Published in final edited form as:

J Fam Psychol. 2012 June ; 26(3): 316–327. doi:10.1037/a0028319.

Linking questionnaire reports and observer ratings of young couples' hostility and support

Frederick O. Lorenz,
Iowa State University

Janet N. Melby,
Iowa State University

Rand D. Conger, and
University of California, Davis

Florenzia F. Surjadi
University of Northern Illinois

Abstract

Past studies have correlated observer ratings with questionnaire self- and partner-reports of behaviors in close relationships. However, few studies have actually proposed and tested longitudinal models that link observer ratings to past behaviors and to questionnaire self- and partner-reports of behaviors during an observational task. Using data from a panel of 324 young couples, we demonstrate that (1) observer ratings of hostility and support are significantly related to couple reports of the same behavior in the relationship two years earlier, and (2) respondent and partner questionnaire reports of hostility and support during the observational task converge with observer ratings of the same behavior even after controlling for earlier self- and partner-reports. These findings demonstrate that observer reports based on brief discussion tasks reflect the tenor of the relationship over a relatively long period of time. They also demonstrate that couple reports of hostile interactions reflect observable behaviors beyond that attributed to earlier self- and partner-reports. Consistent with previous research, effect sizes are larger for hostility than support but there are few differences between men and women.

Two contrasting data collection methods are frequently used when documenting behaviors in close relationships. One method is to observe and record behaviors; the other is to ask participants to report their own behaviors, or the behaviors of others, by responding to questionnaire items in structured surveys (Olson, 1977). Both methods have their advocates and critics. Questionnaire reports are relatively inexpensive and can be easily incorporated into representative sample surveys (Amato, 2007). However, they are often met with skepticism, not only by family scholars (e.g., Miller, Perlman, & Brehm, 2007; Wampler & Halverson, 1993) but also by scholars from other disciplines. In a review of Laumann, Gagnon, Michael and Michaels' *The Social Organization of Sexuality* (1995), geneticist Lewontin (1995) pointed to the problem of self-report as "the fundamental methodological difficulty" facing social and behavioral researchers: "How do we know what is true if we must depend on the interested party to tell us?"

Observer ratings, in contrast, promise direct evidence of visible behaviors (Baumeister, Vohs, & Funder, 2007; Gottman & Notarius, 2000), but the laboratories or observational

settings used to elicit and record behaviors are often deemed artificial, contrived and unrelated to the “natural” interactions of everyday life. Again, this skepticism is expressed not only by family scholars (Olson, 1977), but more generally from a range of disciplines (e.g., Zelditch, 1969). Observers can be trained to classify and calibrate visible behaviors, but they are seldom privy to the everyday world in which individuals interact with each other (Noller & Callan, 1988; Olson, 1977).

Given these broad expressions of skepticism, we examine the correspondence between questionnaire self-reports and observer ratings by addressing two related questions: (1) Do observer ratings of behaviors during an observational task reflect interactions that are acknowledged by couples themselves at an earlier point in their relationship, and (2) Do responses to questionnaire items mirror observable behaviors, or are they primarily reflections of the past histories, sentiments, and attributions individuals recall about themselves and their partners? We frame answers to these two questions by testing a model with two waves of panel data from a sample of young couples.

Background

Skeptics express concern about questionnaire reports and observer ratings because researchers seldom have infallible measures of theoretically important concepts. In the absence of unequivocal “gold standards,” researchers rely on consistency among multiple measures of the same concepts to provide evidence of convergent validity (Campbell & Fiske, 1959; Campbell & Russo, 2001). In family research, consistency is often established by correlating two or more reports of the same behavior from close family members; for example, husbands and wives may report on their own and their partners’ or children’s behavior (Aquilino, 2005; Janssens, DeBruyn, Manders, & Scholte, 2005; Konold & Pianta, 2009; Mikelson, 2008; Rhoades & Stocker, 2006; Saffrey, Bartholomew, Scharfe, Henderson, & Koopman, 2003). These “insider” reports are sometimes complemented by “outsider” reports, where self-, spouse-, or child-reports are corroborated by non-family members, often trained observers who rate visible behaviors (Floyd & Markman, 1983; Furman, Jones, Buhrmester, & Adler, 1989; Hampson, Beavers, & Hulgus, 1989; Melby, Conger & Puspitawati, 1999; Noller & Callan, 1988; Olson, 1977).

In past research, correlations between family member (insider) reports of specific behaviors have been relatively strong, probably because family members often have long-shared histories (Noller & Callan, 1988; Olson, 1977) and because they are asked to respond to inventories of similarly worded questionnaire items having similar response categories (Melby, Conger, Ge, & Warner, 1995). In contrast, correlations between questionnaire reports and observer ratings have often been weak. This is especially well-documented in studies of children (Coie & Dodge, 1988; Feinberg, Neiderhiser, Howe, & Hetherington, 2001; Furman et. al., 1989; Pellegrini & Bartini, 2000; Schwarz, Barton-Henry, & Pruzinsky, 1985). For example, Achenbach, McConaughy, and Howell’s (1987) meta-analysis estimated average correlations of 0.270 between parents and observers of children and adolescents. This same pattern of weak correlations has been found between adults reports of themselves or their partners and observer ratings of personality traits (Bernieri, Zuckerman, Koestner, & Rosenthal, 1994), behaviors such as dominance and friendliness (Moskowitz, 1990), and patterns of communication (Floyd & Markman, 1983; Rhoades & Stocker, 2006).

There are a number of possible reasons for the weak correlations between questionnaire reports and observer ratings. One reason may be that questionnaire items and observer ratings reflect different behaviors or different dimensions of the same behavior. Indeed, observer ratings and questionnaire responses have seldom been based on exactly the same

interactions at the same point in time. Instead, most questionnaire reports have been based on a general recall of past behaviors under what Sanford (2010) refers to as “context-general” circumstances, whereas observer ratings are based on “context-specific” behaviors, or behaviors witnessed in a specific situation at a specific point in time, (e.g., during an observational task). Testing the proposition that *the more similar the context, the higher the correspondence between observer ratings and questionnaire reports*, Lorenz, Melby, Conger, and Xu (2007) reported research in which the context-general and context-specific questionnaire items were carefully established by question preambles. They found that the correlations between observer ratings and questionnaire items measuring hostility were nearly twice as large (0.59 – 0.62) when the preamble to the questionnaire items specified the recently completed observational task as compared to the situation where the preamble asked respondents to recall hostile behaviors over a broader span of time and circumstances (“during the past month”). Similarly, Sanford (2010) reported correlations between questionnaire reports and observer ratings of the same context of between 0.57 and 0.81, about as high as correlations between self and partner (0.63 – 0.75) reports of adversarial and collaborative engagement.

Although these recent studies advance our understanding of the strength of relations between the two methods of collecting information, further progress can be made by proposing and testing a model which addresses the skeptics’ criticisms of questionnaire reports and observer ratings. First, most previous studies have been cross-sectional (e.g., Lorenz et al., 2007) or contain only brief time intervals between assessments (e.g., Sanford, 2010). In this model, we prospectively examine the convergence of questionnaire reports and observer ratings over two waves of data collection, as indicated by Time 1 and Time 2 in Figure 1. This is important for substantive purposes because longer-term effects of behaviors on relationship outcomes are known to be different than the short-term effects of the same behaviors (e.g., Christensen & Heavey, 1990; Gottman & Krokoff, 1989). Methodologically, correlations that are separated by time provide more convincing evidence of convergence than do correlations of the same magnitude between the same variables at a single point in time.

Second, numerous studies have included both respondent and partner reports of behaviors. However, nearly all of these studies have treated self-reports and partner reports of the same behaviors as separate concepts, even though they are usually highly correlated. One repeated finding is that observer ratings correlate more strongly with partner reports than with self-reports, but self- and partner reports correlate even more strongly with each other. The model in Figure 1 takes advantage of this high degree of corroboration to conceive of self- and partner-reports of the same behavior as a single second-order factor, as indicated by factor loadings λ_1 and λ_2 in Figure 1. This begins to address Lewontin’s (1995) “fundamental methodological difficulty” of relying on a single self-report while at the same time overcoming the ambiguities inherent in modeling collinear data.

