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

1Adapted from Charles Dickens (1859), A Tale of Two Cities.
2I 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, & Jongsmas, 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. Furthermore, both partners in distressed relationships characterized by husband-to-

---

³Research 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). 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.

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

---

4Heyman 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).

*Psychol Assess. Author manuscript; available in PMC 2006 April 12.*
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 ... 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. 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). 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

---

9Given 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.

10Christensen 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.

11Because 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 sometimes 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. 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.

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. 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 α = .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, α = .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)

Psychol Assess. Author manuscript; available in PMC 2006 April 12.
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.

---

14To 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. 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

---

15Many 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 obscures 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 (at least) 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. 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 hostiley).

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.

17Clinicians 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]


Psychol Assess. Author manuscript; available in PMC 2006 April 12.


Psychol Assess. Author manuscript; available in PMC 2006 April 12.


Psychol Assess. Author manuscript; available in PMC 2006 April 12.


Psychol Assess. Author manuscript; available in PMC 2006 April 12.


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


Psychol Assess. Author manuscript; available in PMC 2006 April 12.


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)

<table>
<thead>
<tr>
<th>Study</th>
<th>Type of study/participants</th>
<th>Situations/Settings</th>
<th>Observation methods</th>
<th>Psychometrics (reliability/validity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adam &amp; Gingras (1982)</td>
<td>Type: T 38 married couples (19 T, 19 WLCG): C</td>
<td>Setting: Lab Situation: 12-min. 1&lt;sup&gt;1b&lt;/sup&gt;</td>
<td>MICS</td>
<td>Reliability&lt;sup&gt;2&lt;/sup&gt; Validity (Tx sensitivity): Full details on <a href="http://WWW">WWW</a>. (See table note.)</td>
</tr>
<tr>
<td>Andrews, Foster, Capaldi, &amp; Hops (2000)</td>
<td>Type: L 716 adolescents and their parents at T1: C-letters sent to parents of 9–12&lt;sup&gt;th&lt;/sup&gt; graders at local schools T3: 314 couples at 6-year follow-up (153 married, 161 dating for at least 1 month)</td>
<td>Setting: Lab Situation: Two 10-min. 1&lt;sup&gt;b&lt;/sup&gt; 153 couples did interactions</td>
<td>LIFE</td>
<td>Reliability: (ICC); Participant Aversive (.81), Partner Aversive&lt;sup&gt;1&lt;/sup&gt; (.83), Participant Facilitative (.81), Partner Facilitative (.82). Validity (concurrent, postdictive): Full details on WWW</td>
</tr>
<tr>
<td>Aron, Norman, Aron, McKenna, &amp; Heyman (2000)</td>
<td>Type: C 35 married couples: C</td>
<td>Setting: Lab Situation: (1) 5-min. 6&lt;sup&gt;b&lt;/sup&gt; (2) 5-min. 7&lt;sup&gt;b&lt;/sup&gt; Topic: Planning vacation; home improvements. N = 26 couples coded</td>
<td>RMICS</td>
<td>Reliability (kappa): Acceptance (.64), Hostility (.66). Validity&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>Basco, Birchler, Kalal, Talbott, &amp; Slater (1991)</td>
<td>Type: A 36 randomly selected videotapes 10 ND and 24 D couples: C&lt;sup&gt;1&lt;/sup&gt;, Cl&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab Situation: 10-min. 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>CRAC, MICS-II</td>
<td>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&lt;sup&gt;4&lt;/sup&gt; Validity CRAC and MICS-III (convergent): Full details on <a href="http://WWW">WWW</a>. Reliability&lt;sup&gt;4&lt;/sup&gt; Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td>Basco, Prager, Pita, Tamir, &amp; Stephens (1992)</td>
<td>Type: C 34 married couples: (17 depressed*, 17 control): C&lt;sup&gt;1&lt;/sup&gt;, C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab Situation: 8-min. 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>CRAC</td>
<td>Reliability&lt;sup&gt;5&lt;/sup&gt; Validity (Tx sensitivity): Full details on WWW</td>
</tr>
<tr>
<td>Baucom (1982)</td>
<td>Type: T 72 D couples (randomly assigned to BCT and WLCG): C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab Situation: (1) 5–6 min. 1&lt;sup&gt;<em>&lt;/sup&gt; (2) 5–6 min. 5&lt;sup&gt;</em>&lt;/sup&gt;</td>
<td>MICS</td>
<td>Reliability&lt;sup&gt;5&lt;/sup&gt; Validity (Tx sensitivity): Full details on WWW</td>
</tr>
<tr>
<td>Baucom, Sayers, &amp; Sher (1990)</td>
<td>Type: T 60 D couples: C&lt;sup&gt;1&lt;/sup&gt; Randomly assigned to BCT, CR + BCT, BCT + EET, CR + BCT + EET, WLCG</td>
<td>Setting: Lab Situation: Two 7-min. 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>MICS-III</td>
<td>Reliability&lt;sup&gt;5&lt;/sup&gt; Validity (Tx sensitivity): Full details on WWW</td>
</tr>
<tr>
<td>Behrens, Sanders, &amp; Halford (1990)</td>
<td>Type: T 4 D couples (assigned to BCT + Cognitive BCT): C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab and home. Situation: Lab: unspecified 1&lt;sup&gt;b&lt;/sup&gt; Home: Two unspecified 1&lt;sup&gt;b&lt;/sup&gt; Setting: Lab Situation: 15-min. 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>KPI (adapted), Home Audiotaped</td>
<td>Reliability&lt;sup&gt;5&lt;/sup&gt; Validity (Tx sensitivity) Full details on WWW</td>
</tr>
<tr>
<td>Berns, Jacobson, &amp; Gottman (1999a)</td>
<td>Type: C 91 couples (47 PV, 28 D/NV, 16 ND/NV): C&lt;sup&gt;1&lt;/sup&gt;, C&lt;sup&gt;2&lt;/sup&gt;, C&lt;sup&gt;3&lt;/sup&gt; PV: At least 6 moderate acts, 2 severe acts, or 1</td>
<td>Setting: Lab and home. Situation: Lab: unspecified 1&lt;sup&gt;b&lt;/sup&gt; Home: Two unspecified 1&lt;sup&gt;b&lt;/sup&gt; Setting: Lab Situation: 15-min. 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>CRS (adapted), SPAFF</td>
<td>Reliability SPAFF&lt;sup&gt;5&lt;/sup&gt; Reliability CRS (M ICC): Demand (.82), Withdraw (.81), Pos. Communication (.84) Neg. Communication (.85).</td>
</tr>
<tr>
<td>Study</td>
<td>Type of study/ participants</td>
<td>Situations/ Settings</td>
<td>Observation methods</td>
<td>Psychometrics (reliability/ validity)</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-------------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>-------------------------------------------------------------</td>
</tr>
<tr>
<td>Berns, Jacobson, &amp; Gottman (1999b)</td>
<td>Type: C</td>
<td>Setting: Lab Situation: 15-min. 1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>CRS (adapted)</td>
<td>Validity CRS (concurrent, discriminative): Full details on <a href="http://WWW">WWW</a>. Validity SPAFF and CRS (convergent): Full details on <a href="http://WWW">WWW</a>. Reliability&lt;sup&gt;i&lt;/sup&gt; Validity (discriminative): Full details on WWW.</td>
</tr>
<tr>
<td>Biglan, Hops, Sherman, Friedman, Arthur, &amp; Osteen (1985)</td>
<td>47 PV couples: C&lt;sup&gt;1,2,4,8&lt;/sup&gt;. See (1999a) for PV inclusion criteria</td>
<td>Setting: Lab Situation: Two 10-min. 1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>LIFE</td>
<td>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.</td>
</tr>
<tr>
<td>Birchler, Clopton, &amp; Adams (1984)</td>
<td>Type: C</td>
<td>Setting: Lab Situation: 10-min. 1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>MICS</td>
<td>Reliability (discriminative): Full details on <a href="http://WWW">WWW</a>. Reliability&lt;sup&gt;b&lt;/sup&gt; Validity (discriminative): Full details on WWW.</td>
</tr>
<tr>
<td>Birchler, Weiss, &amp; Vincent (1975)</td>
<td>Type: C</td>
<td>Setting: Lab Situation: (1) 4-min. 3&lt;sup&gt;c&lt;/sup&gt; (2) 10-min. 5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>MICS</td>
<td>Reliability&lt;sup&gt;b&lt;/sup&gt; Validity (discriminative): Full details on WWW.</td>
</tr>
<tr>
<td>Boelen, Emmelkamp, MacGillavry, &amp; Markvoort (1980)</td>
<td>Type: T</td>
<td>Setting: Lab Situation: 20-min. 5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>MICS</td>
<td>Reliability&lt;sup&gt;b&lt;/sup&gt; Validity (discriminative): Full details on WWW.</td>
</tr>
<tr>
