: Lab Situation: Length not specified 1[b]	CISS	Reliability (not presented in article but in Gottman, 1979, p. 100): Problem Feeling (.99), Mindreading (.92), Problem Solving (.96), Communication Talk (.94), Agreement (.97), Disagreement (.96), Summarizing Other (.84), Summarizing Self (.90), Question (.94). Validity (discriminative): Full details on WWW
Gottman, Swanson, & Murray (1999)[12]	Type: L 130 newlyweds: C[1]. 130 couples at 3–6-year follow-up. Only analyzed 17 divorced, 20 "stable happily married," 20 "stable unhappily married" at T2	Setting: Lab Situation: 15-min. 1[b]	SPAFF	Reliability[f,g] Validity (discriminative): Full details on WWW

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
Gray-Little, Baucom, & Hamby (1996)[1]	Type: T 53 D couples (randomly assigned to BCT, BCT + CR, BCT + EET, BCT + CR + EET): C[1], Cl[1]	Setting: Lab Situation: Sample 1: (1) 5–6 min. 1[c] (2) 5–6 min. 5[b] Sample 2: Two 7-min. 1[b]	MICS	Reliability (% agreement): Cohort 1: Pos. (87), Neg. (83); Cohort 2: Pos. (90), Neg. (90). Validity[j]
Haber & Jacob (1997)[6]	Type: C 131 couples: (50 H with AD, 16 W with AD, 15 H and W with AD, 50 social drinkers H and W): C[1]	Setting: Lab Situation: Length unspecified 1[b,f]	MICS (adapted)	Reliability[f] Validity (discriminative): Full details on WWW
Hahlweg et al. (1984)	Type: C (Study 1); T (Study 2) Study 1: 41 couples (12 ND, 29 D): C[4,*] referrals from people Study 2: 85 D couples (34 BCT, 35 CT, 17 WLCG): C[9]	Setting: Lab Situation: 10-min. 1[b,c]	KPI	Reliability KPI (kappa): Self-Disclosure (.89), Pos. Solution (.92), Acceptance (.89), Agreement (.89), Criticize (.87), Justification (.85), Disagreement (.93), Problem Description (.86), Metacommunication (.92), Pos. Nonverbal (.86), Neg. Nonverbal (.84), Nonverbal Neu. (.70). Validity KPI (discriminative, Tx sensitivity): Full details on WWW
Hahlweg, Kaiser, Christensen, Fehm-Wolfsdorf, & Groth (2000)	Type: C 81 couples living together for at least 3 years: C[1]	Setting: Lab Situation: 15-min. 1[b]	KPI	Reliability[f,g] Validity (convergent): Full details on WWW
Hahlweg, Markman, Thurmaier, Engl, & Eckert (1998)	Type: L 96 couples randomly assigned: (64 EPL, 32 WLCG): C[1,4]. 64 EPL, 29 WLCG at 1.5-year follow-up; 61 EPL, 24 WLCG at 3-year follow-up	Setting: Lab Situation: Four 10-min. 1[c] (One at pre, post, 15 years, 3 years)	KPI	Reliability[f,g] Validity (Tx sensitivity): Full details on WWW
Hahlweg, Revenstorf, & Schindler (1982)	Type: T 85 D* couples: (17 BCT-C, 16 BCT-CG., 16 C-CT, 19 CG-CT, 17 WLCG): C[2], Cl[2] *based on participants' subjective distress	Setting: Lab Situation: (1) 10-min. 5[b] (2) 10-min. 1[c]	MICS (adapted)	Reliability[l] Validity: No significant results reported for commonly defined constructs
Hahlweg, Revenstorf, & Schindler (1984)	Type: T 50 couples (12 ND, 33 BCT, 17 WLCG): C[2,*], Cl[1] *referrals from personal contacts	Setting: Lab BCT couples: (1) 10-min. 5[b] (2) 10-min. 1[b] (3) 10-min. 1[b] (post); WLCG and ND couples: 10-min. 1[b]	KPI	Reliability[g] Validity (Tx sensitivity): Full details on WWW
Halford, Sanders, & Behrens (1993)	Type: T 26 couples randomly assigned to (13 BCT and 13 enhanced BCT): Cl[1]	Setting: Lab and home Situation (lab): (1) 10-min. 1[b]— Pre (2) 10-min. 1[b]— Post situation (home): 10-min. 1[b] — Follow-up	ICS (KPI) Follow-up Audiotaped	Reliability[f] Validity (Tx sensitivity): Full details on WWW
Harrell & Guerney (1976)	Type: T 60 couples (30 behavioral-exchange, 30 WLCG): C[9]	Setting: Lab Situation: Three length not specified 5[b]	MICS (adapted) Audiotaped	Reliability[h] Validity: No significant results reported for commonly defined constructs.
Haynes et al. (1992)	Type: C 26 elderly married couples: C[6]	Setting: Home or lab Situation: 10-min. 1[b]	CST	Reliability: Correlation between the 2 observers' mean CST scores was .86. Validity (concurrent): Full details on WWW
Haynes, Chavez, & Samuel (1984)	Type: C 67 ND and 33 D couples: C[10]	Setting: Lab Situation: 10-min. 1[b]	MICS (adapted) Audiotaped	Reliability[f,g] Validity (concurrent): Full details on WWW