Third, previous research has demonstrated strong correlations between questionnaire reports and observer ratings of behaviors in the same context, but the strength of these correlations is seldom tested against competing variables. The closest example is Sanford (2010), who examined the relationship between observer ratings and questionnaire reports after controlling for contemporaneous levels of a related concept, relationship quality. We know very little about how much variance in questionnaire reports of context-specific behaviors is explained by those context-specific behaviors, as assessed by trained coders, and how much is explained by the attributions (e.g., Bradbury & Fincham, 1992) and context-general affection or disaffection; for example, sentiment override (e.g., Hawkins, Carrère, & Gottman, 2002; Weiss, 1980), that respondents carry forward with them to the observational task. It is conceivable that questionnaire reports of behaviors in an observational task, as

assessed shortly after the observational task, overwhelm the effects of attributional processes or sentiment override, but it is also conceivable that these processes overwhelm one's ability to judge visible behaviors during recent task. In the model below, observer ratings and context-general reports of past behaviors directly compete for variance in questionnaire reports of context-specific behaviors.

Current Study

We address the skeptics' two questions by focusing on two behaviors, hostility and support, between young men and women who recently married or began cohabiting. Both hostility and support are central to family theory (Fincham & Rogge, 2010), especially in mediating between stressors such as economic hardship and family discord and outcomes such as marital quality (Conger et al., 1990; Kearney & Bradbury, 1995), psychological distress (Cutrona, 1996) and physical health (Friedman, 1991; Lovallo, 2005; Wickrama, Lorenz, Conger, & Elder, 1997). The present study also distinguishes between men and women because there are known gender differences in emotional expressiveness and intensity, especially in reaction to marital conflict and stress (Baucom, McFarland, & Christensen, 2010; Cui, Lorenz, Conger, Melby, & Bryant, 2005; Gottman & Krokoff, 1989). There are also known differences between married and cohabiting couples (Brown & Booth, 1996).

At the center of Figure 1 is observed context-specific behavior (CSB_X) – either hostility or support –during an observational task (denoted by “X”) as rated by trained observers. Path β_{41} links observer ratings of respondents' context-specific behaviors at time 2, during the observational task (CSB_X), to patterns of context-general behaviors (CGB) as self-reported by the respondent (self-report CGB_{30}) and corroborated by his or her partner (partner-reported CGB_{30}) at an earlier point in time (time 1). The subscript “30” refers to a statement in the questionnaire preamble that asks respondents to recall behaviors “during the past month” regardless of when or under what circumstances the behaviors occurred. Path β_{41} establishes the extent to which the common variance shared by the two reports has long-term predictive validity. Conversely, path β_{41} addresses the skeptic's question about whether a brief sampling of behaviors observed during a specific task reflects the patterns of behavior that are typical of couples. We express this relationship with our first hypothesis:

H1: Observer ratings of behaviors during context-specific observational tasks are significantly related to context-general patterns of interactions that reflect behaviors couples themselves acknowledge at an earlier point in time (path $\beta_{41} > 0$).

Couple reports of context-general behavior (CGB) is a second-order latent factor, where λ_1 and λ_2 are factor loadings linking CGB to the respondents' reports of their own behaviors (self-report CGB_{30}) as corroborated by their partners (partner-report CGB_{30}). The advantage of a second-order factor is that it focuses attention on respondents' and their partners' shared variance in a seamlessly integrated manner while disarming otherwise thorny multicollinearity problems that would result if the two reports were kept as separate predictors. From a methodological perspective, the importance of partitioning observed variance into common variance shared by respondents and their partners and specific error variance attributable to each separately was underscored by Cook and Goldstein (1993; see also Melby et. al., 1995) when they distinguished the common variance shared by mothers, fathers and children from the unique variance attributable to each reporter.

Conceptually, paths λ_1 and λ_2 imply that the respondents' (Self-report CGB_{30}) and partners' reports of the respondent's hostility and support during the past month are manifestations of underlying patterns of behavior that date back to their first years as a couple (CGB at time 1). Although questionnaire self- and partner-reports may be self-serving and subject to social desirability, as well as sentiment override (Weiss, 1980) and

attributional processes (Bradbury & Fincham, 1992), it remains that couples are uniquely privy to each others' behaviors in everyday life (Noller & Callan, 1988; Olsen, 1977). We do not expect couples to completely agree in their assessments of each others' hostility and support; however, the common variance shared by the two is reflected in CGB and it is this common variance against which observer ratings are regressed (i.e., $\beta_{41} > 0$).

The two constructs on the right side of Figure 1 record the respondent's hostile and supportive behaviors during the observational task as reported by respondents (self-report CSB_X in Figure 1) and as corroborated by partners (partner-report CSB_X). We hypothesize that

H2: Context-specific self-reports and partner-reports of behaviors during an observational task are significantly related to observer ratings of behaviors during the same task ($\beta_{54} > 0$ and $\beta_{64} > 0$), even after controlling for the context-general sentiments and attributions individuals express about their partners and their own past behaviors.

As a complement, we further hypothesize that

H3: Respondents' questionnaire reports of their behaviors and those of their partners during a context-specific observational task are significantly related to past context-general behaviors recorded at an earlier point in time ($\beta_{52} > 0$ and $\beta_{63} > 0$), even after controlling for observer ratings of the same behaviors during a context-specific observational task.

The questionnaire items used to measure the two self- and partner-reported context-specific behaviors (CSB_X) were administered shortly after the couples completed the observational task. Although the observer ratings are fallible reflections of the underlying hostility and support actually present during the observational task, the relative strength of the hypothesized paths β_{54} and β_{64} provides insight into the extent to which questionnaire reports about hostility and support during the observational task reflect visible hostility and support, some of which is observed, categorized and rated by observers. Meanwhile, β_{52} documents the extent to which the questionnaire reports reflect respondents' context-general self-reports about themselves and β_{63} documents partners' attributions about the respondents' behaviors. Respondents who report that their partners are consistently high on hostility in the past are likely to attribute greater negativity to partner's actions, and to carry forward greater levels of negative sentiment override during a specific task, than will those who regard their partners as being low on hostility (Bradbury & Fincham, 1992; Hawkins et al., 2002; Weiss, 1980). Conversely, respondents' who perceive their own behaviors as generally supportive may construe even their hostile interactions as constructive and encouraging (Malle, 1999; Weiss, 1980). The presence of paths β_{52} and β_{63} rather than direct paths from CGB to self-report CSB_X and from CGB to partner report CSB_X emphasizes that it is the respondents' and their partners' *unique variance*, rather than the shared variance between them (CGB), that affect their responses to questionnaire items about their behavior during the observational task.

Model estimation considerations

From the perspective of skeptics, paths β_{54} and β_{64} may not be significant simply because there may be a poor correspondence between the categories of behavior captured by observers and the aspects of hostility and support reflected in questionnaire items. However, there are at least two competing considerations when modeling multi-informant panel data. First, time 1 in Figure 1 is the first time we observe the couple after becoming a couple; time 2 is the second time we observe them as a couple two years later. As a general rule, the magnitude of correlations decays over time, so that we would expect path coefficients β_{41} ,

β_{52} and β_{63} to be larger had the lag been shorter than two years, and smaller had it been longer. Given multicollinearity between measures of concepts in multivariate models, the magnitude of these correlations also affects the magnitude of other coefficients in the model.

