Psychological Assessment
2005, Vol. 17, No. 4, 423–432

1040-3590/05/$12.00 DOI: 10.1037/1040-3590.17.4.423

Spousal Agreement Regarding Relationship Aggression on the Conflict Tactics Scale–2

Lorelei E. Simpson and Andrew Christensen
University of California, Los Angeles

The current study assessed agreement within 273 treatment-seeking couples that reported on aggression in their relationship using the Conflict Tactics Scale–2 (CTS-2; M. A. Straus, S. L. Hamby, S. Boney-McCoy, & D. B. Sugarman, 1996), the dominant instrument for assessing violence among couples. Results revealed low-to-moderate levels of agreement, consistent with previous studies using earlier versions of the CTS. Both husbands and wives reported a lower level of aggression for themselves than their partners attributed to them, though this discrepancy was generally stronger for husbands. In addition, both husbands and wives showed higher agreement on items rated as more objective and specific. The implications of these findings for researchers and clinicians are discussed.

Keywords: domestic violence, marriage, couple therapy, agreement, CTS-2

Collecting data from both partners in a relationship has been both a triumph and a trial for couple researchers. The benefits of this methodology are numerous, including more in-depth information about couple and individual functioning, examination of the differing views of each partner, and an improved understanding of couple relationships. However, as with every research methodology, this procedure has limitations as well as advantages. One potential concern is the disagreement between partners regarding events in their relationships. On the one hand, obtaining differing perspectives on a relationship can be useful, and statistical methods such as multilevel modeling often rely on reports from multiple sources. However, numerous researchers have also found that couples not only disagree in their opinions about their relationships, but also in reporting on events, especially those that are negative, such as relationship violence.

Studies of several different populations of couples have consistently found that agreement about the occurrence and frequency of aggression ranges from low to moderate levels, though it is usually significantly greater than chance. These varied samples include couples in the community (e.g., Arias & Beach, 1987), couples in which the husband is participating in a batterer treatment program (e.g., Browning & Dutton, 2003), and couples seeking marital therapy (e.g., Heyman & Schlee, 1997). They include dating couples (e.g., Moffitt et al., 1997) and newlyweds (e.g., Heyman & Schlee, 1997), couples drawn from large-scale survey studies (e.g., Schafer, Caetano, & Clark, 1998) and couples from more specific populations, such as military bases (e.g., Bohannon, Dosser, & Lindley, 1995). Despite the consistent findings regarding spousal agreement on relationship violence, there has been little examination of correlates of low agreement or exploration of ways in which to improve our estimates of the prevalence and frequency of violence. Furthermore, almost all of the above studies used the same measure, the Conflict Tactics Scale (CTS; Straus, 1979).

The CTS was originally designed to assess the prevalence and frequency of acts of physical and verbal aggression. Internal consistency, test–retest reliability, and validity have been well established for the CTS (Barling, O'Leary, Jouriles, Vivian, & MacEwen, 1987; Cascardi, Avery-Leaf, O'Leary, & Slep, 1999; Pan, Neidig, & O'Leary, 1994; Schafer, 1996), yet it has limitations. These include a relatively brief assessment of psychological aggression, no assessment of injury or other types of aggression, such as sexual coercion, and a failure to assess for the context and purpose of violent acts (see Archer, 1999, and Dobash, Dobash, Wilson, & Daly, 1992, for reviews). For these reasons, Straus et al. (1996) developed a newer, more comprehensive version, the Revised Conflict Tactics Scales (CTS-2). The CTS-2 was intended to improve the original CTS in several ways: The physical aggression scale was expanded to include additional items. The verbal aggression subscale was renamed the psychological aggression subscale and additional items regarding nonverbal psychological aggression were included. Finally, additional subscales assessing injury and sexual coercion were created (Straus et al., 1996). Straus and his colleagues established good internal consistency, construct validity, and discriminant validity among a sample of college students, and other researchers have used factor analysis to examine the validity of the CTS-2 among other populations (Jones, Ji, Beck, & Beck, 2002; Lucente, Fals-Stewart, Richards, & Goscha, 2001; Newton, Connelly, & Landsverk, 2001). However, while some investigation of this new measure has been undertaken, comparable analyses of spousal agreement using the CTS-2 have

Lorelei E. Simpson and Andrew Christensen, Department of Psychology, University of California, Los Angeles.