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
Haynes, Follingstad, & Sullivan (1979)	Type: C 6 ND and 7 D couples: C[9], Cl[1]	Setting: Home Situation: 25-rnin. 7* *Dinner conversation	MICS (adapted)	Reliability[f] Validity (discriminative): Full details on WWW
Haynes, Jensen, Wise, & Sherman (1981)	Type: C 16 ND and 12 D couples: Cl[1]	Setting: Lab Situation: 10-min. 1[b]	MICS (adapted)	Reliability[f] Validity[k]
Heavey, Christensen, & Malamuth (1995)[12]	Type: L 48 couples (31 married, 17 exclusively dating): C[1,*], Cl[2,4] *local schools 36 couples at 2.5-year follow-up (31 married, 6 exclusively dating, 1 not in original relationship)	Setting: Lab Situation: Two 10-min 1[b,d]	CRS	Reliability (alpha): Withdrawal (.78), Demand (.78). Validity (concurrent and predictive): Full details on WWW
Heavey, Layne, & Christensen (1993)	Type: L 29 married couples: C[3]. 19 couples at approximately 1-year follow-up	Setting: Lab Situation: Two 7-min 1[b,d]	CRS	Reliability (alpha): Demand (.88), Withdrawal (.81), Pos. (.89), Neg. (.84). Validity (concurrent): Full details on WWW
Heyman, Brown, Feldbau-Kohn, & O'Leary (1999)	Type: T 60 married PV couples (randomly assigned to PV couples' T or gender specified T): C[1]. PV: At least 2 acts during the past year	Setting: Lab Situation: 10-min. 1[a]	RMICS	Acceptance (.86), Distress-Maintaining Attributions (.50–.72), Hostility (.67), Humor (.77), Other (.77), Problem Description (.67), Relationship-Enhancing Attributions (.67), Self-Disclosure (.53), Withdrawal (.62). Validity[k]
Heyman et al. (in press)	Type: T, C Sample 1: 197 couples presenting for couples therapy: Cl[1], 50 happily married control couples: C[1]; Sample 2: 157 engaged couples 3 months prior to wedding: Cl[1]	Setting: Lab Situation: 15-min. Sample 1: 1[a]; Sample 2: 1[c]	RMICS	Reliability (kappa): Distress-Maintaining Attributions (.69); Hostility (.71); Dysphoric affect (.89); Withdrawal (.51); Relationship-Enhancing Attributions (.67); Acceptance (.57); Self-Disclosure (.70); Humor (.79); and Constructive Problem Discussion (.69). Coefficient alphas provided for men and women from each sample.
Heyman, Eddy, Weiss, & Vivian (1995)	Type: A 994 couples: All MICS interactions coded at the Oregon Marital Studies Program from 1987 to 1991	Setting: Lab Situation: Varied across investigators Typical procedures: 10–15 min. 1[a,c]	MICS-IV	Reliability[h] Validity (factorial): Full details on WWW
Holtzworth-Munroe, Stuart, Sandin, Smutlzer, & McLaughlin (1997)[5]	Type: C 100 married couples: (25 D/PV, 25 ND/PV, 25 D/NV, and 25 ND/NV): C[1] PV: Both spouses reporting aggression during the lifetime of the relationship and one spouse reporting 3 or more acts in the past year	Setting: Lab. Situation: Two 8-min. 5[b]	SSICS, SSBARS	Reliability SSICS[f], SSBARS[f] Validity SSBARS[k] Validity SSICS: No significant results reported for commonly defined constructs.
Hooley & Hahlweg (1989)[4]	Type: C Study 1: 12 ND and 29 D couples: C[2,4], Cl[1] *referrals from personal contacts. Sample of Hahlweg et al., 1984 Study 2: 30 married unipolar depressed* patients and their spouses: Cl[3]. Sample of Hooley, 1986. *Met criteria for MDD; all were hospitalized unipolar depressives	Setting: Lab (Study 1), Hospital (Study 2) Situation: Study 1: 10-min. 1[b,c] Study 2: 15-min. 5[b]	KPI	Reliability Study 1 (kappa): Nonverbal pos. (.82), Nonverbal neg. (.89), Nonverbal neu. (.52). Only included a range for the remaining codes. Reliability: Study 2[f] Validity (concurrent): Full details on WWW

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
	recruited from psychiatric hospitals			
Iverson & Baucom (1990)	Type: T 48 married couples (randomly assigned to BCT, CR + BCT, BCT + EET, CR + BCT + EET): Cl[1]	Setting: Lab Situation: Two 7-min. 1[b]	MICS-III	Reliability[h] Validity (Tx sensitivity): Full details on WWW
Jacob & Krahn (1987)	Type: C 96 married couples: (37 H with alcohol dependence, 27 H with MDD, and 32 with no psychopathology): C[1]	Setting: Lab Situation: 10–15 min. 1[e]	MICS (adapted)	Reliability[h] Validity (factorial): Too few participants from 3 heterogeneous samples for factor analytic procedure. However, first 3 factors replicated in larger factor analysis by Heyman et al. (1995).
Jacob & Krahn (1988)[6]	Type: C 107 couples: (38 H with AD, 35 MDD H, 34 with no psychopathology): C[1]	Setting: Lab Situation: 9–13 min. 1[b,f]	MICS (adapted)	Reliability[f] Validity (discriminative): Full details on WWW
Jacob & Leonard (1988)[6]	Type: C 49 couples with H with AD: C[1]	Setting: Lab Situation: 10-min. 1[b,f]	MICS (adapted)	Reliability[f] Validity (discriminative): Full details on WWW
Jacob & Leonard (1992)[6]	Type: C 131 couples: (49 H with AD, 40 H with MDD, and 42 control): C[1]	Settings: Lab Situation: Two 10 min. 1 [b,d,f]	MICS (adapted)	Reliability[f] Validity (discriminative): Full details on WWW
Jacobson & Anderson (1980)	Type: T 46 ND and 14 D couples (randomly assigned to BCT + feedback, BCT + behavior rehearsal, BCT + instructions, complete T, WLCG): C[1]	Setting: Lab Situation: pre and post — (1) Two 10-min. 5[b] (2) 10-min. 1[b]	MICS (adapted)	Reliability[g] Validity[k]
Jacobson (1977)	Type: T 10 married couples (randomly assigned to BCT or WLCG): C[1]	Setting: Lab Situation: Pre: Two 5–10 min. 1[c]; Post: Two 5–10 min. 1[c]	MICS (adapted)	Reliability[i] Validity (Tx sensitivity): Full details on WWW
Jacobson (1978)	Type: T 32 couples (17 BCT, 7 no specific treatment, 6 WLCG): C[1,2], Cl[1]	Setting: Lab Situation: Pre: Two 5–10 min. 1[c]; Post: Two 5–10 min. 1[c]	MICS (adapted)	Reliability[f,g] Validity (Tx sensitivity): Full details on WWW
Jacobson et al. (1989)	Type: T 30 married couples (randomly assigned to structured BCT or flexible T): Cl[1]	Setting: Lab Situation: Two 7-min. 1[b,c]	KPI	Reliability[f,g] Validity: No significant results reported for commonly defined constructs.
Jacobson et al. (1994)[7]	Type: C 92 couples: (60 PV and 32 D/NV): C[1,2,4,8] PV: W reporting at least 6 moderate acts, 2 severe acts, or 1 life-threatening act in the past year.	Setting: Lab Situation: Two 15-min. 1[b]	SPAFF	Reliability[f,g] Validity (discriminative): Full details on WWW
Jacobson, Gottman, Gortner, Berns, & Shortt (1996)[7]	Type: L Original sample: 60 D/PV couples: C[1,2,4,8].PV: W reporting at least 6 moderate acts, 2 severe acts, or 1 life-threatening act in the past year. At 2-year follow-up: 45 D/PV couples (28 married, 17 separated/ divorced)	Setting: Lab Situation: 15-min. 1[a,b]	SPAFF	Reliability[f] Validity (discriminative): Full details on WWW
Johnson & Bradbury (1999)[2]	Type: L 60 newlywed couples: C[1]	Setting: Lab Situation: 15-min. 1[b]	VTCS (adapted)	Reliability[f,g] Validity (predictive) Full details on WWW