Second, researchers are concerned that correlations between two distinct concepts may be inflated whenever a single respondent reports on both concepts (e.g., self-report CGB₃₀ and CSB_X and partner-report CGB₃₀ and CSB_X), especially when using the same mode of data collection (questionnaire) to answer similarly worded items having similar response frameworks (Bank, Dishion, Skinner, & Patterson, 1989; Campbell & Russo, 2001; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Concern for this form of method variance is justified in the present study because respondents and partners answer questionnaire items about both their context-general (η_2 and η_3 in Figure 1) and context-specific (η_5 and η_6) behaviors, respectively. This creates the possibility that paths β_{52} and β_{63} are significant simply because of these sources of method variance, even after controlling for paths β_{54} and β_{64} . One way to reduce the problem is to have yet another source of information – preferably something closer to a true gold standard – to measure couples' time 1 context-general behaviors toward each other in general (self- and partner-reported CGB₃₀). Examples of alternative measures might include daily diaries (Bolger, Davis, & Rafaeli, 2003) or some variant on electronic surveillance (Vazire & Mehl, 2008), but they too are fallible measures with problems of their own. Other methods of measurement, ones that are maximally different from questionnaire reports, might make a stronger case (Little, Lindenberger, & Nesselroade, 1999). In the absence of alternative measures, some of the effects of method variance can be addressed by correlating the residual of questionnaire items answered by the same person and having similar wording in a structural equation model (SEM), as we will elaborate shortly.

The practical import of these two considerations is that there is a degree of indeterminacy in our model coefficients. The magnitude of the estimates can vary, depending on time lags and method variance. This is an unsettling prospect that is hardly unique to our study; indeed, it is usually not even acknowledged in cross-sectional or mono-method studies. However, awareness of indeterminacy encourages a more tentative and nuanced interpretation of results, at least compared to cross-sectional or mono-method studies that offer fewer insights into the range of possible estimates.

Method

Sample and procedures

Data to estimate the model in Figure 1 were obtained from young adult participants in the Iowa Family Transition Project, a panel study that began in 1991 when they were 9th graders and continued after they graduated from high school in 1994 (Conger & Elder, 1994; Simons, 1996). Although the sample was limited to families in rural Iowa, the measurement instruments and observational coding scheme developed for this panel have been widely used with a variety of populations, and the strength of relationships between study constructs has been replicated in other studies, including African Americans (Conger & Conger, 2002; Conger et al., 2002; Simons et al., 2002) and Mexican Americans (Parke et al., 2004) and with samples in Finland (Solantaus, Leinonen, & Punamaki, 2004) and the Czech Republic (Lorenz, Hraba, & Pechacova, 2001).

In-home interviews, which include videotaping couples interacting, have been conducted every other year since 1995. By 2007, an estimated 407 of the 550 panel members were either married or cohabiting, and 324 of the 407 were interviewed at least twice since being identified as a couple. For purposes of the present study, we focus on these 324 couples. At the time they were first interviewed together (time 1 in Figure 1), 156 (48%) of the couples

were cohabiting and 168 (52%) were married. The men averaged 24.7 years of age when they were first interviewed together, compared with 22.8 for women. For each in-home interview, couples were first videotaped interacting with each other, after which they answered questions separately about their own and their partners' behaviors during the observational task. Once these questions were completed, interviewers administered a series of questionnaires on a variety of topics, including items on their hostility and support and that of their partner "during the past month."

The observer ratings, and questionnaire reports about behaviors during the observational task, are from the second time the couples were interviewed as couples (Time 2 in Figure 1), while questionnaire reports about their behavior "during the past month" are from the first time they were interviewed together two years earlier (Time 1). Observer ratings of couples' behaviors were obtained from a general discussion task which lasted 20 – 25 minutes. This task was selected because it was successfully used with these targets' parents to elicit conversations about every-day interactions and occurrences in their life as a couple. At the time of the interview, couples were instructed to discuss their life together before a video camera but in the absence of the interviewer. Couples were given questions on cards as a means to encourage discussion. The questions opened with "how long have we known each other?" and then progressed to ask about how the couple handled household responsibilities, how they get along with each others' families, and ended with questions on what frustrates them most and what they valued most. The behaviors seen in the videotapes were categorized and rated by trained coders using the Iowa Family Interaction Rating Scales (IFIRS; Melby & Conger, 2001; Melby et al., 1998), a global rating scale designed to capture the characteristics of each partner's behavior as displayed during the interaction task.

Measurement

Observer ratings (CSB_x) of hostility and support (measured at time 2)—The latent construct of *hostility toward partner* is defined in terms of five distinct but correlated categories of behavior, one of which is labeled "hostility" and defined as the extent to which angry, critical and disapproving behavior appear during the observational task. A second category, "angry coercion," is the extent to which hostile, threatening, or blaming behavior is used to control the partner. Other categories of behavior include "antisocial behavior," plus "escalate hostility" and "reciprocate hostile," which capture the extent to which hostile exchanges build on earlier hostile interactions. Each category of behavior is scored on a 9-point scale from not at all (1) to mainly (9) characteristic.

Support toward partner is similarly composed of five observational categories, each scored on a 9-point scale. They include warmth/support, or the extent to which one person expressed interest, care, concern, support and encouragement toward the other; assertiveness, the extent to which the one clearly expresses oneself to the other in a neutral or positive way; listener responsiveness; positive communication; and prosocial behavior, as demonstrated by helpfulness and sensitivity toward each other.

Questionnaire reports of hostility and support during the observational task (time 2)—After completing the videotaped discussion task, husbands and wives were separated and asked to independently respond to a series of questionnaire items designed to measure the same concepts of hostility and support that were observed during the observational task. The questionnaire items were preceded by preambles that read, "Thinking about the discussion you just had, how much would you agree or disagree that you. . .," and "how much would you agree or disagree that your partner . . ." The items that followed the preambles were scored on a scale from strongly disagree (1) to strongly agree

(6) so that higher scores indicated stronger hostility and support. Examples of items that respondents were asked included how often their partner “was critical of you” and how often you, the respondent, “listened to what your partner had to say.” Other items used to measure hostility tapped into themes relating to anger, arguments, yelling and shouting, and lecturing. Items measuring support addressed themes of caring, affectionate behaviors, understanding and listening to each other, and laughing together.

Questionnaire reports of hostility and support during the past month (time 1)

—To estimate levels of hostility and support of *respondents toward their partners* in a variety of settings over time, the preamble “During the past month when you and your partner have spent time talking or doing things together, how often have you. . . .” was followed by items scored on a 7- point scale from never (1) to always (7). The respondents were asked whether they “got angry with your partner,” “were critical of your partner,” “yelled or shouted,” “hit, pushed, or shoved your partner,” and “argued with your partner.” Similarly, items measuring support during the past month asked about the extent to which the respondent “let him/her know you really care,” “act loving and affectionate toward him/her,” “let him/her know you really appreciate him/her . . .” and “help him/her do something that was important.” Parallel items were asked to measure partners’ hostility and support *toward the respondent*.

Analysis strategy

The concepts in Figure 1 are estimated using structural equations with latent variables (Bollen, 2002). Each of the latent variables is composed of either four or five manifest indicators, as described in the measurement section. Many of the indicators share common themes; for example, the questionnaire item for partner reports of men’s “angry” behavior has the same basic wording and response format as men’s reports of their own “angry” behavior. Further, men’s context-general “angry” questionnaire item at time 1 is repeated in a context-specific “angry” item at time 2. Although the variance shared by these items because of their common theme and similarities in question wording has been modeled in some previous studies as a method factor (e.g., an “angry” method factor), structural equation estimates of multi-trait, multi-method (MTMM) matrices often end in ill-defined solutions (e.g., Lorenz et al., 2007). This led Kenny and Kashy (1992) to recommend a less demanding alternative, which is to correlate the residuals of the manifest variables that would otherwise be brought together under a single common method factor. In the models estimated below, we adopt Kenny and Kashy’s strategy by routinely correlating items with common wording (e.g., “angry” items) within time (e.g., men and partner reports of men’s “angry” items) and between time but within reporter (e.g., men’s time 1 context-general and time 2 context-specific “angry” items). This means the models we estimate take into account both random measurement error and the systematic error associated with common themes expressed in the questionnaire wording.