<td>Bradbury &amp; Fincham (1992)&lt;sup&gt;2,5&lt;/sup&gt;</td>
<td>Type: C</td>
<td>Study 1: Setting: Lab Situation: 15-min. 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Study 1: Developed for study Study 2: VTCS (adapted)</td>
<td>Reliability (developed for study)&lt;sup&gt;d&lt;/sup&gt; Validity (developed for study)&lt;sup&gt;d&lt;/sup&gt; Reliability (VTCS)&lt;sup&gt;f&lt;/sup&gt; Validity VTCS (concurrent, discriminative): Full details on WWW.</td>
</tr>
<tr>
<td>Bradbury &amp; Fincham (1993)&lt;sup&gt;3&lt;/sup&gt;</td>
<td>Type: C</td>
<td>Setting: Lab Situation: 15-min. 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>VTCS</td>
<td>Reliability&lt;sup&gt;b&lt;/sup&gt; Validity (concurrent, discriminative): Full details on <a href="http://WWW">WWW</a>. Reliability&lt;sup&gt;b&lt;/sup&gt; Validity (discriminative): Full details on WWW.</td>
</tr>
<tr>
<td>Bradbury, Beach, Fincham, &amp; Nelson (1996)</td>
<td>Type: C</td>
<td>Setting: Lab Situation 10-min 1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>KPI Audiotaped</td>
<td>Validity (concurrent, discriminative): Full details on WWW.</td>
</tr>
<tr>
<td>Bradbury, Campbell, &amp; Fincham (1995)&lt;sup&gt;4,5&lt;/sup&gt;</td>
<td>Type: L</td>
<td>Setting: Lab Situation: 15-min. 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>VTCS</td>
<td>Reliability&lt;sup&gt;f&lt;/sup&gt; Validity (concurrent): Full details on WWW.</td>
</tr>
<tr>
<td>Burger &amp; Jacobson (1979)</td>
<td>Type: C</td>
<td>Setting: Lab Situation: (1) Two 10-min. 1&lt;sup&gt;b&lt;/sup&gt;, (2) 10-min. 5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>MICS (adapted)</td>
<td>Reliability&lt;sup&gt;d&lt;/sup&gt; Validity&lt;sup&gt;k&lt;/sup&gt;</td>
</tr>
<tr>
<td>Burman, Margolin, &amp; John (1993)&lt;sup&gt;3&lt;/sup&gt;</td>
<td>Type: C</td>
<td>Setting: Home Situation: Two 4&lt;sup&gt;h&lt;/sup&gt; or 1&lt;sup&gt;i&lt;/sup&gt; Lengths not specified</td>
<td>SPAFF (adapted)</td>
<td>Reliability (kappa)&lt;sup&gt;f&lt;/sup&gt; Validity&lt;sup&gt;k&lt;/sup&gt;</td>
</tr>
<tr>
<td>Study</td>
<td>Type of study/participants</td>
<td>Situations/Settings</td>
<td>Observation methods</td>
<td>Psychometrics (reliability/validity)</td>
</tr>
<tr>
<td>------------------------------</td>
<td>----------------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Caceres (1989)</td>
<td>Type: C</td>
<td>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</td>
<td>Setting: Lab Situation: Approximately 14-15-min 1 2</td>
<td>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</td>
</tr>
<tr>
<td>Carrere &amp; Gottman (1999)</td>
<td>Type: L</td>
<td>124 newlywed couples: C: 17 divorced, 107 married at 3-6-year follow-up</td>
<td>Setting: Lab Situation: 15-min. 1 2</td>
<td>Reliability (alpha) 6 7 Validity 6</td>
</tr>
<tr>
<td>Carstensen, Gottman, &amp; Levenson (1995)</td>
<td>Type: C</td>
<td>156 couples: (35 middle aged/ND, 47 middle aged/D, 43 older/ND, and 31 older/D): C1,2,3,4</td>
<td>Setting: Lab Situation: (1) 15-min. 2 (2) 15-min. 1 3 (3) 15-min. 6</td>
<td>Reliability 6 Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td>Christensen &amp; Heavey (1990)</td>
<td>Type: C</td>
<td>31 married couples (18 had son with ADHD; 13 had son without ADHD): C1,2,3,4</td>
<td>Setting: Lab Situation: Two 6-min 1 4</td>
<td>Reliability 6 Validity (concurrent): Full details on WWW</td>
</tr>
<tr>
<td>Cohan &amp; Bradbury (1994)</td>
<td>Type: L</td>
<td>60 newlywed couples: C1, 53 couples at 6-mo. follow-up</td>
<td>Setting: Lab Situation: 15-min. 1 5</td>
<td>Reliability (rs 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</td>
</tr>
<tr>
<td>Cohan &amp; Bradbury (1997)</td>
<td>Type: L</td>
<td>60 newlywed couples: C1, 57 couples at 18-mo. follow-up</td>
<td>Setting: Lab Situation: 15-min. 1 6</td>
<td>Reliability SPAFF (see Cohan &amp; Bradbury, 1994). Validity SPAFF (concurrent): Full details on WWW</td>
</tr>
<tr>
<td>Cohen &amp; Christensen (1980)</td>
<td>Type: T</td>
<td>12 married couples assigned to BCT: C1</td>
<td>Setting: Lab Situation: Four 7-min. 7: Couples communicated as (1) &amp; (4) they typically would at home, (2) at their best, and (3) at their worst.</td>
<td>Reliability 6 Validity 6</td>
</tr>
<tr>
<td>Conger, Reuter, &amp; Elder (1999)</td>
<td>Type: L</td>
<td>383 married couples: C-letters and phone calls to parents of seventh graders. 373 couples at 3-year follow-up</td>
<td>Setting: Home Situation: 1 per year for 3 years 25-min. 7 8</td>
<td>Reliability 6 Validity 6</td>
</tr>
<tr>
<td>Cook et al. (1995)</td>
<td>Type: L</td>
<td>Setting: Lab Situation: 15-min. 1 9</td>
<td>RCISS, MICS-III</td>
<td>Reliability: RCISS 6, MICS 6</td>
</tr>
<tr>
<td>Study</td>
<td>Type of study/participants</td>
<td>Situations/Settings</td>
<td>Observation methods</td>
<td>Psychometrics (reliability/validity)</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Cordova, Jacobson, Gottman, Rushe, &amp; Cox (1993)</td>
<td>79 married couples: C&lt;sup&gt;1&lt;/sup&gt;, 72 couples at 4-year follow-up.</td>
<td>Setting: Lab Situation: Two 15-min.</td>
<td>MICS-III</td>
<td>Validity RCISS and MICS (concurrent). Full details on <a href="http://WWW">WWW</a>. Reliability (kappa): H Aversive (.79), H Facilitative (.64), H Neutral (.63), W Aversive (.75), W Facilitative (.67), and W Neutral (.60). Validity&lt;sup&gt;6&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cousins &amp; Vincent (1983)</td>
<td>Type: C 42 married ND couples: C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab Situation: Two 5-min.</td>
<td>MICS</td>
<td>Reliability (kappa): Vague Complaint (.62), Focused Expression (.67), State Label (.77), Approve/Caring (.59), Legitimize/Empathize (.82), Neg. Behavior (.61). Validity&lt;sup&gt;5&lt;/sup&gt;</td>
</tr>
<tr>
<td>Escudero, Rogers, &amp; Gutierrez (1997)</td>
<td>Type: C 12 ND and 18 D couples: C&lt;sup&gt;1&lt;/sup&gt;, C&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Setting: Lab Situation: No time limit 1&lt;sup&gt;b&lt;/sup&gt; Note: Participants were interrupted 20 min. into the discussion</td>
<td>Spanish translations of RCCCS, CISS</td>
<td>Validity (discriminative): Full details on <a href="http://WWW">WWW</a>. Validity RCCCS&lt;sup&gt;6&lt;/sup&gt;</td>
</tr>
<tr>
<td>Ewart, Taylor, Kraemer, &amp; Agras (1984)</td>
<td>Type: T 20 hypertensive patients and their partners (randomly assigned to Communication Training or WLCG): C&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Setting: Lab Situation: (1) 10-min. 5&lt;sup&gt;c&lt;/sup&gt; (2) 10-min. 1&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>MICS (adapted) Audiotaped</td>
<td>Reliability (median % agreement): Pos. Comments (78%), Neg. Comments (80%), Neu. Comments (82%). Validity (Tx sensitivity): Full details on <a href="http://WWW">WWW</a>. Reliability (median kappa): Hostile (.57), Supportive (.68), Neu. Codes (.66). Validity (concurrent): Full details on <a href="http://WWW">WWW</a>. Reliability&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Ewart, Taylor, Kraemer, &amp; Agras (1991)</td>
<td>Type: C 43 hypertensive patients and their partners: C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab Situation: 10-min. 5&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>MICS (adapted) Audiotaped</td>
<td>Reliability (median kappa): Hostile (.57), Supportive (.68), Neu. Codes (.66). Validity (concurrent): Full details on <a href="http://WWW">WWW</a>. Reliability&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Fals-Stewart &amp; Birchler (1998)</td>
<td>Type: C 34 married couples (17 H with drug abuse, 17 D/no drug abuse): C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab Situation: 10-min. 1&lt;sup&gt;b&lt;/sup&gt; Only 16 of the 34 tapes were coded (6 drug abuse, 10 D/no drug abuse)</td>
<td>CRAC</td>
<td>Reliability (median kappa): Hostile (.57), Supportive (.68), Neu. Codes (.66). Validity (concurrent): Full details on <a href="http://WWW">WWW</a>. Reliability&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>Fehn-Wolfsdorf, Groth, Kaiser, &amp; Hahlweg (1999)</td>
<td>Type: C 80 couples (65 married, 15 living together for at least 3 years): C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab Situation: 15-min. 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>KPI</td>
<td>Reliability&lt;sup&gt;6&lt;/sup&gt; Validity&lt;sup&gt;8&lt;/sup&gt;</td>
</tr>
<tr>
<td>Fichten &amp; Wright (1983)</td>
<td>Study 1 — Type: C 28 couples (38 D, 10 ND): C&lt;sup&gt;1&lt;/sup&gt;, C&lt;sup&gt;2&lt;/sup&gt; Study 2 — Type: C 48 D couples: C&lt;sup&gt;2&lt;/sup&gt;, C&lt;sup&gt;1&lt;/sup&gt; (16 D couples from Study 1)</td>
<td>Setting: Lab Situation: (1) 10-min. 3&lt;sup&gt;ab&lt;/sup&gt; (2) 10-min. 1&lt;sup&gt;ab&lt;/sup&gt; Topic: neu. issue Study 2 — Setting: Lab Situation: Two 10-min. 1&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>MICS (adapted)</td>
<td>Reliability&lt;sup&gt;6&lt;/sup&gt; Validity&lt;sup&gt;8&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