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
	57 couples at 6-mo. follow-up, 54 couples at 1-year follow-up			
Johnson & Jacob (1997)[6]	Type: C 141 couples: (50 H with MDD, 41 W with MDD, and 50 non-MDD): C[1]	Setting: Lab Situation: 10–15 min. 1[b,f]	MICS (adapted)	Reliability[f] Validity (discriminative): Full details on WWW
Johnson & Jacob (2000)[6]	Type: C 140 couples: (49 H with MDD, 41 W with MDD, 50 Non-MDD): C[1]	Setting: Lab Situation: 15-min. 1[b,d,f]	MICS (adapted)	Reliability[f] Validity (concurrent): Full details on WWW
Julien, Markman, & Lindahl (1989)[9]	Type: L 135 premarital ND couples: C[4]. 59 couples at 4-year follow-up (24 not in original relationship, 5 separated/ divorced, 30 did not complete all phases)	Setting: Lab Situation: Two 10–15 min. 1[b,c]. Note: On one interaction only one person was allowed to talk at a time	IDCS, CISS	Reliability CISS[h] Reliability IDCS (Pearson *rs):* Micro codes: Conflict (.47), Dominance (.27), Denial (.66), Withdrawal (.49), Neg. Affect (. 34), Communication Skills (.43), Support-Validation (.51), Problem-Solving (.27), Pos. Affect (.39). Macro Codes: Neg. Escalation (.69), Pos. Escalation (.49). Validity IDCS (concurrent, convergent, predictive): Full details on WWW Validity CISS (predictive): Full details on WWW Validity IDCS and CISS (convergent): Full details on WWW
Kaiser, Hahlweg, Fehm-Wolfsdorf, & Groth (1998)	Type: T 67 couples living together for at least 3 years: C[1] Randomly assigned: (31 EPL II, 36 WLCG).	Setting: Lab Situation: 15-min. 1[b,c]	KPI	Reliability (kappa)[f,g] Validity (Tx sensitivity): Full details on WWW
Karney & Bradbury (1997)[2]	Type: L 60 newlywed couples: C[1] Follow-up every 6 mo. for 4 years; 38 couples at 4-year follow-up	Setting: Lab Situation: 15-min. 1[b]	VTCS (adapted), Audiotaped	Reliability[f] Validity (concurrent, predictive): Full details on WWW
Katz & Gottman (1993)	Type: L 56 married couples: C[1]. 53 couples at 3-year follow-up	Setting: Lab Situation: 15-min 1[b]	SPAFF	Reliability[f,g] Validity (convergent, predictive): Full details on WWW
Kiecolt-Glaser et al. (1993)	Type: C 90 newlywed couples: R	Setting: Lab Situation: 30-min. 1[a,b]	MICS-IV	Reliability[h] Validity (discriminative): Neg., W > H; Avoidance, H > W; Humor, W > H
Kiecolt-Glaser et al. (1996)	Type: C 90 newlywed couples: R	Setting: Lab Situation: 30-min. 1[a,b]	MICS-IV	Reliability[h] Validity (convergent): Full details on WWW
Kobak & Hazan (1991)	Type: C 40 married couples: C[1,2]	Setting: Lab Situation: (1) 10-min. 1[c] (2) Two 7-min. 5[b,c]	IDCS (adapted)	Reliability (interrater *rs*): Rejection (.71), Support-Validation (.77) Validity[k]
Krokoff, Gottman, & Hass (1989)[8]	Type: C 52 married couples: R	Setting: Lab and home Situation: (1) 15-min. 1[b]—Lab. (2) 15-min. 1[b]—Home (Audiotaped)	CISS, SPAFF, RCISS	Reliability CISS (kappa)[f], SPAFF (kappa)[f], RCISS (Pearson correlation)[f]. Reliability CISS (Alpha): H pos. (.95), H Neu. (.97), H Neg. (.91), W Pos. (.95), W Neu. (.98), W Neg. (.90) Validity RCISS and CISS (convergent): Full details on WWW Validity RCISS (discriminative): Full details on WWW
Krokoff, Gottman, & Roy (1988)[8]	Type: C 52 married couples (13 D/H white-collar occupation, 13 ND/H white-collar, 13 D/H	Setting: Home Situation: 15-min. 1[b]	CISS Audiotaped	Reliability (alpha): H Pos. (.95), H Neu. (.97), W Pos. (.95), W Neu. (.98), W Neg. (.90). Validity (discriminative): Full details on WWW