This research was supported by the National Institute of Mental Health through National Research Service Award MH63616 to Lorelei E. Simpson at UCLA and by research grants to Andrew Christensen at UCLA (MH56223) and Neil S. Jacobson at the University of Washington (MH56165) for a two-site clinical trial of couple therapy. After Jacobson's death, William George served as principal investigator at the University of Washington.

Correspondence concerning this article should be addressed to Lorelei E. Simpson, who is now at Psychology Service (116B), Department of Veterans Affairs Medical Center, 3350 La Jolla Village Drive, San Diego, CA 92161. E-mail: lorelei.simpson@med.va.gov

yet to be fully conducted. The current study offers an examination of spousal agreement using this new measure.

This issue of partner agreement is particularly important because the CTS and CTS-2 are often used both in survey studies and studies of specific populations to estimate the prevalence and severity of violence within a group. There are several possible explanations for low interpartner agreement. One or both partners may not report truthfully due to factors such as social desirability, fears of legal or partner repercussions, or a desire to portray their partner in a specific light. Items on measures such as the CTS-2 may not be specific or objective enough to measure "true" rates of aggression (e.g., "beat up" may mean different things to different people). It may be unrealistic to expect respondents to be able to recall the specific frequency of a number of behaviors over a period of one year. If researchers wish to obtain the best estimates of aggression in populations, they need to consider both the agreement between partners on measures of aggression and the possible factors impacting that agreement. This information may also influence decisions about how to calculate estimates of prevalence, such as averaging spousal reports or choosing the highest report of aggression.

In addition to replicating earlier studies of agreement on the CTS with the more comprehensive CTS-2, we also attempt to expand the current knowledge on agreement to address some of the above issues. First, we examine the direction of disagreement, asking such questions as whether partners tend to report their own violence at greater or lesser prevalence than their partner attributes to them, and whether there are any gender differences in reporting patterns. These questions are raised by the issues discussed above and by the research on bias in reporting in other areas. Studies have shown that people vary in their tendency to attribute responsibility to themselves or others, depending on the situation. In general, we are more likely to attribute responsibility for positive events to ourselves and for negative events to others (Baumeister, 1982; Burger & Rodman, 1983; Thompson & Kelley, 1981). Although the CTS-2 does not include items about responsibility for aggression, the structure asks respondents to rate their own and then their partner's aggression. This format may result in an implied responsibility, which combined with the social desirability issues associated with reporting relationship violence, may lead spouses to rank their partner as more violent.

Second, we explore correlates of agreement in two areas: item characteristics and couple characteristics. In terms of item characteristics, we examine the relationship between agreement and the specificity, objectivity, and severity of items. Previous research has revealed that spouses tend to agree more on items regarding nonviolent behaviors that are specific and objective than on those that are nonspecific or subjective (Christensen, Sullaway, & King, 1983; Jacobson & Moore, 1981). There is little research on the impact of the severity of violence on agreement, but we hypothesized that couples would have higher agreement on more severe acts of violence because such events will be more salient and easier to recall.

At the couple level, we examine the association between marital satisfaction and partner agreement. There is evidence that people in happy relationships may actually attribute more responsibility for negative events to themselves and responsibility for positive events to their partners, while the opposite may be true for people in unhappy relationships (Fincham, Bradbury, Arias, Byrne, & Karney, 1997). It may be that partners in happy relationships have an incentive to try to make their partner look good, or to believe good things about their partner, while people in unhappy relationships have a greater desire to blame problems or negative events on their partner.

Finally, we examine these issues among a clinically relevant population: couples seeking therapy. This population is especially important because the prevalence of aggression in it is very high, ranging up to 60–70% of couples (Ehrensaft & Vivian, 1996; Meyer, Vivian, & O'Leary, 1998). Given that violence is quite common among couples seeking therapy, and that researchers often rely upon this population in studies of violence (e.g., O'Farrell, Van Hutton, & Murphy, 1999; O'Leary, Vivian, & Malone, 1992), information about how spousal agreement affects estimates of violence within this population is needed. Furthermore, studies of violence among couples seeking therapy are clinically relevant as well, as researchers disagree about the appropriateness of couple therapy when violence is present (e.g., Bograd & Mederos, 1999; Gauthier & Levendosky, 1996; O'Leary, Heyman, & Neidig, 1999). Further knowledge about interpartner agreement within this population may be useful to clinical researchers using self-report measures to gather information about violence and to clinicians using such self-report measures to determine if and when couple therapy is appropriate for a given couple.