*Note: * Details on methodology and findings are not included here for brevity.
<table>
<thead>
<tr>
<th>Study</th>
<th>Type of study/ participants</th>
<th>Situations/ Settings</th>
<th>Observation methods</th>
<th>Psychometrics (reliability/ validity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitzpatrick, Fullis, &amp; Vance</td>
<td>Type: C 43 couples living together for at least 6 mo.</td>
<td>Setting: Home Situation: Two 10-15 min.</td>
<td>VTCS (adapted)</td>
<td>Reliability&lt;sup&gt;a&lt;/sup&gt;; Validity&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Floyd (1988)&lt;sup&gt;10&lt;/sup&gt;</td>
<td>Type: C 40 ND premarital couples (27 engaged couples, 13 dating couples): C&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Setting: Lab Situation: (1) 15-min 5&lt;sup&gt;a&lt;/sup&gt; (2) 15-min 1&lt;sup&gt;b&lt;/sup&gt;; only one partner could talk at a time</td>
<td>CST</td>
<td>Reliability&lt;sup&gt;g&lt;/sup&gt;; Validity (convergent): Full details on WWW</td>
</tr>
<tr>
<td>Floyd, O’Farrell, &amp; Goldberg</td>
<td>Type: C 47 couples (22 with H with alcohol dependence [AD] and 25 with non-AD H): C&lt;sup&gt;1&lt;/sup&gt;, Cl&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Setting: Lab Situation: 10-min 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>MICS</td>
<td>Reliability&lt;sup&gt;3&lt;/sup&gt;; Reliability MICS&lt;sup&gt;j&lt;/sup&gt;; Reliability CST&lt;sup&gt;i&lt;/sup&gt;,&lt;sup&gt;j&lt;/sup&gt;; Reliability MICS&lt;sup&gt;j&lt;/sup&gt;; Validity CST (concurrent, discriminative): Full details on WWW Validity MICS&lt;sup&gt;j&lt;/sup&gt;</td>
</tr>
<tr>
<td>Follingstad &amp; Haynes (1981)</td>
<td>Type: T (within-subjects design) 7 married couples assigned to BCT: C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Home Situation: Three 30-min. 3&lt;sup&gt;c&lt;/sup&gt;</td>
<td>MICS (Adapted)</td>
<td>Reliability&lt;sup&gt;g&lt;/sup&gt;; Validity (Tx sensitivity): Full details on WWW</td>
</tr>
<tr>
<td>Frankenstein, Hay, &amp; Nathan</td>
<td>Type: C 2 W with AD, 6 H with AD and their spouses: C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab Situation: 10-min 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>MICS</td>
<td>Reliability&lt;sup&gt;b&lt;/sup&gt;; Validity&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Gill, Christensen, &amp; Fincham</td>
<td>Type: L 40 married couples (29 ND, 11D): C&lt;sup&gt;1&lt;/sup&gt;, Cl&lt;sup&gt;1&lt;/sup&gt;, 30 couples at 1-year follow-up</td>
<td>Setting: Lab Situation: 15-min 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>RCRS, VTCS</td>
<td>Reliability RCRS (alpha): H Negativity (r = .90), W Negativity (r = .88), H Avoidance (r = .80), W Avoidance (r = .73). Validity RCRS&lt;sup&gt;3&lt;/sup&gt; Reliability VTCS&lt;sup&gt;j&lt;/sup&gt;; Validity VTCS (concurrent): Full details on WWW</td>
</tr>
<tr>
<td>Ginges, Adam, &amp; Chagnon (1983)</td>
<td>Type: T 19 married couples assigned to BCT: C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab Situation: 12-min. 1&lt;sup&gt;i&lt;/sup&gt;</td>
<td>MICS (adapted)</td>
<td>Reliability&lt;sup&gt;j&lt;/sup&gt;; Validity (postdictive): Full details on WWW</td>
</tr>
<tr>
<td>Gottman (1979, pp. 237–248)</td>
<td>Type: C 17 married, moderately adjusted couples (10 lab, 7 home): C&lt;sup&gt;1&lt;/sup&gt; Moderate adjustment: MRI scores 85–102</td>
<td>Setting: Lab and home Situation: Group 1 : 1&lt;sup&gt;b&lt;/sup&gt; in lab Group 2: 1&lt;sup&gt;b&lt;/sup&gt; at home Lengths not specified</td>
<td>CISS (Nonverbal only); Group 2 Audiotaped</td>
<td>Reliability (alpha): Voice Pos. (.93); Voice Neu. (.997); Voice Neg. (.988) Validity (Content: Generalizability of interaction): Full details on WWW</td>
</tr>
<tr>
<td>Gottman (1979, pp. 278–288)</td>
<td>Type: T (Phase 3 only) 27 married couples (assigned to BCT): C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab Situation: No time limit 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td>CISS</td>
<td>Reliability (alpha): Problem Feeling (.90), 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 Reliability (kappa)&lt;sup&gt;2&lt;/sup&gt;, Reliability (alpha): Pos. Affect (.90), Neu. Affect (.96), Neg. Affect (.99), Validity (content): Full details on WWW</td>
</tr>
<tr>
<td>Gottman (1980)</td>
<td>Type: C 19 ND &amp; 19 D couples: C&lt;sup&gt;1&lt;/sup&gt;, Cl&lt;sup&gt;1&lt;/sup&gt; Marital satisfaction based on marital relationship Inventory; D: &lt;85, ND: &gt;102.</td>
<td>Setting: Lab Situation: (1) 15-min. 6&lt;sup&gt;a&lt;/sup&gt; (2) Length not specified 5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>CISS</td>
<td>Reliability MICS&lt;sup&gt;1&lt;/sup&gt;, SPAFF&lt;sup&gt;f&lt;/sup&gt;, RCISS&lt;sup&gt;4&lt;/sup&gt; Validity CISS (predictive): Full details on WWW Reliability&lt;sup&gt;d&lt;/sup&gt; Validity&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Gottman (1993)&lt;sup&gt;11&lt;/sup&gt;</td>
<td>Type: L 79 married couples: C&lt;sup&gt;3&lt;/sup&gt;, 72 couples at 4-year follow-up</td>
<td>Setting: Lab Situation: 15-min. 1&lt;sup&gt;h&lt;/sup&gt;</td>
<td>MICS-III, SPAFF, RCISS</td>
<td>Reliability MICS&lt;sup&gt;1&lt;/sup&gt;, SPAFF&lt;sup&gt;f&lt;/sup&gt;, RCISS&lt;sup&gt;4&lt;/sup&gt; Validity CISS (predictive): Full details on WWW Reliability&lt;sup&gt;d&lt;/sup&gt; Validity&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Gottman, Cuan, Carrere, &amp; Swanson (1998)&lt;sup&gt;12&lt;/sup&gt;</td>
<td>Type: L 130 newlyweds: C&lt;sup&gt;3&lt;/sup&gt;, 130 couples at 3–6 yr. follow-up. Only analyzed 17 divorced, 20 “stable happily</td>
<td>Setting: Lab Situation: One interaction per year/6 years: 15-min. 1&lt;sup&gt;h&lt;/sup&gt;</td>
<td>SPAFF</td>
<td>Reliability MICS&lt;sup&gt;1&lt;/sup&gt;, SPAFF&lt;sup&gt;f&lt;/sup&gt;, RCISS&lt;sup&gt;4&lt;/sup&gt; Validity CISS (predictive): Full details on WWW Reliability&lt;sup&gt;d&lt;/sup&gt; Validity&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Study</td>
<td>Type of study/ participants</td>
<td>Situations/ Settings</td>
<td>Observation methods</td>
<td>Psychometrics (reliability/ validity)</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>-------------------------------------</td>
</tr>
</tbody>
</table>
| Gottman et al. (1995) | Type: L  
61 PV couples: C<sup>1,2,4,8</sup>  
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.  
½ of couples completed the same interaction at 2-year follow-up. | SPAFF | Reliability<sup>f</sup>  
Validity (discriminative): Full details on WWW |
| Gottman & Krokoff (1989) | Type: L  
Sample 1: 25 married couples: C<sup>2,8</sup>  
Sample 2: 30 married couples: C<sup>1</sup>.  
19 married couples at 3-year follow-up. Sample of Levenson & Gottman, 1983 | Setting: Lab and home  
Situation: (1) Lab, 15-min.  
Ⅰ<sup>2</sup>(2) Home, 15-min.  
Ⅰ<sup>b</sup>  
(Audiotaped Sample 1 only) | SPAFF, CISS, MICS-III | Reliability CISS<sup>f</sup>, SPAFF<sup>f</sup>, MICS<sup>b</sup>  
Validity CISS (concurrent): Full details on WWW  
Validity MICS (concurrent, predictive): Full details on WWW |
| Gottman & Levenson (1992)<sup>11</sup> | Type: L  
79 married couples: C<sup>1</sup>.  
72 couples at 4-year follow-up | Setting: Lab  
Situation: First interaction: 15-min.  
1<sup>a,b</sup>  
4-year follow-up: (1) 15-min 7<sup>Ⅰa,b</sup>; (2) 15-min 1<sup>Ⅰa,b</sup> = Topic: events of last 4 years | MICS-III, SPAFF, RCISS | Reliability MICS<sup>b</sup>; SPAFF<sup>f</sup>; RCISS<sup>f</sup>  
Validity (discriminative, predictive): Full details on WWW |
| Gottman & Levenson (1999d)<sup>11</sup> | Type: L  
79 married couples: C<sup>1</sup>.  
42 couples at 4-year follow-up | Setting: Lab  
Situation: See Gottman & Levenson (1999a) | SPAFF | Reliability<sup>1</sup>  
Validity SPAFF (postdictive): Full details on WWW |
| Gottman & Levenson (1999b)<sup>11</sup> | Type: L  
79 married couples: C<sup>1</sup>.  
42 couples at 4-year follow-up | Setting: Lab  
Situation: See Gottman & Levenson (1999a) | SPAFF | Reliability<sup>1</sup>  
Validity (predictive): Full details on WWW |
| Gottman & Levenson (1999c) | Type: L  
85 married couples: C<sup>1</sup>.  