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
	blue-collar, 13 ND/H blue-collar): R			
Leonard & Jacob (1997)	Type: C 99 married couples: (50 control, 30 H with steady AD, 19 H with episodic AD): C[1]	Setting: Lab Situation: Length not specified 1[b,f]	MICS (adapted)	Reliability[f] Validity: No significant differences between episodic versus steady AD groups on MICS behaviors
Leonard & Roberts (1998)	Type: C 135 newlywed couples: (60 PV and 75 NV): C[1] PV: At least 2 moderate acts since marriage, at least 1 severe act since marriage, or at least 1 moderate act since marriage and a serious act prior to marriage	Setting: Lab Situation: (1) 15-min. 1[b,c] (2) 5-min. 3[c] (3) 15-min. 1[b,c] Note: Some Hs were randomly given alcohol	MICS-IV	Reliability (% agreement): Negativity Summary Code (70%), Criticize (54%), Disagree (78%), Disapprove (66%), Interrupt (69%), Mind Read Neg. (63%), Noncomply (70%), Put Down (72%), Turn Off (64%), Command (79%); Problem Solving (83%): Compromise (64%), Neg. Solution (63%), Pos. Solution (70%), Problem Description (82%), Question (90%); Positivity Summary Code (79%): Humor (64%), Smile/Laugh (80%) Validity[k]
Liberman, Levine, Wheeler, Sanders, & Wallace (1976)	Type: T 9 married couples (4 BCT, 5 IT): C[5], Cl[2,4]	Setting: Lab Situation: (1) 10-min. 1[b] (2) 10-min. 1[b,c]	MICS	Reliability[g] Validity (Tx sensitivity): Full details on WWW
Lindahl, Clements, & Markman (1997)[10]	Type: L 25 premarital ND couples: C[4]. 25 married couples at 5-year follow-up.	Setting: Lab Situation: 10-min. 1[b,c]	IDCS	Reliability (alpha): Supportive Communication (.75), Conflictural Communication (.81). Reliability (Pearson *r):* Neg. Escalation (.77). Validity (concurrent): Full details on WWW
Margolin (1978a)[14]	Type: C 27 married D couples: C[1,2]	Setting: Lab Situation: Two 10-min. 1[c]	MICS	Reliability[f] Validity (convergent): Full details on WWW
Margolin (1978b)[14]	Type: C (assessment phase of T study) 27 married D couples: C[1,2]	Setting: Lab Situation: Two 10-min. 1[c]	MICS	Reliability[f] Validity (convergent): Full details on WWW
Margolin & Wampold (1981)	Type: C 17 ND and 22 D couples: Cl[1], C[9]	Setting: Lab Situation: Two 10-min. 1[e]	MICS	Reliability (% agreement): Pos. Behaviors (68%), Neg. Behaviors (57%), Neu. Behaviors (74%) Validity (discriminative): Full details on WWW
Margolin & Weiss (1978a, 1978b)[14]	Type: T 27 married D couples (randomly assigned to BCT, Attitudinal Restructuring + BCT, No specific T): C[1,2]	Setting: Lab Situation: Two 10-min. 1[c]	MICS	Reliability[f] Validity (Tx sensitivity): Full details on WWW
Markman, Renick, Floyd, Stanley, & Clements (1993)[9]	Type: T Original sample: 114 couples: C[9]; 4-year follow-up: (15 PREP, 24 WLCG); 5-year follow-up: (12 intervention, 18 WLCG)	Setting: Lab Situation: Two 10–15 min. 1[b]	IDCS	Reliability[f] Validity (Tx sensitivity): Full details on WWW
Matthews, Wickrama, & Conger (1996)	Type: L 436 married couples: C-letters sent home to seventh graders in school. 436 couples at 5-year follow-up	Setting: Home Situation: (1) 25-min. 7* per year/4 years. *Topics: Enjoyable times, relationships, and conflict.	IFIRS	Reliability (ICC): Hostility (H = .72, W = .73), Angry-Coercive (H = .57, W = .56), Antisocial (H = .60, W = .59) Validity[k]
McKnight, Nelson-Gray, & Gullick (1989)	Type: C 24 couples: (8 ND, 8 D, and 8 bipolar patients with varying levels of marital adjustment): Cl[3]	Setting: Lab Situation: (1) 10-min. 6[a, b] (2) 10-min. 1[a,b]	MICS	Reliability[f] Validity (discriminative): Full details on WWW