Results

The model in Figure 1 was estimated for each combination of men’s and women’s hostility and support. For men’s hostility toward their partner, the overall chi-squared statistic was 543.0 with 290 degrees of freedom. The Lewis-Tucker non-normed fit index (NNFI) was 0.930 and the root mean square error of approximation (RMSEA) was 0.052 with a 90% confidence interval of 0.045 to 0.059. Factor loadings for the model are summarized in Table 1, where the abbreviation “M:M→P(30)” means men’s (M:) report of their behavior (M) toward their partner (P) during the past month (30). For men’s reports of their hostility toward their partner over the past month at time 1, factor loadings ranged from a low of 0.51 for “hit, pushed, and shoved” (hit) to a high of 0.80 for “yelled or shouted” (yell). Similar

ranges (0.54 – 0.84) are reported for partner's reports of men's behavior (P:M→P(30)). Factor loadings for observer ratings of men's hostility (X:M→P) ranged from 0.68 for "reciprocate hostile" to 0.89 for the specific category labeled "hostility."

The same statistics for women's hostility toward their partners were χ^2 (290 df) = 552, NNFI = 0.944, RMSEA = 0.053 (0.045, 0.060). The factor loadings ranged from 0.41 and 0.53 (hit) to 0.97 (observed hostility). The summary statistics for men's and women's support were [χ^2 (247 df) = 420.0; NNFI = 0.972; RMSEA = 0.047 (0.039, 0.055)] and [χ^2 (247 df) = 389; NNFI = 0.972; RMSEA = 0.042 (0.034, 0.050)], respectively.

Descriptive statistics and correlations

Table 2 provides the summary statistics and correlations for men (top of Table 2) and women (bottom). For both men and women, the first five rows and five columns of data are the correlations between latent variables for support toward partner (above the diagonal) and hostility toward partner (below the diagonal). The estimated means, standard deviations and reliabilities (Cronbach's alpha) for support are to the right of the correlation matrix while the same estimates for hostility are just below the correlation matrix.

The means and standard deviations were derived by summing responses to each item in the construct and dividing by the number of items. Mean scores for men's self report of hostility toward partners "during the past month" averaged 2.05 on the scale from 1 to 7, with a standard deviation of 0.76. The index had an estimated reliability of 0.82. Similarly, women's report of their support toward their partner during the past month averaged 6.34 at time 1, while their partners' reports of women's support averaged 5.98 at time 1 and 4.92 at time 2. Women averaged higher levels of observed hostility than men (3.14 vs. 2.62 at time 2), which is consistent with previous findings (e.g., Cui et al., 2005).

The correlations among constructs are instructive. First, the context-specific correlations between questionnaire reports and observer ratings of men's and women's hostility and support are strong, thus replicating earlier findings by Lorenz and colleagues (2007) and Sanford (2010). For example, observer ratings of men's hostility correlate 0.480 with men's reports of their own hostility during the observational task; the same correlation for women is 0.589. Second, both men's and women's reports of their own hostility during the observational task also correlate strongly with their own context-general questionnaire reports two years earlier (0.409 and 0.453, respectively), thus indicating a high degree of consistency over time. This may imply that men's and women's context-specific behavior during an observational task has a strong trait-like component so that their behavior in a context-specific situation is strongly predicted from their behaviors in general. Alternatively, the context-general and context-specific questionnaire items are similar in wording and format, so that at least some of the correlation may be attributed to the effects of method variance.

Finally, there is evidence of temporal decay. In data not in tabular form, the correlations between observer ratings (time 1) and men's and women's reports of hostility and support during the observational task at the same time (time 1) were greater (avg. = 0.406) than were the correlations of observer ratings (time 1) with husband and wife reports of hostility and support two years later (avg. = 0.356).

Evidence linking observer ratings to past context-general behavior

The first two columns of data in both Tables 3 and 4 address hypothesis H1 (path $\beta_{41} > 0$) about *whether a brief sampling of behaviors observed during a specific task (CSB_X) reflect patterns of interactions that couples themselves acknowledged two years earlier (CGB)*. Focusing first on observed hostility during the observational task (the 1st two columns of

Table 3(a)), men's observed hostility (η_4 : MHOST_X) is significantly related to our corroborated measure of men's context-general hostility (η_1 : MHost), as indicated by the standardized regression coefficient of (0.403; $t = 5.54$). The standardized factor loadings linking MMHost₃₀ and PMHost₃₀ to their latent variable (η_1 : HHost), not shown in tabular form, were 0.70 and 0.71, respectively. Table 3(b) reports that women's observed hostility (η_4 : WHOST_X) is also significantly related (0.332; $t = 4.92$) to their context-general hostility (η_1 : WHOST), again indicating that a significant portion of the variance in women's observed hostility can be linked back in time to both women's and partner's reports of women's context-general hostility at time 1. The factor loadings linking WHOST to WWHost₃₀ and HWHost₃₀ were 0.78 and 0.75, respectively. In addition, women's observed hostility was higher among both cohabiting couples (-0.129 ; $t = -2.39$) and among younger women (-0.125 ; $t = -2.12$).

The R-squared estimates for these two equations indicate that 20.1% of the variance in men's observed hostility and 18.0% of the variance in women's observed hostility were explained by the three predictor variables, mostly by the context-general measure of hostility among couples (η_1 : MHost men's hostility, and η_1 : WHost women's hostility). Although the effects sizes are modest, the results provide evidence that observer ratings are significantly linked to at least one measure of "real world" interactions as reported by participants at an earlier point in time (time 1), well before the observers rated husbands' and wives' hostility.

To explore methodological concerns about the length of the lag between observer ratings and hostility "during the past month," we re-estimated the model using only time 1 cross-sectional data. In this case, questionnaire reports on the past 30 days were actually collected in the same interview but after asking respondents about their hostility during the observational task. For this cross-sectional model, the estimated magnitude of path β_{41} , not shown in tabular form, was marginally smaller for men's hostility (0.395 instead of 0.403) and clearly stronger for women's (0.459 instead of 0.332). We will return to this theme in the Discussion.

The first two columns of Table 4(a) and 4(b) address the same question as it applies to men's (η_4 : MSpt_X) and women's (η_4 : WSpt_X) reports of their context-specific supportive behaviors. The path linking observer ratings of men's support to men's context-general support two years earlier (η_1 : MSpt) was significant but more modest than that recorded for hostility (0.286; $t = 3.31$ compared with 0.403 in Table 3(a)). For women (Table 4(b)), the path from earlier context-general support (η_1) to observed support (η_4) was stronger (0.396; $t = 5.48$) than the estimates for either men's support (Table 4(a)) or for women's hostility (Table 3(b)), although not dramatically so. Men's support appeared to be stronger among married than cohabiting couples (0.173; $t = 3.02$) and women's support was higher among those who were older at time 1 (0.125; $t = 2.13$). For men, the factor loadings linking MSpt to MMSpt₃₀ and to PMSpt₃₀ were 0.89 and 0.43, respectively, while for women the loadings linking WSpt to WWSpt₃₀ and PWSpt₃₀ were 0.67 and 0.68, respectively.

Evidence linking questionnaire reports of hostility to observer ratings

The second pair of columns in Table 3 address the 2nd and 3rd hypotheses, which jointly address the question, *Do self-reports of hostility during the observational tasks correspond with the actual hostile behaviors as rated by observers, or are they primarily a reflection of the past histories and self-appraisals men and women bring to the observational task?*

Similarly, the third pair of columns in Table 3 addresses the question, *Do partner reports of hostility during the observational task correspond with the behaviors as rated by observers, or are they primarily a reflection of the sentiment override and attributions they bring to the observational task?* The magnitude of coefficients linking men's reports of their hostility

during the observational task (η_5 : MMHost_X) to observed hostility (η_4 : MHost_X) is a relatively strong ($\beta_{54}\beta_5$ 0.378; $t = 7.07$), but MMHost_X is also significantly predicted by men's own reports of their behavior two years earlier (0.299; $t = 5.54$). Judging from the relative magnitude of these coefficients, men are able to recount their actual behaviors during the observational task and provide reports that correspond significantly with observer ratings, but they are not able to isolate their assessment of that behavior from their perceptions about their longer history of interactions with their partners.