69 couples at 4-year follow-up | Setting: Lab  
Situation: (1) 15-min.  
1<sup>Ⅰa</sup>, (2) 15-min.  
6<sup>b,c</sup> | SPAFF | Reliability<sup>1</sup>  
Validity (predictive): Full details on WWW |
| Gottman, Markman, & Notarius (1977) | Type: C  
14 ND and 14 D couples: C<sup>1</sup>, Cl<sup>1</sup> | Setting: Lab  
Situation: Length not specified | 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  
Reliability<sup>1</sup>  
Validity (discriminative): Full details on WWW |
| Gottman, Swanson, & Murray (1999)<sup>12</sup> | Type: L  
130 newlyweds: C<sup>1</sup>.  
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.  
Ⅰ<sup>d</sup> | SPAFF | Reliability<sup>1</sup>  
Validity (discriminative): Full details on WWW |
<table>
<thead>
<tr>
<th>Study</th>
<th>Type of study/ participants</th>
<th>Situations/ Settings</th>
<th>Observation methods</th>
<th>Psychometrics (reliability/ validity)</th>
</tr>
</thead>
</table>
| Gray-Little, Baucom, & Hamby (1996) | Type: T 53 D couples (randomly assigned to BCT, BCT + CR, BCT + EET, BCT + CR + EET); C, C1                                               | Setting: Lab                      | MICS                | Reliability (% agreement): Cohort 1: Pos. (87), Neg. (83); Cohort 2: Pos. (90), Neg. (90). Validity  
|                       |                                                                                             | Situation: Sample                 |                     |                                                        |
|                       |                                                                                             | 1: (1) 3–6 min, 1b                |                     |                                                        |
|                       |                                                                                             | (2) 3–5 min. 3b                   |                     |                                                        |
|                       |                                                                                             | Sample 2: Two 7-min. 1b            |                     |                                                        |
| Hahlew et al. (1984)  | Type: C (Study 1); T (Study 2) Study 1: 41 couples (12 ND, 29 D); C1 referrals from people  | Setting: Lab                      | MICS (adapted)      | Reliability  
|                       | Study 2: 85 D couples (34 BCT, 35 CT, 17 WLCG); C1                                           | Situation: Length                 |                     | discriminative): Full details on WWW                  |
|                       |                                                                                             | unspecified 1b                    |                     |                                                        |
| Hahlew, Kaiser, Christensen, Fehm, Wolffsdorf, & Groth (2000) | Type: C 81 couples living together for at least 3 years: C1                                   | Setting: Lab                      | KPI                 | Reliability  
|                       |                                                                                             | Situation: Lab                    |                     | (discriminative, Tx sensitivity): Full details on WWW |
|                       |                                                                                             | 15-min. 1b                        |                     |                                                        |
| Hahlew, Markman, Thurmuer, Engl, & Eckert (1998) | Type: L 96 couples randomly assigned: (64 EPL, 32 WLCG): C1, C1.6, 64 EPL, 29 WLCG at 1.5-year follow-up; (61 EPL, 24 WLCG at 3-year follow-up) | Setting: Lab                      | KPI                 | Validity (Tx sensitivity): Full details on WWW        |
|                       |                                                                                             | Situation: Four                   |                     |                                                        |
|                       |                                                                                             | 10-min. 1b (One at pre, post, 15 |                     |                                                        |
|                       |                                                                                             | years, 3 years)                   |                     |                                                        |
| Hahlew, Revenstorf, & Schindler (1984) | Type: T 85 D* couples: (17 BCT-C, 16 BCT-CG, 16 CT-17 WLCG-C); C1, C1.6 based on participants’ subjective distress | Setting: Lab                      | MICS (adapted)      | Reliability  
|                       |                                                                                             | BCT couples: (1)                  |                     | (discriminative): No significant results reported for commonly defined constructs |
|                       |                                                                                             | 10-min. 5b (2) 10-min. 1c         |                     |                                                        |
| Halford, Sanders, & Behrens (1993) | Type: T 26 couples randomly assigned to (13 BCT and 13 enhanced BCT); C1                        | Setting: Lab and home             | ICS (KPI)           | Reliability  
|                       |                                                                                             | Setting (lab): (1)                 |                     | (Tx sensitivity): Full details on WWW                 |
|                       |                                                                                             | 10-min. 1b — Pre                   |                     |                                                        |
|                       |                                                                                             | (2) 10-min. 1b — Post              |                     |                                                        |
|                       |                                                                                             | home: 10-min. 1b                   |                     |                                                        |
|                       |                                                                                             | — Follow-up Audiotaped             |                     |                                                        |
| Harrell & Guerney (1976) | Type: T 60 couples (30 behavioral-exchange, 30 WLCG); C9                                     | Setting: Lab                      | MICS (adapted)      | Reliability  
|                       |                                                                                             | Situation: Three                   |                     | (Tx sensitivity): No significant results reported for commonly defined constructs. |
|                       |                                                                                             | length not specified 5b            |                     |                                                        |
| Haynes et al. (1992)  | Type: C 26 elderly married couples: C6                                                    | Setting: Home or lab               | CST                 | Reliability: Correlation between the 2 observers’ mean CST scores was .86. Validity (concurrent): Full details on WWW |
|                       |                                                                                             | Situation: 10-min. 1b              |                     |                                                        |
| Haynes, Chavez, & Samuel (1984) | Type: C 67 ND and 33 D couples: C10                                                         | Setting: Lab                      | MICS (adapted)      | Reliability  