The purpose of the current study is to add to the knowledge about couple agreement regarding violence in several ways. First, we examine the degree of interpartner agreement on items and subscales of the CTS-2, hypothesizing that couples will evidence statistically significant but relatively low levels of agreement, as has been found with the shorter, earlier versions of the CTS. Second, we examine direction of disagreement in reporting aggression, hypothesizing that disagreement occurs because both husbands and wives report that their partner has committed more acts of aggression than the partner endorses. Specifically, disagreements will occur because the recipient of the behavior says that it occurred while the partner who committed the behavior says that it did not, rather than the reverse. Third, we hypothesize that couples will be more likely to agree on CTS-2 items that are specific, objective, and severe, than on items that are vague, subjective, or mild. Finally, we explore the association between marital satisfaction and agreement. We hypothesize that couples with lower marital satisfaction will be less likely to agree on the history of violence in their relationship than couples with higher marital satisfaction. Our examination was conducted on a sample of treatment-seeking couples, potentially one of the most relevant populations for clinicians and many researchers.

Method

Sample

The sample consisted of 273 married, heterosexual couples that had expressed interest in participating in a study of marital therapy. The average age was 39.7 years ($SD = 9.0$) for wives and 41.6 years ($SD = 9.2$) for husbands. Sixty-eight percent of wives were Caucasian, 5.1% were Asian American, 8.7% were African American, 8.0% were Latina, 0.4% were Native American, 3.6% were of some other ethnicity, and 5.8% declined to report their ethnicity. Seventy-two percent of husbands were Caucasian, 3.6% were Asian American, 9.5% were African American,

6.9% were Latino, 0.4% were Native American, 2.2% were of some other ethnicity, and 5.5% declined to report their ethnicity. The average length of marriage was 9.4 years ($SD = 7.8$), couples had an average of 1.4 ($SD = 1.2$) children, and both husbands and wives had an average of 16 years of education ($SD = 3.0$). Sixty-five percent of wives and 85% of husbands were employed at the time of the study.

Measures

The CTS-2 (Straus et al., 1996) was used to assess aggression in the couples. The CTS-2 consists of 78 items comprising 15 subscales measuring minor, severe, and overall rates of physical (e.g., slapping, pushing/shoving, kicking, or beating), psychological (e.g., yelling/screaming, insults or name calling, threatening), and sexual aggression (e.g., coercing or forcing a partner to have sex); rates of injury (e.g., bruises, scrapes, broken bones, or needing a doctor as a result of a fight); and rates of nonviolent negotiation behaviors (e.g., expressing opinions, supporting partner's view). All but the negotiation items and subscales were used in the current study, though the participants completed the entire scale. Husbands and wives reported on both their own and their partner's behavior. Participants rated whether a behavior had never occurred, occurred once, twice, 3–5, 6–10, 11–20, or more than 20 times in the past year, or had occurred, but not in the past year. The CTS-2 was scored according to guidelines outlined by Straus et al. (1996) which provide subscales measuring frequency of aggression in the past year and lifetime prevalence of aggression. Frequency is calculated by converting the above categories into single digits (i.e., never = 0, once = 1, twice = 2, 3–5 times = 4, 6–10 times = 8, 11–20 times = 15, more than 20 times = 25), and summing the items within each subscale. Lifetime prevalence is calculated by converting the never category to 0 and all other categories to 1, and then summing the items within subscales to indicate if a type of violence had ever occurred in the relationship. In the current study we examined only frequency and prevalence of aggression in the past year, as recent data on aggression is of greater interest to clinicians and researchers and more likely to be preserved in the memory of the participants.

Each of the CTS-2 violence and injury items was rated by eight coders on two Likert scales ranging from 1 to 5 on objectivity and specificity. Objectivity was defined as the degree to which an objective outside observer, rather than an involved participant, could judge an act as having occurred or not. For example, "did something to spite partner" was rated as one of the least objective items, as it can be a matter of subjective opinion as to whether an individual has deliberately done something to spite another. In contrast, an item such as "had a broken bone as a result of fight with partner" is clearly an occurrence that could be judged by both participants or objective observers equally well. Similarly, specificity was defined as the extent to which an item defined a clear, discrete act of aggression. Items such as "shouted or yelled at partner" were defined as the most vague, as it is not as clear from the item which behaviors might fall into that category, while items such as "choked partner" refer to a clearly definable, discrete act. Although the two scales are conceptually distinct, though related, in practice they were highly correlated ($r = .88$, $p < .001$), so we combined them to reduce the number of analyses and avoid Type I error inflation. The coders included the authors, five graduate students, and one lab assistant with a BA in psychology. There were four male and four female coders. The interrater reliability was excellent, with a coefficient alpha of .92 for the combined scale across all eight raters demonstrating good consistency between raters. Severity of violence was measured using the minor/severe categories of items provided in the scoring system for the CTS-2.