The variances in partner reports of men's hostility (PMHost_X) are similarly partitioned. Continuing with the coefficients in the 3rd pair of columns in Table 3(a), partner reports of men's hostility are even more strongly congruent with the observer ratings (0.486; $t = 9.90$) than are men's reports, but their responses too are shaped by their sentiments and attributions about their husbands' behaviors over the past years (0.293; $t = 5.75$). For both men and women, a substantial portion of the variance in their reports of men's hostility (35.2% and 49.2%, respectively) are explained by observer ratings and earlier questionnaire reports, although some of the variance in women's reports were related to their marital status (-0.131 ; $t = -2.62$) and age at time of marriage (-0.108 ; $t = -2.29$).

The coefficients for women's hostility (Table 3(b)) are roughly the same as for men's hostility. Women's reports of their hostility during the observational task (η_5 : WWHost_X) are strongly reflective of observers reports (0.468; $t = 9.79$) but are also shaped by their history as a couple (0.337; $t = 6.80$). Partner reports of women's hostility (η_6 : MWHost_X) follow similar patterns, and the R-squares for both women's and partner reports of women's hostility during the observational task are relatively high (44.5% and 48.0%, respectively).

Again, we re-estimated the models using only time 1 cross-sectional data. The results for men's hostility (not in tabular form) show that path $\beta_{54} = 0.276$ ($t = 5.68$) rather than 0.378 and $\beta_{52} = 0.436$ instead of 0.299. Similarly, path $\beta_{64} = 0.278$ ($t = 5.79$) rather than 0.486 and $\beta_{63} = 0.384$ ($t = 8.33$) instead of 0.293. Differences in coefficient estimates were about the same magnitude for the other models. This gives us some indication of the range of values the coefficients take when different lags are assumed and data are collected in a different sequence.

Evidence linking questionnaire reports of support to observer ratings

The general pattern of coefficients reported in Table 3 are repeated for supportive behaviors in Table 4, but the coefficients are consistently weaker in magnitude. For example, men's reports of their support during the observational task (η_5 : MMSpt_X) were significantly related to both observer ratings (0.275; $t = 4.87$) and their self-report (0.225; $t = 4.14$), similar to men's self-reports of hostility in Table 3, although the proportion of variance explained is much less (19.7% instead of 35.2%). The differences are even more dramatic for the partners' reports: for example, the magnitude of the coefficients linking partner reports of women's support (η_6 : PWSpt_X in Table 4(b)) during the observational task to observer ratings (0.324; $t = 5.37$) was relatively modest when compared with the parallel coefficient (0.543; $t = 12.2$) in Table 3(b), and the proportion of variance explained was 29.4% compared to 48.0% in Table 3(b). Taken together, one conclusion may be that it is more difficult to achieve correspondence between observer ratings and questionnaire reports of supportive behaviors than hostile behaviors.

Discussion

Our purpose in this study is to address two common expressions of skepticism in modern social and behavioral research, one regarding questionnaire self-reports of behaviors and one relating to the extent to which observer ratings of behaviors can be traced back to couples'

patterns of behavior in everyday life. Our approach to these two concerns was to acknowledge that popular skepticism about research findings often arises because social and behavioral researchers, to a more obvious degree than many other disciplines, do not have unambiguous “gold standards” of measurement. In the absence of a convincing gold standard, our approach to validating a measure is to establish its consistency with other measures of the same concept. One widely accepted approach to describing consistency is to display multiple measures of the same concepts in a MTMM matrix (e.g., Campbell & Fiske, 1959), and one approach to analyzing a MTMM matrix is with confirmatory factor models (e.g., Bollen, 2002; Lorenz et al., 2007).

Our study moved beyond the traditional MTMM analysis to the structural equation model shown in Figure 1. One distinctive feature of this model is that it did not rely on the interested respondent alone to tell us about how he or she behaved; we corroborated respondent reports of behaviors with partner reports of the respondent’s behavior in the context-general situation. Clearly, neither report provides gold standards for the other. We might be more certain of our results if we had maximally different indicators of past behaviors; for example, from daily diaries or electronic surveillance, as discussed earlier, but we are more convinced about behaviors that respondents and partners agree on than either reported alone, and our use of the 2nd-order factor model focuses attention specifically on couples’ shared variance. One recommendation that derives from this experience is that the use of 2nd-order factors to isolate common variance is both a practical and theoretically justified approach to advancing knowledge on the correspondence between observer ratings and questionnaire reports of behaviors.

One conclusion we draw from this approach is that, at a minimum, observer ratings of context-specific behaviors do not seem to be *unrelated* to “natural” behaviors in a larger context-general environment. We found modest to moderate standardized regression coefficients linking observer ratings of behaviors during an observational task to the common variance shared by respondents and partners two year earlier ($\hat{\beta}_{41} = 0.286$ to 0.403). Although it is difficult to judge how large coefficients should be, the magnitude of our estimates are larger than most zero-order correlations reported in past literature. Further, we don’t expect patterns of context-general behaviors to be perfectly stable over time, especially in the early years of couples’ life together, so estimates based on a lag of two years between measurements may represent a lower bound on the strength of the relationships. We have some evidence that shorter lags would produce larger coefficients.

Our model in Figure 1 also gave us an opportunity to examine whether and to what extent context-specific questionnaire reports of behaviors can be traced back to visible behaviors during the context-specific observational task. Is it possible for respondents and their partners to see past the personal histories, attributions, and sentiments they bring with them to the observational setting and make judgments about their behaviors here and now? Again, we are limited by the lack of a gold standard and we are making the strong assumption that observers really can be trained to accurately see hostility and support. But that said, we can make at least three observations. First, respondent and partner questionnaire reports about hostility and support during the observational task were at least moderately related to observer ratings during the same task (the standardized coefficients ranged from 0.275 ($t = 4.87$) to 0.468 ($t = 9.79$)), even after controlling for their earlier reports of context-general behaviors “during the past month.” These coefficients are sufficiently large in magnitude to dispel concerns that respondents are “clueless” about their own behaviors or unable to judge the behaviors of their partners.

Second, respondents’ context-general histories and personal characteristics continue to color questionnaire responses about context-specific behaviors even though they were first

measured two years earlier and regardless of how hard researchers may try to focus respondents' attention on specific behaviors during a specific task at a specific point in time. We know of no previous research that so graphically documents the extent to which recent behaviors are shaped by the power of persistent trait-like personal histories, sentiments and attributional processes that took shape years ago.

A third lesson relates to the overall pattern of responses we found in Tables 3 and 4. There were very few systematic differences between response patterns of men and women but, as a general observation, consistency in reporting hostile behaviors was greater than in reporting supportive behaviors. The variance explained (R^2) in men's and women's self- and spouse-reports of hostility during the observational setting ranged from 35.2% to 49.2% (Table 3) compared with a range of 19.7% to 32.4% for support (Table 4). Similarly, the smallest path coefficient linking self-reports (β_{54}) and spouse reports (β_{64}) to observer ratings of hostility were larger than the largest path coefficient linking self- and spouse-reports to observer ratings of support. This is consistent with previous literature (e.g., Cui et al., 2005) and we suspect that hostile behaviors are more likely to be remembered by participants when reporting past events. Further, hostility may be easier to identify by both trained observers and participants, perhaps because supportive behaviors are more idiosyncratic and both trained raters and partners have to work harder to identify them.

A challenge for future research is to close the gap between the predictive power of questionnaire reports of hostility and support while strengthening the relationship between the categories of behaviors used by observers and the questionnaire items to which respondents and their partners react. Ideally, an iterative program of research which actively re-writes questionnaire items and re-conceptualizes behavioral categories could achieve incremental improvements in the correspondence between questionnaire items and coding schemes. The goal would be reach a point where the two approaches can substitute for one another so that neither makes unique contributions to important outcomes such as relationship quality or marital stability. In the meantime, at a more practical level, survey researchers typically do not augment large scale sample surveys with procedures to videotape families interacting, but there may be opportunities to include observational components to a subset of respondents. Recent developments in research designs with "planned missingness" may offer one scheme to encourage systematic studies linking observer ratings and questionnaire reports.