<p>|                       |                                                                                             | Situation: 10-min. 1b              |                     | (concurrent): Full details on WWW                      |
|                       |                                                                                             | Audiotaped                        |                     |                                                        |</p>
<table>
<thead>
<tr>
<th>Study</th>
<th>Type of study/ participants</th>
<th>Situations/ Settings</th>
<th>Observation methods</th>
<th>Psychometrics (reliability/ validity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haynes, Follingstad, &amp; Sullivan (1979)</td>
<td>Type: C 6 ND and 7 D couples: C(^1), C(^2)</td>
<td>Setting: Home Situation: 25-min.</td>
<td>MICS (adapted)</td>
<td>Reliability(^1) Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td>Haynes, Jensen, Wise, &amp; Sherman (1981)</td>
<td>Type: C 16 ND and 12 D couples: C(^1)</td>
<td>Setting: Lab Situation: 10-min.</td>
<td>MICS (adapted)</td>
<td>Reliability(^1) Validity(^2)</td>
</tr>
<tr>
<td>Heavey, Christensen, &amp; Malamuth (1995)</td>
<td>Type: L 48 couples (31 married, 17 exclusively dating): C(^1), C(^2)<em>local schools</em> 36 couples at 2.5-year follow-up (31 married, 6 exclusively dating, 1 in original relationship)</td>
<td>Setting: Lab Situation: Two 10-min 1(^{rd})</td>
<td>CRS</td>
<td>Reliability (alpha): Withdrawal (.78), Demand (.78). Validity (concurrent and predictive): Full details on WWW</td>
</tr>
<tr>
<td>Heavey, Layne, &amp; Christensen (1993)</td>
<td>Type: L 29 married couples: C(^1). 19 couples at approximately 1-year follow-up</td>
<td>Setting: Lab Situation: Two 7-min 1(^{rd})</td>
<td>CRS</td>
<td>Validity (alpha): Demand (.88), Withdrawal (.81), Pos. (.84), Neg. (.84). Validity (concurrent): Full details on WWW</td>
</tr>
<tr>
<td>Heyman, Brown, Feldbau-Kohn, &amp; O'Leary (1999)</td>
<td>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</td>
<td>Setting: Lab Situation: 10-min. 1(^{st})</td>
<td>RMICS</td>
<td>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 (alpha): Demand (.78), Withdrawal (.78), Pos. (.78), Neg. (.78). Validity (concurrent): Full details on WWW</td>
</tr>
<tr>
<td>Heyman et al. (in press)</td>
<td>Type: T, C Sample 1: 197 couples presenting for couples therapy: C(^1), 50 happily married control couples: C(^1); Sample 2: 157 engaged couples 3 months prior to wedding: C(^1)</td>
<td>Setting: Lab Situation: 15-min. Sample 1: 1(^{st}); Sample 2: 1(^{st})</td>
<td>RMICS</td>
<td>Validity (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. Validity (alpha): Demand (.88), Withdrawal (.81), Pos. (.84), Neg. (.84). Validity (concurrent): Full details on WWW</td>
</tr>
<tr>
<td>Heyman, Eddy, Weiss, &amp; Vivian (1995)</td>
<td>Type: A 994 couples: All MICS interactions coded at the Oregon Marital Studies Program from 1987 to 1991</td>
<td>Setting: Lab Situation: Varied across investigators Typical procedures: 10–15 min. 1(^{st})</td>
<td>MICS-IV</td>
<td>Validity (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. Validity (alpha): Demand (.88), Withdrawal (.81), Pos. (.84), Neg. (.84). Validity (concurrent): Full details on WWW</td>
</tr>
<tr>
<td>Holtzworth-Munroe, Stuart, Sandin, Smutlizer, &amp; McLaughlin (1997)</td>
<td>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</td>
<td>Setting: Lab Situation: Two 8-min. 5(^{th})</td>
<td>SSICS, SSBARS</td>
<td>Validity SSICS(^5), SSBARS(^5) Validity SSBARS: Validity SSICS: No significant results reported for commonly defined constructs.</td>
</tr>
<tr>
<td>Hooley &amp; Hahlweg (1989)</td>
<td>Type: C Study 1: 12 ND and 29 D couples: C(^2), C(^1) <em>referrals from personal contacts. Sample of Hahlweg et al., 1984 Study 2: 30 married unipolar depressed</em> patients and their spouses: C(^1). Sample of Hooley, 1986. *Met criteria for MDD; all were hospitalized unipolar depressives</td>
<td>Setting: Lab (Study 1), Hospital (Study 2) Situation: Study 1: 10-min. 1(^{st}) Situation: Study 2: 15-min. 5(^{th})</td>
<td>KPI</td>
<td>Reliability Study 1 (kappa): Nonverbal pos. (.82), Nonverbal neg. (.89), Nonverbal neu. (.52). Only included a range for the remaining codes. Validity: Study 2(^7) Validity (concurrent): Full details on WWW</td>
</tr>
<tr>
<td>Study</td>
<td>Type of study/ participants</td>
<td>Situations/ Settings</td>
<td>Observation methods</td>
<td>Psychometrics (reliability/ validity)</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>----------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Iverson &amp; Baucom (1990)</td>
<td>recruited from psychiatric hospitals</td>
<td>48 married couples (randomly assigned to BCT, CR + BCT, BCT + EET; CR + BCT + EET): C</td>
<td>Setting: Lab Situation: Two 7-min. 1b</td>
<td>MICS-III</td>
</tr>
<tr>
<td>Jacob &amp; Krahn (1987)</td>
<td>Type: C 96 married couples: (37 H with alcohol dependence, 27 H with MDD, and 32 with no psychopathology): C1</td>
<td>Setting: Lab Situation: 10–15 min. 1c</td>
<td>MICS (adapted)</td>
<td>Reliabilityb 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).</td>
</tr>
<tr>
<td>Jacob &amp; Krahn (1988)</td>
<td>Type: C 107 couples: (38 H with AD, 35 MDD H, 34 with no psychopathology): C1</td>
<td>Setting: Lab Situation: 9–13 min. 1c1</td>
<td>MICS (adapted)</td>
<td>Reliabilityc Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td>Jacob &amp; Leonard (1988)</td>
<td>Type: C 49 couples with H with AD: C1</td>
<td>Setting: Lab Situation: 10-min. 1b</td>
<td>MICS (adapted)</td>
<td>Reliabilityc Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td>Jacob &amp; Leonard (1992)</td>
<td>Type: C 131 couples: (49 H with AD, 40 H with MDD, and 42 control): C1</td>
<td>Setting: Lab Situation: Two 10 min. 1b,d1</td>
<td>MICS (adapted)</td>
<td>Reliabilityc Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td>Jacobson &amp; Anderson (1980)</td>
<td>Type: T 46 ND and 14 D couples (randomly assigned to BCT + feedback, BCT + behavior rehearsal, BCT + instructions, complete T, WLCG): C1</td>
<td>Setting: Lab Situation: pre and post — (1) Two 10-min. 5c (2) 10-min. 1b</td>
<td>MICS (adapted)</td>
<td>Reliabilityc Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td>Jacobson (1977)</td>
<td>Type: T 10 married couples (randomly assigned to BCT or WLCG): C1</td>
<td>Setting: Lab Situation: Pre: Two 5–10 min. 1c; Post: Two 5–10 min. 1c</td>
<td>MICS (adapted)</td>
<td>Reliabilityc Validity (Tx sensitivity): Full details on WWW</td>
</tr>
<tr>
<td>Jacobson (1978)</td>
<td>Type: T 32 couples (17 BCT, 7 no specific treatment, 6 WLCG): C1,2,1, Cl1</td>
<td>Setting: Lab Situation: Pre: Two 5–10 min. 1c; Post: Two 5–10 min. 1c</td>
<td>MICS (adapted)</td>
<td>Reliabilityc Validity (Tx sensitivity): Full details on WWW</td>
</tr>
<tr>
<td>Jacobson et al. (1989)</td>
<td>Type: T 30 married couples (randomly assigned to structured BCT or flexible T): C1</td>
<td>Setting: Lab Situation: Two 7-min. 1c</td>
<td>KPI</td>
<td>Reliabilityc Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td>Jacobson et al. (1994)</td>
<td>Type: C 92 couples: (60 PV and 32 D/NV): C1,2,4,8 PV; W reporting at least 6 moderate acts, 2 severe acts, or 1 life-threatening act in the past year.</td>
<td>Setting: Lab Situation: Two 15-min. 1b</td>
<td>SPAFF</td>
<td>Reliabilityc Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td>Jacobson, Gottman, Gortner, Berns, &amp; Shortt (1996)</td>
<td>Type: L Original sample: 60 D/ PV couples: C1,4,8, 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)</td>
<td>Setting: Lab Situation: 15-min. 1b</td>
<td>SPAFF</td>
<td>Reliabilityc Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td>Johnson &amp; Bradbury (1999)</td>
<td>Type: L 65 newlywed couples: C1</td>
<td>Setting: Lab Situation: 15-min. 1b</td>
<td>VTCS (adapted)</td>
<td>Reliabilityc Validity (predictive) Full details on WWW</td>
</tr>
<tr>
<td>Study</td>
<td>Type of study/participants</td>
<td>Situations/Settings</td>
<td>Observation methods</td>