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
Melby, Conger, Ge, & Warner (1995)	Type: L 412 married couples: C-letters sent home to seventh graders in local schools. 391 at 1-year follow-up	Setting: Home Situation: 20-min. 7* Topic: Enjoyable events and disagreements in past year	IFIRS	Reliability[h] Validity[k]
Mendolia, each, & Tesser (1996)	Type: C 66 married couples: C[1,7]	Setting: Lab Situation: 20-min. 1[c]	MICS-IV	Reliability[i] Validity (convergent): Full details on WWW
Miller & Bradbury (1995)[2]	Type: C 60 newly wed couples: C[1]	Setting: Lab Situation: (1) 15-min. 1[b] (2) Two 10-min. 5[b]	VTCS SSICS	Reliability VTCS[f,g], Reliability SSICS[f] Validity VTCS (concurrent discriminative): Full details on WWW Validity SSICS (concurrent): Full details on WWW
Miller, Dopp, Myers, Stevens, & Fahey (1999)	Type: C 41 married couples: C[1] 18 high hostility (Ho), 18 low Ho groups based on the Cook-Medley Hostility scale (Cook & Medley, 1954)	Setting: Lab Situation: 15-min. 1[b]	SPAFF	Reliability[f] Validity (convergent, discriminative): Full details on WWW
Murphy & O'Farrell (1997)	Type: C 90 married couples: (60 PV and 30 NV; all with H with AD): Cl[2]	Setting: Lab Situation: 10-min. 1[c]	MICS-IV	Reliability[h] Validity (discriminative): Full details on WWW
Nelson & Beach (1990)	Type: C 60 married couples: (20 ND/Non-MDD, 20 D/ Non-MDD, and 20 D/ MOD): C[1], Cl[1], C[1]	Setting: Lab Situation: Approximately 10-min. 1[b]	KPI (adapted) Audiotaped	Reliability (alpha): Facilitative, (H = .85, W = .82), Aversive (H = .77, W = 78), Depressive (H = .70, W = .72), W: Depressive → H: Facilitative (.69), W: Depressive → H: Aversive (.78), W: Depressive → H: Aversive (.63). Validity[k]
Newton, Kiecolt-Glaser, Glaser, & Malarkey (1995)	Type: C 83 newlywed couples: C[10]	Setting: Hospital Situation: 30-min. 1[a,b]	MICS-IV	Reliability[i] Validity[h]
Noel, McCrady, Stout, & Fisher-Nelson (1991)	Type: C 45 married couples (12 W with AD, 33 H with AD): C[1,2, 6], Cl[3]	Setting: Hospital Situation: 10-min. 1[b]	MICS	Reliability[h] Validity
Noller (1982)	Type: C 48 married couples (16 D, 16 ND, 16 moderate adjustment): C[1,7,]*, Cl[1]. *referrals from other people D: MAT > 120; ND: <95; Moderate Adj.: MAT = 95–120	Setting: Lab Situation: 10-min. 7[c]	CISS, CSIC	Reliability (test-retest, kappa): Test-retest was assessed by comparing one coder's rating on two couples at two different time periods (time period not reported). Verbal Channel (H = .94, W = .92), Visual Channel (H = .88, W = .86), Vocal Channel (H = .84, W = .84). Reliability interrater, kappa): Verbal Channel (H = .92, W = .92), Visual Channel (H = .92, W = .86), Vocal Channel (H = .86, W = .90). Validity: CISS[k], CSIC[k]
Notarius & Johnson (1982)	Type: C 6 ND couples: C[9]	Setting: Lab Situation: 30-min. time limit 1[e]	CISS	Reliability[f] Validity (discriminative): Full details on WWW
Notarius, Benson, Sloane, Vanzetti, & Hornyak (1989)	Type: C 18 couples (9 ND, 9 D): C[1,2]; Cl[1]	Setting: Lab Situation: Two maximum of 20-min 1[a,c] Note: One partner talks at a time	SPAFF (adapted) Audiotaped	Reliability[f] Validity[k]
O'Farrell & Birchler (1987)	Type: C	Setting: Lab Situation: 10-min. 1[b]	MICS	Reliability[g] Validity (discriminative): Full details on WWW

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
	78 married couples: (26 H with AD, 26 D, 26 ND): Cl[1], C[i]			
O'Farrell, Cutter, & Floyd (1985)	Type: T 34 couples with H with AD (10 BCT, 12 IT, 12 WLCG): Cl[2]	Setting: Lab Situation: (1) 10-min. 1[b] (2) 10-min. 5[b]	MICS	Reliability[f] Validity[d]
Paley, Cox, Burchinal, & Payne (1999)	Type: C 138 married couples: C-prenatal classes	Setting: Home Situation: 15-min. 1[b,c]	IDCS	Reliability (interrater [unspecified metric]): Pos. Affect (.67), Neg. Affect (.88), Withdrawal (.90). Validity (discriminative): Full details on WWW
Pasch & Bradbury (1998)[2]	Type: L 60 newlywed couples: C[1]. 57 couples at 24-mo. follow-up	Setting: Lab Situation: (1) 15-min. 1[b] (2) Two 10-min. 5[b]	VTCS, SPAFF SSICS Audiotaped	Reliability VTCS (ICC): Integrative Tactics (H = .89, W = .83), Distributive Tactics (H = .80, W = .89), Avoidant Tactics (H = .91, W = .94). SPAFF (ICC): Pos. Affect (H = .80, W = .92), Neg. Affect (H = .81, W = .94). SSICS (ICC): Helpers: (Pos. = .88, Neg. = .84, Neu. = .90, Off-task = .99); Helpees (Pos. = .98, Neg. = .96, Neu. = .90, Off-task = .98). Validity SPAFF (concurrent): Full details on WWW Validity SSICS (concurrent): Full details on WWW Validity VTCS (concurrent): Full details on WWW Validity SSICS and VTCS (convergent): Full details on WWW Validity SPAFF and SSICS (convergent): Full details on WWW
Pasch, Bradbury, & Davila (1997)[2]	Type: C 60 newlywed couples: C[1]	Setting: Lab Situation: Two 10-min. 5[b]	SSICS Audiotaped	Reliability (% agreement): Helpers (79%), Helpees (87%). Reliability (kappa): Helpers (.71), Helpees (.79). Validity (discriminative, concurrent): Full details on WWW
Pasupathi, Carstensen, Levenson, & Gottman (1999)	Type: C 79 married couples: (20 D/elderly, 20 ND/ elderly, 20 D/middle aged, 19 ND/middle aged): R	Setting: Lab Situation: (1) 15-min. 2 (2) 15-min. 1[b,c]	RCISS	Reliability (kappa): Presence/ Absence of Eye Contact (H = .74; W = .80), Presence/Absence of Facial Movement (H = .34; W = .38), Presence/Absence of Backchannels (H = .88; W = .88), Pos. or Neg. Facial Expression (H = .74; W = .72). Validity[k]
Patterson, Hops, & Weiss (1975)	Type: T 10 difficult couples (e.g., divorced, separated, affairs, prolonged conflict) assigned to BCT: Cl[1]	Setting: Lab Situation: Four approximately 10-min. 1[e]	MICS	Reliability[i] Validity[k] Validity (Tx sensitivity): Full details on WWW
Revenstorf, Hahlweg, Schindler, & Vogel (1984)[4]	Type: T 40 couples: (20 BCT, 10 WLCG, 10 ND): C[9]	Setting: Lab Situation: 15-min. 5[b,c]	MICS (Adapted)	Reliability[f] Validity (discriminative): Full details on WWW
Revenstorf, Vogel, Wegener, Hahlweg, & Schindler (1980)	Type: C 10 ND and 10 D couples: C[9]	Setting: Lab Situation: 10-min. 5[b,c]	MICS (Adapted)	Reliability[h] Validity: Results reported not clear indicators of validity
Robinson & Price (1980)	Type: C 8 couples (4 ND and 4D couples): C[7]	Setting: Home Situation: Two 1-hr 1[b]	MICS (adapted)	Reliability[f] Validity[k]