The Marital Satisfaction Inventory, Revised (MSI-R; Snyder, 1997) was administered to measure marital functioning. The MSI-R is a well-validated and reliable measure that has been used frequently to assess marital satisfaction and functioning in a number of areas. It consists of 150 true/false items that comprise 12 subscales. Of these 12, only the Global Distress Subscale (GDS), which assesses overall marital satisfaction, was used for the current analyses, although couples completed the full MSI-R. The GDS has been well-validated as a measure of global relationship satisfaction.

Procedure

Participants in this study were recruited through the Couple Therapy Project, a dual-site study at the University of California, Los Angeles, and the University of Washington, Seattle, comparing two types of couple therapy (cf. Christensen et al., 2004). Couples were recruited through newspaper, radio, and TV advertisements, clinical referrals, and flyers. All couples in this study completed two stages of eligibility screening. In the first, both spouses completed a phone interview in which they reported on demographic information, amount of desire for seeking couple therapy, and marital satisfaction. Eligibility requirements at the first screening included the following: spouses were required to be married and living together, to be between the ages of 18 and 65, have at least a high school diploma or GED, be fluent in English, have not recently started or changed a psychotropic medication, not currently be in another form of psychotherapy, and score in the clinical range for marital distress (an average score of less than 100 on the Marital Adjustment Test; Locke & Wallace, 1959).

At the second stage of eligibility screening, both partners were mailed the MSI-R and the CTS-2, among other measures, and were asked to complete them separately and return them to the project. For the purposes of the larger therapy study, couples who were not within the clinical range of marital distress and couples in which the wife reported moderate-to-severe husband domestic violence were told they were ineligible and were provided appropriate referrals. Those couples that passed the exclusionary criteria at the second stage of eligibility were asked to come to the university for a final stage of eligibility assessment. For the purposes of the current study, all couples in which both partners completed and returned the MSI-R and the CTS-2 at the second stage of eligibility were included in the sample. The current sample consisted of all couples who were included in the therapy study and all couples who were excluded from the therapy study as a result of moderate-to-severe violence or because they were not within the clinical range of marital distress.

Results

Prior to examining agreement, we calculated the prevalence and frequency of aggression within the sample. Prevalence estimates were calculated by combining husband and wife reports of aggression in the past year and are shown in Table 1. As can be seen in

Table 1
Combined Spouse Reports of Prevalence of Aggression in the Past Year

Type of aggression	*N* (%)
Minor psychological	270 (98.9%)
Severe psychological	192 (70.3%)
Minor physical	148 (54.2%)
Severe physical	73 (26.7%)
Minor sexual coercion	134 (49.1%)
Severe sexual coercion	29 (10.6%)
Minor injury	67 (24.5%)
Severe injury	15 (5.5%)
Any aggression (physical aggression, sexual coercion, injury)	203 (74.4%)

Note. If either wife or husband reported aggression for self or other, it is listed in the table.

this table, the majority of couples had experienced at least one act of aggression in the past year, but relatively few couples had experienced severe types of aggression. In terms of frequency, both husbands and wives committed an average of 30 acts of psychological aggression and an average of 4 acts of physical aggression in the past year. Further discussion of the rates and types of violence in this sample can be found elsewhere (Simpson, Doss, Wheeler, & Christensen, 2005), but it should be noted here that the overall severity of violence in this sample was relatively low. Although some cases of more severe aggression (i.e., battering) were present, the majority of violence in this sample appeared to be milder and to occur on a mutual basis, consistent with the description of common couple violence provided by Johnson (1995; Johnson & Ferraro, 2000). In fact, the overall frequency rates are considerably lower than those found in studies of abuse, which range as high 65 acts of violence in a year (Johnson, 1995). As such, the term "aggression" is probably more descriptive for the acts present in this sample than "domestic violence," as defined by the APA Task Force on Violence and the Family (American Psychological Association, 1996) and Walker (1999), who both define "domestic violence" as a pattern of abusive behaviors intended to dominate and control a partner.