<td>Psychometrics (reliability/validity)</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Johnson &amp; Jacob (1997)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>57 couples at 6-mo. follow-up, 54 couples at 1-year follow-up</td>
<td>Setting: Lab Situation: 10–15 min.</td>
<td>MICS (adapted)</td>
<td>Reliability&lt;sup&gt;f&lt;/sup&gt; Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td>Johnson &amp; Jacob (2000)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>141 couples: (50 H with MDD, 41 W with MDD, and 50 non-MDD): C&lt;sub&gt;1&lt;/sub&gt;</td>
<td>Setting: Lab Situation: 15-min.</td>
<td>MICS (adapted)</td>
<td>Reliability&lt;sup&gt;f&lt;/sup&gt; Validity (concurrent): Full details on WWW</td>
</tr>
<tr>
<td>Julien, Markman, &amp; Lindahl (1989)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>135 premarital ND couples: C&lt;sub&gt;4&lt;/sub&gt;; 59 couples at 4-year follow-up (24 not in original relationship, 5 separated/divorced, 30 did not complete all phases)</td>
<td>Setting: Lab Situation: Two 10–15 min.</td>
<td>IDCS, CISS</td>
<td>Reliability&lt;sup&gt;k&lt;/sup&gt; Validity (convergent, predictive): Full details on WWW</td>
</tr>
<tr>
<td>Kaiser, Hahlweg, Fehn, Wolfsdorf, &amp; Groth (1998)</td>
<td>67 couples living together for at least 3 years: C&lt;sub&gt;1&lt;/sub&gt; Randomly assigned: (31 EPL II, 36 WLCG).</td>
<td>Setting: Lab Situation: 15-min.</td>
<td>KPI</td>
<td>Reliability&lt;sup&gt;k&lt;/sup&gt; Validity (Tx sensitivity): Full details on WWW</td>
</tr>
<tr>
<td>Karney &amp; Bradbury (1993)</td>
<td>56 newlywed couples: C&lt;sub&gt;1&lt;/sub&gt; Follow-up every 6 mo. for 4 years; 38 couples at 4-year follow-up</td>
<td>Setting: Lab Situation: 15-min.</td>
<td>VTCS (adapted), Audiotaped</td>
<td>Reliability&lt;sup&gt;k&lt;/sup&gt; Validity (concurrent, predictive): Full details on WWW</td>
</tr>
<tr>
<td>Katz &amp; Gottman (1993)</td>
<td>53 couples at 3-year follow-up</td>
<td>Setting: Lab Situation: 15-min</td>
<td>SPAFF</td>
<td>Reliability&lt;sup&gt;k&lt;/sup&gt; Validity (convergent, predictive): Full details on WWW</td>
</tr>
<tr>
<td>Kiecolt-Glaser et al. (1993)</td>
<td>90 newlywed couples: R</td>
<td>Setting: Lab Situation: 30-min.</td>
<td>MICS-IV</td>
<td>Reliability&lt;sup&gt;g&lt;/sup&gt; Validity (discriminative): Neg., W &gt; H; Avoidance, H &gt; W; Humor, W &gt; H</td>
</tr>
<tr>
<td>Kiecolt-Glaser et al. (1996)</td>
<td>90 newlywed couples: R</td>
<td>Setting: Lab Situation: 30-min.</td>
<td>MICS-IV</td>
<td>Reliability&lt;sup&gt;g&lt;/sup&gt; Validity (convergent): Full details on WWW</td>
</tr>
<tr>
<td>Kobak &amp; Hazan (1991)</td>
<td>40 married couples: C&lt;sub&gt;1&lt;/sub&gt;,&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Setting: Lab Situation: (1) 10-min. (2) Two 7-min.</td>
<td>IDCS (adapted)</td>
<td>Reliability (intrerrater rs): Rejection (.71), Support-Validation (.77)</td>
</tr>
<tr>
<td>Kroff, Gottman, &amp; Haas (1989)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>52 married couples: R</td>
<td>Setting: Lab and home Situation: (1) 15-min. — Lab. (2) 15-min.</td>
<td>CISS, SPAFF, RCIS</td>
<td>Reliability CISS (kappa&lt;sup&gt;f&lt;/sup&gt;, SPAFF (kappa&lt;sup&gt;f&lt;/sup&gt;, RCIS (Pearson correlation)&lt;sup&gt;j&lt;/sup&gt;, Reliability CISS (Alpha): H pos. (.95), H Neu. (.97), H Neg. (.91), W Pos. (.95), W Neu. (.98), W Neg. (.90) Validity RCIS and CISS (convergent): Full details on WWW</td>
</tr>
<tr>
<td>Kroff, Gottman, &amp; Roy (1988)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>52 married couples (13 D/H white-collar occupation, 13 ND/H white-collar, 13 D/H</td>
<td>Setting: Home Situation: 15-min.</td>
<td>CISS Audiotaped</td>
<td>Reliability (alpha): H Pos. (.95), H Neu. (.97), W Pos. (.95), W Neu. (.98), W Neg. (.90) Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td>Study</td>
<td>Type of study/participants</td>
<td>Situations/Settings</td>
<td>Observation methods</td>
<td>Psychometrics (reliability/validity)</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------</td>
<td>------------------</td>
<td>------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Leonard &amp; Jacob (1997)</td>
<td>blue-collar, 13 ND/H blue-collar): R</td>
<td>Setting: Lab</td>
<td>MICS (adapted)</td>
<td>Reliability: No significant differences between episodic versus steady AD groups on MICS behaviors</td>
</tr>
<tr>
<td></td>
<td>Type: C</td>
<td>Situation: Length not specified 1&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>99 married couples: (50 control, 30 H with steady AD, 19 H with episodic AD): C&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leonard &amp; Roberts (1998)</td>
<td>135 newlywed couples; (60 PV and 75 NV): C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab</td>
<td>MICS-IV</td>
<td>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: Reliability&lt;sup&gt;4&lt;/sup&gt; Validity (Tx sensitivity): Full details on WWW</td>
</tr>
<tr>
<td></td>
<td>PV: At least 2 moderate acts since marriage, at least 1 severe act since marriage, or at least 1 moderate act prior to marriage</td>
<td>Situation: (1) 15-min. 1&lt;sup&gt;b&lt;/sup&gt; (2) 5-min. 3&lt;sup&gt;b&lt;/sup&gt; (3) 15-min. 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9 married couples (4 BCT, 5 IT): C&lt;sup&gt;1&lt;/sup&gt;, C&lt;sup&gt;1&lt;/sup&gt;-&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Situation: (1) 10-min. 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lindahl, Clements, &amp; Markman (1997)&lt;sup&gt;10&lt;/sup&gt;</td>
<td>Type: L</td>
<td>Setting: Lab</td>
<td>IDCS</td>
<td>Reliability (% agreement): Pos. Behaviors (68%), Neg. Behaviors (57%), Neu. Behaviors (74%); Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td></td>
<td>25 premarital ND couples: C&lt;sup&gt;1&lt;/sup&gt;-&lt;sup&gt;4&lt;/sup&gt;, 25 married couples at 5-year follow-up.</td>
<td>Situation: Lab</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Margolin (1978a)&lt;sup&gt;14&lt;/sup&gt;</td>
<td>Type: C</td>
<td>Setting: Lab</td>
<td>MICS</td>
<td>Reliability (% agreement): Pos. Behaviors (68%), Neg. Behaviors (57%), Neu. Behaviors (74%); Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td></td>
<td>27 married D couples: C&lt;sup&gt;1&lt;/sup&gt;-&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Situation: Two 10-min. 1&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Margolin (1978b)&lt;sup&gt;14&lt;/sup&gt;</td>
<td>Type: C (assessment phase of T study)</td>
<td>Setting: Lab</td>
<td>MICS</td>
<td>Reliability (% agreement): Pos. Behaviors (68%), Neg. Behaviors (57%), Neu. Behaviors (74%); Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td></td>
<td>27 married D couples: C&lt;sup&gt;1&lt;/sup&gt;-&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Situation: Two 10-min. 1&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17 ND and 22 D couples: C&lt;sup&gt;1&lt;/sup&gt;, C&lt;sup&gt;9&lt;/sup&gt;</td>
<td>Situation: Two 10-min. 1&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Margolin &amp; Weiss (1978a, 1978b)&lt;sup&gt;14&lt;/sup&gt;</td>
<td>Type: T</td>
<td>Setting: Lab</td>
<td>MICS</td>
<td>Reliability (% agreement): Pos. Behaviors (68%), Neg. Behaviors (57%), Neu. Behaviors (74%); Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td></td>
<td>27 married D couples (randomly assigned to BCT, Attitudinal Restructuring + BCT, No specific T): C&lt;sup&gt;1&lt;/sup&gt;-&lt;sup&gt;2&lt;/sup&gt;</td>
<td>Situation: Two 10-min. 1&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Markman, Renick, Floyd, Stanley, &amp; Clements (1993)&lt;sup&gt;3&lt;/sup&gt;</td>
<td>Type: T</td>
<td>Original sample: 114 couples: C&lt;sup&gt;9&lt;/sup&gt;, 4-year follow-up: (15 PREP, 24 WLCG); 5-year follow-up: (12 intervention, 18 WLCG)</td>
<td>Setting: Lab</td>
<td>IDCS</td>
</tr>
<tr>
<td></td>
<td>Situation: Two 10–15 min. 1&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matthews, Wickrama, &amp; Conger (1996)</td>
<td>Type: L</td>
<td>Setting: Home</td>
<td>IFIRS</td>
<td>Reliability (ICC): Hostility (H = .72, W = .73), Angry-Coercive (H = .57, W = .56), Antisocial (H = .60, W = .59); Validity&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>436 married couples: 436 couples sent home to seventh graders in school. 436 couples at 5-year follow-up</td>