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
Rogers, Castleton, & Lloyd (1996)	Type: C 25 married couples: C[1]	Setting: Lab Situation: (1) 10-min. 7* (2) 10-min. 2 (3) 10-min. 1[c] (4) 10-min. 7** *Topic: How they met and why they married. **Topic: What it takes to have a good marriage	RCCCS Audiotaped	Reliability[h] Validity[k]
Rogge & Bradbury (1999)[2]	Type: L 60 newlywed couples: C[1] 56 couples (22 ND, 16 D, 18 divorced/ separated) at 4-year follow-up	Setting: Lab. Situation: 15-min. 1[c]	SPAFF	Reliability (Pearson *r*): Anger (H = .79, W = .88), Contempt (H = .81, W = .99), Whining (H = .69, W = .81), Sadness (H = .95, W = .61), Humor (H = .83, W = .92), Affection (H = .55, W = .56). Validity (convergent, discriminative/predictive): Full details on WWW
Royce & Weiss (1975)[7]	Type: C 40 undergraduate judges rated 24 couples (12 D, 12 ND) on videotape: C[1]	Setting: Lab Situation: 10-min. 5[b]	MICS	Reliability[f] Validity (concurrent, content): Full details on WWW
Ruscher & Gotlib (1988)	Type: C 22 couples (11 with a partner EDS on BDI [EDS], 11 non-EDS): C[1]	Setting: Lab Situation: 15-min 1[b]	Developed by Gotlib & Kowalik (1985), CISS	Reliability (Gotlib & Kowalik)[g] Validity (Gotlib & Kowalik)[k], Reliability CISS[g] Validity CISS (discriminative): Full details on WWW
Sagrestano, Heavey, & Christensen (1999)	Type: C 42 married couples: C[4]	Setting: Lab Situation: Two 10-min. 1[d]	CRS	Reliability (*M* internal consistency — alpha): Demand (.88), Withdraw (.73). Reliability (Interobserver — alpha): Demand (.93), Withdraw (.81). Validity (concurrent): Full details on WWW
Sayers & Baucom (1991)[1]	Type: C 60 D couples: CI[1]	Setting: Lab Situation: Two 7-min. 1[c]	MICS-III	Reliability[l] Validity (concurrent): Full details on WWW.
Sayers, Baucom, Sher, Weiss, & Heyman (1991)[1]	Type: T 60 D couples (48 BCT, 12 WLCG): CI[1]	Setting: Lab Situation: Two 7-min. 1[c]	MICS-III	Reliability[l] Validity (Tx sensitivity, concurrent): Full details on WWW
Schaap (1984)	Type: C 27 married couples (9 D, 9 ND, 9 conflict): C[1].	Setting: Lab (1) 5-min. 7* (2) 25-min. 1[c] *Discussion regarding beginning of relationship.	MICS (Adapted), CISS-only AC used Audiotaped	Reliability MICS[g] Validity MICS (discriminative): Full details on WWW Validity CISS (discriminative): Full details on WWW
Schaap & Jansen-Nawas (1987)	Type: C 18 married couples (9 ND and 9 D couples): C[1].	Setting: Lab (1) 5-min. 7* (2) 25-min. 1[c] *Discussion regarding beginning of relationship.	CISS, MICS (Adapted)	Reliability CISS[f], MICSg Validity MICS (discriminative): Full details on WWW Validity CISS (discriminative): Full details on WWW
Schafer, Birchler, & Fals-Stewart (1994)	Type: C 31 married couples (H is a recovering polysubstance abuser): C[1], CI[2]	Setting: Lab Situation: 10-min. 7* Topic: Impact of polysubstance use on the relationship.	MICS-IV	Reliability[h] Validity
Schmaling & Jacobson (1990)	Type: C 126 couples: (32 D/W with MDD, 34 ND/ MDD; W, 36 D/Non-MDD, and 24 ND/Non-MDD): C[1,2,4], CI[1,2]	Setting: Lab Situation: (1) 5–10 min. 2 (2) Two 7-min. 1[c]	KPI	Reliability[f,g] Validity (discriminative, content): Full details on WWW
Schmaling et al. (1996)	Type: C	Setting: Lab Situation: (1) 7-min. 1[c] (2) time	LIFE	Reliability[h] Validity. No correlations between LIFE constructs