There are a number of limitations to this study. The obvious one is that our community sample is not drawn from a random sample of a known population. There are logistic difficulties in collecting observational data on a large scale, and the best evidence of generalizability may be to continue replications with a wide array of subpopulations as we discussed earlier.

Another limitation of our study and studies like ours is now more evident. Although our purpose was to quantify the magnitude of linkages between observer ratings and questionnaire reports of both context-general and context-specific behaviors, one consequence has been to heighten our awareness of the inherent indeterminacy of the modeling process that is not simply due to sampling. When examining bivariate correlations or drawing inferences from mono-method or cross-sectional data, researchers readily acknowledge that alternative ordering of concepts in path models or alternative selection of measures can lead to different conclusions. When moving to multi-informant and panel designs, conclusions can additionally be affected by decisions about the length of the lag and the choice of informants. In our model, the magnitude of coefficients linking observer ratings (β_{54} and β_{64}) to self- and spouse reports of hostility and support were greater than the stability coefficients (β_{52} and β_{63}), which leads to one interpretation over another. Those

coefficients could change in relative magnitude, however, if researchers used longer or shorter lags, or if patterns of past behaviors were based on alternative measurement methods, such as daily diaries or some variant on electric surveillance, rather than questionnaire reports. In the end, models such as the one we estimated both undermine our absolute confidence in our results, but they also offer more angles from which to draw conclusions.

Acknowledgments

The authors appreciate the reviewers many useful comments and criticisms. This research is currently supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Institute of Mental Health, and the American Recovery and Reinvestment Act (HD064687, HD051746, MH051361, and HD047573). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. Support for earlier years of the study also came from multiple sources, including the National Institute of Mental Health (MH00567, MH19734, MH43270, MH59355, MH62989, and MH48165), the National Institute on Drug Abuse (DA05347), the National Institute of Child Health and Human Development (HD027724), the Bureau of Maternal and Child Health (MCJ-109572), and the MacArthur Foundation Research Network on Successful Adolescent Development Among Youth in High-Risk Settings.

References

- Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*. 1987; 101:213–232. [PubMed: 3562706]
- Amato, P. Hofferth, SL.; Casper, LM. Handbook of measurement issues in family research. Mahwah, NJ: Lawrence Erlbaum; 2007. Studying marriage and commitment with survey data; p. 53–66.
- Aquilino WS. Two views of one relationship: Comparing parents' and young adult children's reports of the quality of intergenerational relations. *Journal of Marriage and Family*. 2005; 61:858–870.
- Bank, L.; Dishion, T.; Skinner, ML.; Patterson, GR. Method variance in structural equation modeling: Living with "glop". In: Patterson, GR., editor. *Aggression and Depression in Family Intervention*. Hillsdale, NJ: Lawrence Erlbaum Assoc; 1990. p. 247–279.
- Baucom BR, McFarland PT, Christensen A. Gender, topic, and time in observed demand-withdraw interaction in cross- and same-sex couples. *Journal of Family Psychology*. 2010; 24:233–242. [PubMed: 20545396]
- Baumeister RF, Vohs KD, Funder DC. Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*. 2007; 2:396–402.
- Bernieri FJ, Zuckerman M, Koestner R, Rosenthal R. Measuring person perception accuracy: Another look at self-other agreement. *Personality and Social Psychology Bulletin*. 1994; 20(4):367–378.
- Bolger N, Davis A, Rafaeli E. Diary methods: Capturing life as it is lived. *Annual Review of Psychology*. 2003; 54:579–616.
- Bollen KA. Latent variables in psychology and the social sciences. *Annual Review of Psychology*. 2002; 53:605–634.
- Bradbury TN, Fincham FD. Attributions and behavior in marital interaction. *Journal of Personality and Social Psychology*. 1992; 63:613–628. [PubMed: 1447688]
- Brown SL, Booth A. Cohabitation versus marriage: A comparison of relationship quality. *Journal of Marriage and Family*. 1996; 58:668–678.
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 1959; 56:81–105. [PubMed: 13634291]
- Campbell, DT.; Russo, MJ. *Social measurement*. Thousand Oaks: Sage; 2001.
- Christensen A, Heavey CL. Gender and social structure in the demand/withdraw pattern of marital conflict. *Journal of Personality and Social Psychology*. 1990; 59(1):73–81. [PubMed: 2213491]
- Coie JD, Dodge KA. Multiple sources of data on social behavior and social status in the school: A cross-age comparison. *Child Development*. 1988; 59:815–829. Retrieved from <http://www.jstor.org/stable/10.2307/1130578>. [PubMed: 3383681]

- Conger RD, Conger KJ. Resilience in Midwestern families: Selected findings from the first decade of a prospective, longitudinal study. *Journal of Marriage and the Family*. 2002; 65:361–373.
- Conger, RD.; Elder, GH, Jr. Families in troubled times: Adapting to change in rural America. Mahwah, NJ: Lawrence Erlbaum Associates; 1994.
- Conger RD, Elder GH Jr, Lorenz FO, Conger KJ, Simons RL, Whitbeck LB, Melby JN. Linking economic hardship to marital quality and instability. *Journal of Marriage and the Family*. 1990; 52:643–656.
- Conger RD, Wallace LE, Sun Y, Simons RL, McLoyd VC, Brody GH. Economic pressure in African American families: A replication and extension of the family stress model. *Developmental Psychology*. 2002; 38:179–193. [PubMed: 11881755]
- Cook WL, Goldstein MJ. Multiple perspectives on family relationships: A latent variables model. *Child Development*. 1993; 64(5):1377–1388. [PubMed: 8222878]
- Cui M, Lorenz FO, Conger RD, Melby JN, Bryant CM. Observer, self- and partner reports of hostile behaviors in romantic relationships. *Journal of Marriage and Family*. 2005; 67:1169–1181.
- Cutrona, C. Social support in couples. Thousand Oaks, CA: Sage; 1996.
- Feinberg M, Neiderhiser J, Howe G, Hetherington EM. Adolescent, parent, and observer perceptions of parenting: Genetic and environmental influences on shared and distinct perceptions. *Child Development*. 2001; 72:1266–1284. [PubMed: 11480946]
- Fincham FD, Rogge R. Understanding relationship quality: Theoretical challenges and the new tools for assessment. *Journal of Family Theory & Review*. 2010; 2:227–242.
- Floyd FJ, Markman HJ. Observational biases in spouse observation: Toward a cognitive/behavioral model of marriage. *Journal of Consulting and Clinical Psychology*. 1983; 51(3):450–457. [PubMed: 6863707]
- Friedman, HS. Understanding hostility, coping, and health. Washington, D.C.: American Psychological Association; 1991.
- Furman, W.; Jones, L.; Buhrmester, D.; Adler, T. Sibling interaction across cultures: Theoretical and methodological issues. Zukow, PG., editor. New York: Springer-Verlag; 1989. p. 163-183.
- Gottman JM, Krokoff LJ. Marital interaction and satisfaction: A longitudinal view. *Journal of Consulting and Clinical Psychology*. 1989; 57(1):47–52. [PubMed: 2487031]
- Gottman JM, Notarius CI. Decade review: Observing marital interactions. *Journal of Marriage and the Family*. 2000; 62:927–947.
- Hampson RB, Beavers WR, Hulgus YF. Insiders' and outsiders' view of family: The assessment of family competence and style. *Journal of Family Psychology*. 1989; 36:118–136.
- Hawkins MW, Carrère S, Gottman JM. Marital sentiment override: Does it influence couples' perceptions? *Journal of Marriage and Family*. 2002; 64(1):193–201.
- Janssens JMAM, DeBruyn EEJ, Manders WA, Scholte RHJ. The multitrait-multimethod approach in family assessment: Mutual parent-child relationships assessed by questionnaires and observations. *European Journal of Psychological Assessment*. 2005; 21:232–239.
- Kearney BR, Bradbury TN. The longitudinal course of marital quality and stability: A review of theory, method and research. *Psychological Bulletin*. 1995; 118:3–34. [PubMed: 7644604]
- Kenny DA, Kashy DA. Analysis of the multi-trait, multi-method matrix by confirmatory factor analysis. *Psychological Bulletin*. 1992; 112:165–172.
- Konold TR, Pianta RC. The influence of informants on ratings of children's behavioral functioning. *Journal of Psychoeducational Assessment*. 2009; 25:222–236.
- Lewontin RC. Sex lies, and social science. [Review of the books *Science in the bedroom: A history of sex research*, by V. L. Bullough, *The social organization of sexuality: Sexual practices in the United States*, by E. O. Laumann, J. H. Gagnon, R. T. Michael, & S. Michaels, and *Sex in America*, by R. T. Michael, J. H. Gagnon, E. O. Laumann, & G. Kolata]. *New York Review of Books*. 1995 Apr 20.:24–29.
- Little TD, Lindenberger U, Nesselroade JR. On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods*. 1999; 4:192–211.