<td>Situation: (1) 25-min. 7&lt;sup&gt;a&lt;/sup&gt; per year/4 years. *Topics: Enjoyable times, relationships, and conflict.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McKnight, Nelson-Gray, &amp; Gulluck (1989)</td>
<td>Type: C</td>
<td>Setting: Lab</td>
<td>MICS</td>
<td>Reliability: Validity (discriminative): Full details on WWW</td>
</tr>
<tr>
<td></td>
<td>24 couples: (8 ND, 8 D, and 8 bipolar patients with varying levels of marital adjustment): C&lt;sup&gt;3&lt;/sup&gt;</td>
<td>Situation: (1) 10-min. 6&lt;sup&gt;a&lt;/sup&gt; (b) 2) 10-min. 1&lt;sub&gt;ab&lt;/sub&gt;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Studies and details on WWW
<table>
<thead>
<tr>
<th>Study</th>
<th>Type of study/ participants</th>
<th>Situations/ Settings</th>
<th>Observation methods</th>
<th>Psychometrics (reliability/ validity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melby, Conger, Ge, &amp; Warner (1995)</td>
<td>Type: L</td>
<td>Setting: Home</td>
<td>IFIRS</td>
<td>Reliability&lt;sup&gt;a&lt;/sup&gt; Validity&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>412 married couples: C-</td>
<td>Situation: 20-min.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>letters sent home to</td>
<td>7&lt;sup&gt;a&lt;/sup&gt; Topic:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>seventh graders in local</td>
<td>Enjoyable events</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>schools. 391 at 1-year</td>
<td>and disagreements in</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>follow-up</td>
<td>past year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mendolia, euch. &amp; Tesser (1996)</td>
<td>Type: C</td>
<td>Setting: Lab</td>
<td>MICS-IV</td>
<td>Reliability&lt;sup&gt;1&lt;/sup&gt; Validity&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>66 married couples: C&lt;sup&gt;1,7&lt;/sup&gt;</td>
<td>Situation: 20-min. 1&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miller &amp; Bradbury (1995)</td>
<td>Type: C</td>
<td>Setting: Lab</td>
<td>VTCS</td>
<td>Reliability VTCS&lt;sup&gt;2,4&lt;/sup&gt;, Reliability SSICS&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>60 newly wed couples: C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Situation: (1) 15-min. 1&lt;sup&gt;p&lt;/sup&gt; (2)</td>
<td>SSICS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Two 10-min. 5&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miller, Dopp, Myers, Stevens, &amp; Failey (1999)</td>
<td>Type: C</td>
<td>Setting: Lab</td>
<td>SPAFF</td>
<td>Reliability&lt;sup&gt;1&lt;/sup&gt; Validity&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>41 married couples: C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Situation: 15-min. 1&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18 high hostility (Ho),</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18 low Ho groups based</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>on the Cook-Medley Hostility scale (Cook &amp; Medley, 1954)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Murphy &amp; O’Farrell (1997)</td>
<td>Type: C</td>
<td>Setting: Lab</td>
<td>MICS-IV</td>
<td>Reliability&lt;sup&gt;a&lt;/sup&gt; Validity&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>90 married couples: (60 PV and 30 NV; all with H with AD): C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Situation: 10-min. 1&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nelson &amp; Beach (1990)</td>
<td>Type: C</td>
<td>Setting: Lab</td>
<td>KPI (adapted)</td>
<td>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&lt;sup&gt;k&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>60 married couples: (20 ND/Non-MDD, 20 D/Non-MDD, and 20 D/MOD): C&lt;sup&gt;1&lt;/sup&gt;, C&lt;sup&gt;3&lt;/sup&gt;, C&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Approximately 10-min. 1&lt;sup&gt;i&lt;/sup&gt;</td>
<td>Audiotaped</td>
<td></td>
</tr>
<tr>
<td>Newton, Kiecolt, Glaser, Glaser, &amp; Malarkey (1995)</td>
<td>Type: C 83 newlywed couples: C&lt;sup&gt;10&lt;/sup&gt;</td>
<td>Setting: Hospital</td>
<td>MICS-IV</td>
<td>Reliability&lt;sup&gt;a&lt;/sup&gt; Validity&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Situation: 30-min. 1&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noel, McCrady, Stout, &amp; Fisher-Nelson (1991)</td>
<td>Type: C</td>
<td>Setting: Hospital</td>
<td>MICS</td>
<td>Reliability&lt;sup&gt;a&lt;/sup&gt; Validity&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>45 married couples (12 W with AD, 33 H with AD): C&lt;sup&gt;1,2,6, Cl&lt;sup&gt;3&lt;/sup&gt;&lt;/sup&gt;</td>
<td>Situation: 10-min. 1&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noller (1982)</td>
<td>Type: C</td>
<td>Setting: Lab</td>
<td>CISS, CSIC</td>
<td>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&lt;sup&gt;k&lt;/sup&gt;, CSIC&lt;sup&gt;k&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>48 married couples (16 D, 16 ND, 16 moderate adjustment): C&lt;sup&gt;1,3,5&lt;/sup&gt;, C&lt;sup&gt;1,1&lt;/sup&gt;</td>
<td>Setting: Lab</td>
<td>CISS</td>
<td>Reliability&lt;sup&gt;1&lt;/sup&gt; Validity&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>*referrals from other people</td>
<td>Situation: 10-min. 7&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D: MAT &gt; 120; ND: &lt;95; Moderate Adj.: MAT = 95–120</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notarius &amp; Johnson (1982)</td>
<td>Type: C</td>
<td>Setting: Lab</td>
<td>CISS</td>
<td>Reliability&lt;sup&gt;4&lt;/sup&gt; Validity&lt;sup&gt;4&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>6 ND couples: C&lt;sup&gt;9&lt;/sup&gt;</td>
<td>Situation: 30-min. time limit 1&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notarius, Benson, Sloan, Vanzetti, &amp; Hornyk (1989)</td>
<td>Type: C</td>
<td>Setting: Lab</td>
<td>SPAFF (adapted)</td>
<td>Reliability&lt;sup&gt;1&lt;/sup&gt; Validity&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>18 couples (9 ND, 9 D): C&lt;sup&gt;1,2&lt;/sup&gt;, Cl</td>
<td>Situation: Two maximum of 20-min 1&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Audiotaped</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Note: One partner talks at a time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’Farrell &amp; Birchler (1987)</td>
<td>Type: C</td>
<td>Setting: Lab</td>
<td>MICS</td>
<td>Reliability&lt;sup&gt;1&lt;/sup&gt; Validity&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Situation: 10-min. 1&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Type of study/participants</td>
<td>Situations/Settings</td>
<td>Observation methods</td>
<td>Psychometrics (reliability/validity)</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
</tbody>
</table>
| O’Farrell, Cutter, & Floyd (1985) | Type: T 34 couples with H with AD (10 BCT, 12 IT, 12 WLCG); C1 | Setting: Lab Situation: (1) 10-min. 1" (2) 10-min. 5" | MICS | Reliability
| Pasch & Bradbury (1998) | Type: L 60 newlywed couples: C1, 57 couples at 24-mo. follow-up | Setting: Lab Situation: (1) 15-min. 1 (2) Two 10-min. 5 | VTCS, SPAFF | Reliability
| Pasch, Bradbury, & Davila (1997) | Type: C 60 newlywed couples: C1 | Setting: Lab Situation: Two 10-min. 5 | SSICS Audiotaped | Validity
| 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. 1b/c | RCISS | Reliability
| Patterson, Hops, & Weiss (1975) | Type: T 10 difficult couples (e.g., divorced, separated, affairs, prolonged conflict) assigned to BCT; C1 | Setting: Lab Situation: Four approximately 10-min. 1" | MICS | Reliability
| Revenstorf, Halilweg, Schindler, & Vogel (1984) | Type: T 40 couples: (20 BCT, 10 WLCG, 10 ND): C1 | Setting: Lab Situation: 15-min. 5" | MICS (Adapted) | Validity
| Revenstorf, Vogel, Wegener, Halilweg, & Schindler (1980) | Type: C 10 ND and 10 D couples: C1 | Setting: Lab Situation: 10-min. 5" | MICS (Adapted) | Reliability
| Robinson & Price (1980) | Type: C 8 couples (4 ND and 4 D couples): C1 | Setting: Home Situation: Two 1-hr 1b | MICS (adapted) | Reliability
| | | | | |