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
	6 married patients suffering from asthma and their partners: Cl[4]	unspecified 7* Topic: discuss a recent asthma attack		(Aversive, Dysphoric, Facilitative, Problem Solving) were significant given the extraordinarily low power of an *n* = 6 sample
Sher, Baucom, & Larus (1990)[1]	Type: T 47 couples: (14 D/with a partner elevated depressive symptomatology, 12 D/ Psychopathology, 9D, 12 WLCG): C[1]	Setting: Lab Situation: Two 7-min. 1[b]	MICS-III	Reliability[f] Validity[k]
Shoham, Rohrbaugh, Stickle, & Jacob (1998)	Type: T 63 couples with male with AD (37 married, 26 living together for at least 1 year): Cl[2]	Setting: Lab Situation: (1) 10-min. 1[b](2) 10-min. 7* Topic: the effects of alcohol dependence on the relationship	MICS	Reliability CRS[h]; Reliability MICS[h]; Validity CRS (Predictive): Full details on WWW Validity MICS[k]
Snyder & Wills (1989)	Type: T 79 couples (29 BCT, 30 insight-oriented couples therapy, 20 WLCG): C[1]	Setting: Lab Situation: Two 7-min 1[c]	CISS	Reliability[g,j] Validity (Tx sensitivity): Full details on WWW
Snyder, Mangrum, & Wills (1993)	Type: T 55 couples (originally 29 BCT, 30 insight-oriented couples therapy): C[1]	Setting: Lab Situation: Two 7-min. 1[c]	CISS	Reliability[h] Validity (Tx sensitivity/ predictive validity): Full details on WWW
Snyder, Trull, & Wills (1987)	Type: C 42 couples (30 clinic, 12 nonclinic): Cl[1], C[1,4]	Setting: Lab Situation: Two 7-min 1[c]	CISS	Reliability[f,g] Validity (convergent): Full details on WWW
Stein, Giordo, & Dotzenroth (1982)	Type: C 26 married couples: C[1,2]	Setting: Lab Situation: 10-min. 1[b]	MICS	Reliability[f] Validity (convergent): Full details on WWW
Vincent, Friedman, Nugent, & Messerly (1979)	Type: C 20 ND and 20 D couples: C[8]	Setting: Lab Situation: (1) 10-min. 5[b,*]: Neu. (2) 10-min. 5[b,*]: Faking good (3) 10-min. 5[b,*]. Faking bad	MICS	Reliability (ICC): Neu. condition: Problem-Solving Behaviors (.90), Problem Descriptive Behaviors (.60), Neg. Verbal Behaviors (.85), Pos. Verbal Behaviors (41), Pos. Nonverbal Behaviors (.88), Neg. Nonverbal Behaviors (.46); Faking conditions: Problem-Solving Behaviors (.76), Problem Descriptive Behaviors (.61), Neg. Verbal Behaviors (. 96), Pos. Verbal Behaviors (.64), Pos. Nonverbal Behaviors (.89), Neg. Nonverbal Behaviors (.78). Validity (discriminative): Full details on WWW
Vincent, Weiss, & Birchler (1975)[13]	Type: C 12 ND and 12 D couples: C[1]	Setting: Lab Situation: (1) 4-min. 3[c] (2) length not specified 5[b]	MICS	Reliability[i] Validity (discriminative): Pos. Behavior, ND > D
Walker, Johnson, Manion, & Cloutier (1996)	Type: T 32 married couples: (randomly assigned to 16 EFT, 16 WLCG): C[4]*; *letters and phone calls to parents with ill children at a pediatric hospital	Setting: Lab Situation: 15-min. 1[e] Note: Eight (four for each group) interactions were not coded.	CST	Reliability[f,g] Validity (Tx sensitivity): Across time, Neg. Communication EFT < WLCG.
Weiss & Tolman (1990)	Type: A 26 ND and 24 D couples from five U.S.A. studies.	Setting: Lab Situation: 4 studies: 10-min. l[a,b], 1 study: 7-min. 1[a,b]	MICS-III, MICS-G	Reliability MICS-G (% agreement): Withdrawal (Hs = 82%, Ws = 86%). Reliability MICS-G (ICC): Withdrawal (Hs = .56, Ws = .54). Reliability MICS[g] Validity (MICS, MICS-G convergent): Full details on WWW

Study	Type of study/ participants	Situations/ Settings	Observation methods	Psychometrics (reliability/ validity)
				Validity MICS-G (discriminative, concurrent): Full details on WWW Validity MICS (concurrent): Full details on WWW
Weiss, Hops, & Patterson (1973)	Type: T 5 couples assigned to BCT: C[9]	Setting: Unspecified Situation: Unspecified	MICS	Reliability[l] Validity (Tx sensitivity): Full details on WWW
Wieder & Weiss (1980)	Type: C 14 married couples: Cl[1]	Setting: Lab Situation: Two 10-min. 1[b]	MICS, Audiotaped	Reliability[g] Validity (generalizability): Full details on WWW
Wilson, Bornstein, & Wilson (1988)	Type: T 15 D couples: (5 CG, 5 C, 5 WLCG): C[9]	Setting: Lab Situation: Length unspecified 1[b]	MICS, Audiotaped	Reliability (kappa): Pos. Verbal Behavior (.78), Neg. Verbal Behavior (.59) Validity[k]
Witkin & Rose (1978)	Type: T 14 D couples (assigned to BCT): C[13], Cl[1]. Based on revised-MAT with cut-off scores with a range of 105–110 (Kimmel & VanDerVeen, 1974)	Setting: Lab Situation: Unspecified 1[c]	MICS	Reliability[f] Validity[k]

Notes. Table 1 is a severely abridged version of the complete table (i.e., psychometrics of all observational studies of couples). Coding systems used by fewer than three studies were censored. Because validity inferences can only be drawn from the relations of specific code constructs to specific dependent variables in specific situations, validity details were censored as well. The full table is available on the World Wide Web (WWW) at http://www.psy.sunysb.edu/marital, at http://www.aabtcouples.org, or from the author on request. BCT = behavioral couples therapy; C = conjoint; CG = conjoint-group; CTS = Conflict Tactics Scale (Straus, 1979); EPL = German version of PREP; H = husband; neg. = negative; neu. = neutral; pos. = positive; MDD = Major Depressive Disorder; PREP = Premarital Relationship Enhancement Program; T1 = Time 1; T2 = Time 2, etc.; Tx = treatment; W = wife; WLCG = wait list control group; CR = cognitive restructuring; EET = emotional expressive training; AD = alcohol dependence; IT = interactional training; CT = communication training; EDS = elevated depression symptomatology; EFT = emotionally focused therapy.