- Lorenz FO, Hrabá J, Pechacova Z. Effects of spouse support and hostility on trajectories of Czech couples' marital satisfaction and instability. *Journal of Marriage and Family*. 2001; 63:1068–1082.
- Lorenz FO, Melby JN, Conger RD, Xu X. The effects of context on the correspondence between observational ratings and questionnaire reports of hostile behavior: A multitrait, multimethod approach. *Journal of Family Psychology*. 2007; 21:498–509. [PubMed: 17874935]
- Lovullo, WR. *Stress & health: Biological and psychological interactions*. 2nd ed.. Thousand Oaks: Sage; 2005.
- Malle BF. How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*. 1999; 3:23–48. [PubMed: 15647146]
- Melby, JN.; Conger, RD. Kerig, PK.; Lindahl, KM. Family observational coding systems: Resources for systematic research. Mahwah, NJ: Lawrence Erlbaum Associates; 2001. The Iowa Interactional Rating Scale: Instrument summary; p. 33-57.
- Melby, JN.; Conger, RD.; Book, R.; Rueter, M.; Lucy, LD.; Repinski, D.; Scaramella, L. The Iowa family interaction rating scales. 5th ed.. Ames, IA: Institute for Social and Behavioral Research, Iowa State University; 1998.
- Melby JN, Conger RD, Ge X, Warner TD. The use of structural equation modeling in assessing the quality of marital observations. *Journal of Family Psychology*. 1995; 9:280–293.
- Melby, JN.; Conger, KJ.; Puspitawati, H. Insider, participant observer, and outsider perspectives on adolescent sibling relationships. In: Berardo, FM.; Shehan, CL., editors. *Contemporary perspectives on family research: Vol. 1. Through the eyes of the child: Revisioning children as active agents in family life*. Stanford, CT: JAI Press; 1999. p. 329-351.(Series Ed.) & (Vol. Ed.)
- Mikelson KS. He said, she said: Comparing mother and father reports of father involvement. *Journal of Marriage and Family*. 2008; 70(3):613–624.
- Miller, RS.; Perlman, D.; Brehm, SS. *Intimate relationships*. 4th ed.. Boston: McGraw-Hill; 2007.
- Moskowitz DS. Convergence of self-reports and independent observers: Dominance and friendliness. *Journal of Personality and Social Psychology*. 1990; 58:1096–1106.
- Muthén, LK.; Muthén, BO. *Mplus user's guide: Statistical analysis with latent variables*. Los Angeles, CA: Muthén & Muthén; 2007.
- Noller P, Callan VJ. Understanding parent-adolescent interactions: Perceptions of family members and outsiders. *Developmental Psychology*. 1988; 24(5):707–714.
- Olson, DH. Insiders' and outsiders' views of relationships: Research studies. In: Levinger, G.; Raush, HL., editors. *Close relationships: Perspective on the meaning of intimacy*. Amherst, MA: University of Massachusetts Press; 1977. p. 115-135.
- Parke RD, Coltrane S, Duffy S, Buriel R, Dennis J, Powers J, Widaman KF. Economic stress, parenting and child adjustment in Mexican American and European American families. *Child Development*. 2004; 75:1632–1656. [PubMed: 15566370]
- Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*. 2003; 88:879–903. [PubMed: 14516251]
- Pelligrini AD, Bartini M. An empirical comparison of methods of sampling aggression and victimization in school settings. *Journal of Educational Psychology*. 2000; 92:350–366.
- Rhoades GK, Stocker CM. Can spouses provide knowledge of each other's communication patterns? A study of self-report, spouses' reports, and observational coding. *Family Process*. 2006; 45:499–511. [PubMed: 17220117]
- Saffrey C, Bartholomew K, Scharfe E, Henderson AJZ, Koopman R. Self- and partner-perceptions of interpersonal problems and relationship functioning. *Journal of Social and Personal Relationships*. 2003; 20:117–139.
- Sanford K. Assessing conflict communication in couples: Comparing the validity of self-report, partner-report, and observer ratings. *Journal of Family Psychology*. 2010; 24(2):165–174. [PubMed: 20438192]
- Schwarz JC, Barton-Henry ML, Pruzinsky T. Assessing child-rearing behaviors: A comparison of ratings made by mother, father, child, and sibling on the CRPBI. *Child Development*. 1985; 56:462–479. Retrieved from: <http://www.jstor.org/stable/1129734>. [PubMed: 3987419]

- Simons, RL. Understanding differences between divorced and intact families: Stress, interaction, and child outcome. Thousand Oaks: Sage; 1996.
- Simons RL, Murray V, McLoyd V, Lin K, Cutrona C, Conger RD. Discrimination, crime, ethnic identity, and parenting as correlates of depressive symptoms among African American children: A multilevel analysis. *Development and Psychopathology*. 2002; 14:271–393.
- Solantaus T, Leinonen J, Punamaki RL. Children's mental health in times of economic recession: Replication and extension of the family economic stress model in Finland. *Developmental Psychology*. 2004; 40:412–429. [PubMed: 15122967]
- Vazire S, Mehl MR. Knowing me knowing you: The accuracy and unique predictive validity of self-rating and other ratings of daily behavior. *Journal of Personality and Social Psychology*. 2008; 95:1202–1212. [PubMed: 18954202]
- Wampler, KS.; Halverson, CF, Jr. Quantitative measurement in family research. In: Boss, PB.; Doherty, WJ.; LaRossa, R.; Schumm, WR.; Steinmetz, SK., editors. *Sourcebook of family theory and methods: A conceptual approach*. New York: Springer; 1993. p. 181-194.
- Wickrama KAS, Lorenz FO, Conger RD, Elder GH Jr. Marital quality and physical illness: A latent growth curve analysis. *Journal of Marriage and the Family*. 1997; 59:143–155.
- Weiss, RL. Strategic behavioral marital therapy: toward a model for assessment and intervention. In: Vincent, JP., editor. *Advances in family intervention, assessment and theory*. Vol. Vol. 1. Greenwich, Ct.: JAI Press; 1980. p. 229-271.
- Zelditch, M, Jr. Can you really study an army in the laboratory?. In: Etzioni, A., editor. *A sociological reader on complex organizations*. New York: Holt, Rinehart & Winston; 1969. p. 528-539.

Figure 1. Theoretical model

Note. CGB = Context-general behavior; CSB = Context-specific behavior

Table 1Standardized factor loadings for men's and women's hostility and support ($N = 324$).