Psychol Assess. Author manuscript; available in PMC 2006 April 12.
<table>
<thead>
<tr>
<th>Study</th>
<th>Type of study/ participants</th>
<th>Situations/ Settings</th>
<th>Observation methods</th>
<th>Psychometrics (reliability/ validity)</th>
</tr>
</thead>
</table>
| Rogers, Castleton, & Lloyd (1996) | Type: C 25 married couples: C<sup>1</sup> | Setting: Lab  
Situation: (1) 10-min. 7<sup>)*</sup> (2) 10-min. 2 (3) 10-min. 1<sup>)*</sup> (4) 10-min. 7<sup>**)</sup>  
*Topic: How they met and why they married.  
**Topic: What it takes to have a good marriage | RCCCS Audiotaped | Reliability<sup>h</sup> Validity<sup>k</sup> |
| Rogge & Bradbury (1999)<sup>2</sup> | Type: L 60 newlywed couples: C<sup>1</sup> 56 couples (22 ND, 16 D, 18 divorced/separated) at 4-year follow-up | Setting: Lab  
Situation: 15-min. 1<sup>)*</sup> | 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)<sup>3</sup> | Type: C 40 undergraduate judges rated 24 couples (12 D, 12 ND) on videotape: C<sup>1</sup> | Setting: Lab  
Situation: 10-min. 5<sup>)*</sup> | MICS | 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<sup>1</sup> | Setting: Lab  
Situation: 15-min. 1<sup>)*</sup> | Developed by Gotlib & Kowalik (1985), CISS | Reliability (Gotlib & Kowalik)<sup>g</sup> Validity (Gotlib & Kowalik)<sup>h</sup>, Reliability CISS (discriminative): Full details on WWW |
| Sagrestano, Heavey, & Christensen (1999) | Type: C 42 married couples: C<sup>4</sup> | Setting: Lab  
Situation: Two 10-min. 1<sup>)*</sup> | CRS | Reliability CISS<sup>g</sup>, Validity CISS (discriminative): Full details on WWW |
| Sayers & Baucom (1991)<sup>4</sup> | Type: C 60 D couples: C<sup>1</sup> | Setting: Lab  
Situation: Two 7-min. 1<sup>)*</sup> | MICS-III | Reliability<sup>l</sup> Validity (concurrent): Full details on WWW. Reliability<sup>l</sup> Validity (Tx sensitivity, concurrent): Full details on WWW |
| Sayers, Baucom, Sher, Weiss, & Heyman (1991)<sup>4</sup> | Type: T 60 D couples (48 BCT, 12 WLCG): C<sup>1</sup> | Setting: Lab  
Situation: Two 7-min. 1<sup>)*</sup> | MICS-III | Reliability<sup>l</sup> Validity MICS<sup>g</sup>, MICS<sup>g</sup> Validity (discriminative): Full details on WWW, Validity CISS (discriminative): Full details on WWW |
| Schaap (1984) | Type: C 27 married couples (9 D, 9 ND, 9 conflict): C<sup>1</sup>, CI<sup>1</sup> | Setting: Lab  
Situation: (1) 5-min. 7<sup>)*</sup> (2) 25-min. 1<sup>)*</sup>  
*Discussion regarding beginning of relationship. | MICS (Adapted), CISS-only AC used Audiotaped | Reliability MICS<sup>g</sup> 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<sup>1</sup>, CI<sup>1</sup> | Setting: Lab  
Situation: (1) 5-min. 7<sup>)*</sup> (2) 25-min. 1<sup>)*</sup>  
*Discussion regarding beginning of relationship. | CISS, MICS (Adapted) | Reliability CISS<sup>l</sup>, MICS<sup>g</sup> 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<sup>1</sup>, CI<sup>2</sup> | Setting: Lab  
Situation: 10-min. 7<sup>)*</sup>  
*Topic: Impact of polysubstance use on the relationship. | MICS-IV | Reliability<sup>l</sup> 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<sup>1</sup>, CI<sup>2</sup>, CI<sup>3</sup> | Setting: Lab  
Situation: (1) 5–10 min. 2 (2) Two 7-min. 1<sup>)*</sup> | KPI | Reliability<sup>l</sup> Validity (discriminative, content): Full details on WWW |
| Schmaling et al. (1996) | Type: C | Setting: Lab  
Situation: (1) 7-min. 1<sup>)*</sup> (2) time | LIFE | Reliability<sup>l</sup> Validity. No correlations between LIFE constructs |
<table>
<thead>
<tr>
<th>Study</th>
<th>Type of study/ participants</th>
<th>Situations/ Settings</th>
<th>Observation methods</th>
<th>Psychometrics (reliability/ validity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sher, Baucom, &amp; Larus (1990)</td>
<td>6 married patients suffering from asthma and their partners: Cl 4</td>
<td>unspecified 7*</td>
<td>Setting: Lab Situation: Two 7-min. 1*</td>
<td>MICS-III (Aversive, Dysphoric, Facilitative, Problem Solving) were significant given the extraordinarily low power of an n = 6 sample Reliability 7* Validity 8</td>
</tr>
<tr>
<td>Shoham, Rohrbaugh, Stickle, &amp; Jacob (1998)</td>
<td>47 couples: (14 D/with a partner elevated depressive symptomatology, 12 D/ Psychopathology, 9D, 12 WLCG): C 1</td>
<td>Topic: discuss a recent asthma attack (Aversive, Dysphoric, Facilitative, Problem Solving) were significant given the extraordinarily low power of an n = 6 sample Reliability 8 Validity 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snyder, Wills (1989)</td>
<td>63 couples with male with AD (37 married, 26 living together for at least 1 year): Cl 2</td>
<td>Setting: Lab Situation: Two 7-min. 1*</td>
<td>CISS Reliability 8 Validity (Tx sensitivity): Full details on WWW</td>
<td></td>
</tr>
<tr>
<td>Snyder, Mangrum, &amp; Wills (1993)</td>
<td>55 couples (originally 29 BCT, 30 insight-oriented couples therapy): C 1</td>
<td>Setting: Lab Situation: Two 7-min. 1*</td>
<td>CISS Reliability 8 Validity (Tx sensitivity/ predictive validity): Full details on WWW</td>
<td></td>
</tr>
<tr>
<td>Snyder, Trull, &amp; Wills (1987)</td>
<td>42 couples (30 clinic, 12 nonclinic): Cl 1, C 1,4</td>
<td>Setting: Lab Situation: Two 7-min. 1*</td>
<td>CISS Reliability 8 Validity (convergent): Full details on WWW</td>
<td></td>
</tr>
<tr>
<td>Stein, Giordo, &amp; Dotzenroth (1982)</td>
<td>26 married couples: C 1,2</td>
<td>Setting: Lab Situation: Ten-min. 1*</td>
<td>MICS Reliability 1 Validity (convergent): Full details on WWW</td>
<td></td>
</tr>
<tr>
<td>Vincent, Friedman, Nugent, &amp; Messerly (1979)</td>
<td>20 ND and 20 D couples: C 8</td>
<td>Setting: Lab Situation: (1) 10-min. 5*: Neu. (2) 10-min. 5*: Faking good (3) 10-min. 5*: Faking bad</td>
<td>MICS Reliability 1 Validity (discriminative): Pos. Behavior, ND &gt; D</td>
<td></td>
</tr>
<tr>
<td>Vincent, Weiss, &amp; Birchler (1975)</td>
<td>12 ND and 12 D couples: C 1</td>
<td>Setting: Lab Situation: (1) 4-min. 3* (2) length not specified 5*</td>
<td>MICS Reliability 1 Validity (discriminative): Pos. Behavior, ND &gt; D</td>
<td></td>
</tr>
<tr>
<td>Walker, Johnson, Manion, &amp; Cloutier (1996)</td>
<td>32 married couples: (randomly assigned to 16 EFT, 16 WLCG): C 8*: *letters and phone calls to parents with ill children at a pediatric hospital</td>
<td>Setting: Lab Situation: 15-min. 1* Note: Eight (four for each group) interactions were not coded.</td>
<td>CST Reliability 1 Validity (Tx sensitivity): Across time, Neg. Communication EFT &lt; WLCG.</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Type of study/participants</td>
<td>Situations/Settings</td>
<td>Observation methods</td>
<td>Psychometrics (reliability/validity)</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Weiss, Hops, &amp; Patterson (1973)</td>
<td>Type: T 5 couples assigned to BCT; C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Unspecified Situation: Unspecified</td>
<td>MICS</td>
<td>Validity MICS-G (discriminative, concurrent); Full details on WWW</td>
</tr>
<tr>
<td>Wieder &amp; Weiss (1980)</td>
<td>Type: C 14 married couples: Cl&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab Situation: Two 10-min. 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>MICS, Audiotaped</td>
<td>Reliability&lt;sup&gt;1&lt;/sup&gt; Validity (Tx sensitivity); Full details on WWW</td>
</tr>
<tr>
<td>Wilson, Bornstein, &amp; Wilson (1988)</td>
<td>Type: T 15 D couples: (5 CG, 5 C, 5 WLCG): C&lt;sup&gt;1&lt;/sup&gt;</td>
<td>Setting: Lab Situation: Length unspecified 1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>MICS, Audiotaped</td>
<td>Reliability&lt;sup&gt;1&lt;/sup&gt; Validity (generalizability); Full details on WWW</td>
</tr>
<tr>
<td>Winkin &amp; Rose (1978)</td>
<td>Type: T 14 D couples (assigned to BCT): C&lt;sup&gt;1&lt;/sup&gt;, Cl&lt;sup&gt;1&lt;/sup&gt;, C&lt;sup&gt;2&lt;/sup&gt;. Based on revised-MAT with cut-off scores with a range of 105–110 (Kimmel &amp; VanDerVeen, 1974)</td>
<td>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.</td>
<td>MICS</td>
<td>Reliability&lt;sup&gt;1&lt;/sup&gt; Validity&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

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 few than three studies were censored. Because validity inferces 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.


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

Participants: Sampling used: Community: C<sup>1</sup> = advertising—newspapers; C<sup>2</sup> = advertising—TV, radio; C<sup>3</sup> = advertising—flyers; C<sup>4</sup> = advertising—other; C<sup>5</sup> = church; C<sup>6</sup> = nonchurch community group; C<sup>7</sup> = college students; C<sup>8</sup> = partly through random sampling; C<sup>9</sup> = unspecified; C<sup>10</sup> = marriage licenses; C<sup>11</sup> = representative sample (recruited from a specified sampling frame, with all appropriate participants having an equal chance of inclusion); C<sup>12</sup> = clinical, outpatient, marital treatment; C<sup>13</sup> = clinical, outpatient, other treatment; C<sup>14</sup> = clinical, inpatient, marital treatment; C<sup>15</sup> = 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); N = 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 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 unbranched 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); Gottlieb & Kowalic (1985) coding system: (1); IBRS = Interpersonal Behavior Rating System: (1); ICS (KPI) =

Psychol Assess. Author manuscript; available in PMC 2006 April 12.
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.