Study: Observational data was used in more than one study included in this table. Primary paper describing data set: [1] Baucom, Sayers, & Sher, 1990; [2]Bradbury, 1994; [3]Burman, John, & Margolin, 1987; [4]Hahlweg, Schindler, Revenstorf, & Brengelmann, 1984; [5]Holtzworth-Munroe, Stuart, Sandin, Smutlzer, & McLaughlin, 1997; [6]Jacob, Seilhamer, & Rushe, 1989; [7] Jacobson, Gottman, Waltz, Rushe, Babcock, & Holtzworth-Munroe, 1994; [8]Krokoff, 1987; [9]Markman, Duncan, Storaasli, & Howes, 1987; [10]Markman, Floyd, Stanley, & Storaasli, 1988; [11]Gottman & Levenson, 1992; [12]Gottman, Coan, Carrere, & Swanson, 1998; [13]Birchler, Weiss, & Vincent, 1975; [14]Margolin & Weiss, 1978a, 1978b; [15]Bradbury & Fincham, 1992.

Type of Study: A = Archival, L = Longitudinal; C = Cross-sectional; T = Treatment.

Participants: Sampling used: Community: C[1] = advertising—newspapers; C[2] = advertising—TV, radio; C[3] = advertising—flyers; C[4] = advertising—other; C[5] = church; C[6] = nonchurch community group; C[7] = college students; C[8] = partly through random sampling; C[9] = unspecified; C[10] = marriage licenses; C[R] = representative sample (recruited from a specified sampling frame, with all appropriate participants having an equal chance of inclusion); Cl[1] = clinical, outpatient, marital treatment; Cl[2] = clinical, outpatient, other treatment; Cl[3] = clinical, inpatient; Cl[4] = clinical, outpatient, other medical setting.

Couple status: D = distressed (typically ≤ 100 on Marital Adjustment Test [MAT], Locke & Wallace, 1959, or ≤ 97 on Dyadic Adjustment Scale [DAS], Spanier, 1976); ND = nondistressed (typically ≥ 100 on MAT or ≥ 98 on DAS); PV = partner (i.e., husband → wife) violent; NV = not partner violent.

Situations: 1 = conflict; 2 = events-of-the day; 3 = unstructured discussion; 4 = reenactment of prior conflict; 5 = role play of standard scenario; 6 = pleasant conversation; 7 = other; [a] = Topic picked by experimenter from list of possible conflicts; [b] = Topic was narrowed down from general topic (e.g., "money") through interview (e.g., Gottman's play-by-play interview, Gottman, 1996) or specificity of topic selected (e.g., Areas-of-Change questionnaire, Weiss, Hops, & Patterson, 1973); [c] = Topic picked by participants; [d] = Topic picked was top problem for specified gender; [e] = How topic was chosen was not specified; [f] = Alcohol made available during one of the conversations.

Observation methods: Numbers in parentheses indicates number of times coding system used in unabridged table (Total: 231): CISS = Couples Interaction Scoring System: (17); COMFI = Codebook of Marital and Family Interaction: (2); Communication Box: (2); COS = Category Observation System: (2); CRAC = Clinical Rating of Adult Communication Scale: (3); CRS = Conflict Rating System: (6); CSIC = Coding Scheme for Interpersonal Conflict: (2); CST = Communication Skills Test: (4); Developed for this study: (33); DISC = Dyadic Interaction Scoring Code: (2); FAMISS = Family and Marital Interaction Scoring System II: (2); Gotlib & Kowalik (1985) coding system: (1); IBRS = Interpersonal Behavior Rating System: (1); ICS (KPI) =

Interactional Coding System (KPI) (1); IDCS = Interactional Dimensions Coding System: (5); IFIRS = Iowa Family Interaction Rating Scales: (3); KPI = Kategoriensystem für Partnerschaftliche Interaktion: (13); LIFE = Living In Family Environments coding system: (3); MICS = Marital Interaction Coding System: (76); MICS-G = Marital Interaction Coding System—Global: (1); Rapid-KPI = Rapid Kategoriensystem für Partnerschaftliche Interaktion: (1); RCCCS = Relational Communication Control Coding System: (2); RCISS = Rapid Couples Interaction Scoring System: (5); RCRS = Rapid Conflict Rating System: (1); RMICS = Rapid Marital Interaction Coding System: (3); SPAFF = Specific Affect Coding System: (23); SSBARS = Social Support Behavior/Affect Rating System: (1); SSBC = Social Support Behavior Coding system: (2); SSICS = Social Support Interaction Coding System: (5)-VTCS = Verbal Tactics Coding Scheme: (10).

Reliability: When interrater agreement for specific constructs is not provided: [f] = interrater agreement provided for overall system; [g] = range of interrater agreements provided, but no specific information for constructs of interest; [h] = No interrater agreement provided; [i] = No specific interrater agreement provided, but a minimum criterion for agreement during training and/or coding is mentioned. ICC = Intraclass Correlation Coefficients.

Validity: [j] = codes constituting constructs not specified or only some of the codes are specified; [k] = Constructs as constructed not used in any other published study, [l] Andrews et al. used *aversive* for a construct that had been previously labeled in marital depression studies as *aggressive* (e.g., Biglan et al., 1985). Because *aversive* more clearly describes the component codes (and to avoid confusion with verbal or physical aggression), we use the term *aversive* for this construct in all depression studies where it was used.