<u>Men's hostility toward their partner (Table 3(a)):</u>	
M: M $\rightarrow$ P(30):	angry 0.78; criticize 0.71; yell 0.80; hit 0.51; argue 0.71.
P: M $\rightarrow$ P(30):	angry 0.84; criticize 0.71; yell 0.83; hit 0.54; argue 0.71.
X: M $\rightarrow$ P :	hostile 0.89; angry/coercion 0.74; antisocial 0.74; escalate 0.83; reciprocate 0.68.
M: M $\rightarrow$ P(X):	angry 0.86; criticize 0.61; yell 0.65; lecture 0.79; argue 0.87.
P: M $\rightarrow$ P(X):	angry 0.83; criticize 0.76; yell 0.55; lecture 0.70; argue 0.81.
<u>Women's hostility toward their partner (Table 3(b)):</u>	
W: W $\rightarrow$ P(30):	angry 0.80; criticize 0.70; yell 0.87; hit 0.41; argue 0.81.
P: W $\rightarrow$ P(30):	angry 0.78; criticize 0.71; yell 0.84; hit 0.53; argue 0.84.
X: W $\rightarrow$ P :	hostile 0.97; angry/coercion 0.77; antisocial 0.89; escalate 0.85; reciprocate 0.59.
W: W $\rightarrow$ P(X):	angry 0.85; criticize 0.77; yell 0.53; lecture 0.75; argue 0.87.
P: W $\rightarrow$ P (X):	angry 0.85; criticize 0.68; yell 0.58; lecture 0.75; argue 0.87.
<u>Men's support toward their partner (Table 4(a)):</u>	
M: M $\rightarrow$ P(30):	care 0.74; affectionate 0.74; appreciates 0.94; helps 0.76.
P: M $\rightarrow$ P(30):	care 0.89; affectionate 0.83; appreciates 0.83; helps 0.74.
X: M $\rightarrow$ P:	warm/support 0.63; assert 0.85; responsive 0.89; positive comm. 0.89; prosocial 0.87.
M: M $\rightarrow$ P(X):	care 0.89; affectionate 0.80; understands 0.85; listens 0.83; laugh together 0.71.
P: M $\rightarrow$ P(X):	care 0.87; affectionate 0.71; understands 0.88; listens 0.83; laugh together 0.73.
<u>Women's support toward their partner (Table 4(b)):</u>	
W: W $\rightarrow$ P(30):	care 0.86; affectionate 0.85; appreciates 0.76; helps 0.66.
P: W $\rightarrow$ P(30):	care 0.71; affectionate 0.63; appreciates 0.89; helps 0.77.
X: W $\rightarrow$ P:	warm/support 0.58; assert 0.82; responsive 0.87; positive comm. 0.90; pro-social 0.83.
W: W $\rightarrow$ P(X):	care 0.92; affectionate 0.84; understands 0.82; listens 0.81; laugh together 0.72.
P: W $\rightarrow$ P(X):	care 0.87; affectionate 0.81; understands 0.83; listens 0.76; laugh together 0.70.

Table 2

Correlations for men's and women's hostile (below diagonal) and supportive (above diagonal) behaviors ($N = 324$).

		Correlations						
		1.	2.	3.	4.	5.	Mean	SD Alpha
Men's hostility and support toward partner:								
1. Men's self-report, past 30 days (t0)		—	0.392	0.291	0.329	0.270	6.00	0.85 0.88 (5)
2. Partner report, past 30 days (t0)		0.493	—	0.159	0.171	0.268	6.22	0.88 0.88 (5)
3. Observer rating (t2)		0.251	0.339	—	0.370	0.437	5.49	1.49 0.92 (5)
4. Men's self-report of obs task (t2)		0.409	0.302	0.480	—	0.626	4.89	0.89 0.91 (5)
5. Partner-report of obs task (t2)		0.315	0.478	0.613	0.570	—	5.13	0.79 0.90 (5)
Mean		2.05	2.00	2.62	2.02	1.75		
St. Deviation		0.76	0.74	1.42	0.97	0.82		
Alpha		0.82	0.83	0.89	0.86	0.84		
Women's hostility and support toward partner:								
1. Women's self-report, past 30 days (t0)		—	0.431	0.188	0.331	0.165	6.34	0.75 0.87 (4)
2. Partner report, past 30 days (t0)		0.589	—	0.337	0.282	0.410	5.98	0.86 0.86 (4)
3. Observer rating (t2)		0.307	0.286	—	0.511	0.459	5.72	1.33 0.90 (5)
4. Women's self-report of obs task (t2)		0.453	0.293	0.589	—	0.636	5.11	0.80 0.90 (5)
5. Partner-report of obs task (t2)		0.263	0.416	0.638	0.621	—	4.92	0.80 0.90 (5)
Mean		2.06	2.38	3.02	1.76	1.95		
St. Deviation		0.78	0.84	1.60	0.84	0.87		
Alpha		0.84	0.84	0.91	0.86	0.84		

Note. All correlations are significant at the $p < 0.01$ level.

Table 3

Standardized regression coefficients for men's and women's hostility ($N = 324$).

(a) Equations for men's hostility						
	η_4 : MHost _X		η_5 : MMHost _X		η_6 : PMHost _X	
	beta	t-ratio	beta	t-ratio	beta	t-ratio
Couples' consensus (η_1 : MHost)	0.403	5.54				
Observer ratings (η_4 : MHost _X):			0.378	7.07	0.486	9.90
Men's self-report (η_2 : MMHost)			0.299	5.54		
Partner's report (η_3 : PMHost)					0.293	5.75
Age at marriage/cohabitation	-.056	-0.87	-.095	-1.71	-.131	-2.62
Married (1) vs. cohabiting (0)	-.118	-1.68	-.085	-1.68	-.108	-2.29
R ²	20.1%		35.2%		49.2%	
(b) Equations for women's hostility						
	η_4 : WHost _X		η_5 : WWHost _X		η_6 : PWHost _X	
	beta	t-ratio	beta	t-ratio	beta	t-ratio
Couples' consensus (η_1 : WHost)	0.332	4.92				
Observer ratings (η_4 : WHost _X):			0.468	9.79	0.543	12.2
Women's self-report (η_2 : WWHost)			0.337	6.80		
Partner's report (η_3 : PWHost)					0.268	5.45
Age at marriage/cohabitation	-.125	-2.12	-.042	-0.81	-0.042	-0.82
Married (1) vs. cohabiting (0)	-.129	-2.39	-.002	-0.39	-0.074	-1.52
R ²	18.0%		44.5%		48.0%	

Note. In Tables 3 and 4 all t-ratios larger than |2.00| are significant at the $p < 0.05$ level.

Table 4

Standardized regression coefficients for men's and women's support ($N = 324$).

(a) Equations for men's support									
	η_4 : MSpt _x		η_5 : MMSpt _x		η_6 : PMSpt _x				
	beta	t-ratio	beta	t-ratio	beta	t-ratio	beta	t-ratio	
Couples' consensus (η_1 : MSpt)	0.286	3.31							
Observer ratings (η_4 : MMSpt _x):			0.275	4.87	0.372	7.06			
Men's self-report (η_2 : HHSpt)			0.225	4.14					
Partner's report (η_3 : WHSpt)					0.200	3.97			
Age at marriage/cohabitation	0.091	1.58	0.068	1.13	0.034	0.61			
Married (1) vs. cohabiting (0)	0.173	3.02	0.113	2.08	0.100	1.86			
R ²	13.0%		19.7%		22.8%				
(b) Equations for women's support									
	η_4 : WSpt _x		η_5 : WWSpt _x		η_6 : PWSpt _x				
	beta	t-ratio	beta	t-ratio	beta	t-ratio	beta	t-ratio	
Couples' consensus (η_1 : WSpt)	0.396	5.48							
Observer ratings (η_4 : WSpt _x):			0.435	8.83	0.324	5.37			
Women's self-report (η_2 : WWSpt)			0.254	5.15					
Partner's report (η_3 : HWSpt)					0.303	5.85			
Age at marriage/cohabitation	0.125	2.13	−0.002	−0.39	0.104	1.92			
Married (1) vs. cohabiting (0)	0.077	1.36	0.100	2.00	0.083	1.60			
R ²	19.2%		32.4%		29.4%				