Item and Subscale Agreement

We used multiple methods to examine spousal agreement for husband and wife aggression. To examine agreement on the occurrence of items we used a measure of percent agreement and two statistical measures of agreement that take into account chance agreement, Cohen's kappa (Cohen, 1960) and Yule's *Y* (Spitznagel & Helzer, 1985). Percent occurrence agreement was calculated by dividing the number of couples in which both partners had endorsed a behavior by the number of couples in which either partner had endorsed the behavior. This measure of agreement was thought to be more informative than total percent agreement (i.e., number of couples that agreed that a behavior occurred plus number of couples that agreed that the behavior did not occur divided by total couples), as the emphasis is on agreement among those couples who actually experienced aggression and is a more stringent measure of agreement than total percent agreement, especially when there are small cell sizes, as in many cases here. Although any measure of percent agreement is biased by the proportion of couples/individuals reporting a behavior, occurrence agreement is more easily interpreted than other measures of agreement (see below) and is often reported in other studies of agreement.

In contrast, Cohen's kappa (Cohen, 1960) and Yule's *Y* (Spitznagel & Helzer, 1985) both provide a measure of agreement that accounts for chance and is less affected by the proportion of individuals endorsing a behavior, but they are constrained by the number of responses per cell. For items and subscales on which husbands or wives reported fewer than 10 occurrences (e.g., "burned or scalded on purpose") or nonoccurrences (e.g., "yelled at partner") neither kappa nor *Y* was calculated, as the statistics becomes meaningless as any cell size approaches zero (Christensen et al., 1983). As a result, neither kappa nor Yule's *Y* was calculated for approximately half of the items. Kappa was used because it is one of the more common measures of interrater agreement and is often used in studies of agreement (e.g., Jouriles & O'Leary, 1985; Moffitt et al., 1997; Schafer et al., 1998; Schafer, Caetano, & Clark, 2002). It does, however, result in problematic estimates of agreement when prevalence of a behavior is very low or very high, which can be corrected using Yule's *Y*. *Y* is on the same scale as kappa, ranging from −1 to 1, with a score of 0 indicating agreement at the level of chance, negative scores indicating agreement less than chance and positive scores indicating agreement greater than chance. Finally, we calculated agreement on the frequency of events using tau-b correlations, as the data were not normally distributed. We correlated wife report of husband behavior to husband self-report and husband report of wife behavior to wife self-report.

Occurrence and frequency agreement for the subscales of the CTS-2 are reported in Tables 2 and 3. For occurrence agreement,

Table 2
Occurrence Agreement for the Past Year

	% occurrence agreement		Kappa		*Y*	
Subscale	Wife	Husband	Wife	Husband	Wife	Husband
Minor psychological	91.48%	94.03%	—	—	—	—
Severe psychological	46.84%	47.59%	0.37	0.41	0.39	0.43
Any psychological	91.88%	94.05%	—	—	—	—
Minor physical	62.30%	58.85%	0.64	0.59	0.65	0.61
Severe physical	38.10%	37.78%	0.47	—	0.58	—
Any physical	61.60%	57.38%	0.62	0.58	0.63	0.60
Minor sexual coercion	26.25%	44.63%	0.29	0.44	0.38	0.46
Severe sexual coercion	0.00%	5.00%	—	—	—	—
Any sexual coercion	26.51%	43.20%	0.29	0.41	0.38	0.43
Minor injury	54.72%	37.50%	0.66	—	0.75	—
Severe injury	18.18%	0.00%	—	—	—	—
Any injury	54.24%	35.59%	0.64	0.45	0.73	0.59
Any aggression	59.76%	60.67%	0.52	0.49	0.53	0.49

Note. Dashes indicate that agreement was not calculated because there were cells with less than 10 cases. Agreement data for individual items are available from Lorelei E. Simpson on request.

Table 3
Agreement on Frequency of Aggression in the Past Year

Item/subscale	Frequency (Tau-B correlation)	
	Wife	Husband
Minor psychological aggression	.40	.40
Severe psychological aggression	.37	.45
Any psychological aggression	.41	.42
Minor physical aggression	.62	.57
Severe physical aggression	.47	.53
Any physical aggression	.61	.57
Minor sexual coercion	.30	.40
Severe sexual coercion	−.01	.09
Any sexual coercion	.29	.38
Minor injury	.64	.47
Severe injury	.31	−.01
Any injury	.62	.46

Note. Agreement data for individual items are available from Lorelei E. Simpson on request.

couples were considered to agree on a given subscale if both of them reported that at least one behavior on the subscale had occurred or if both reported that no behavior on the subscale occurred. As a result, a couple in which one partner reported six minor physically aggressive items in the past year and the other reported one minor physically aggressive item would be rated as agreeing minor physical aggression had occurred. Frequency agreement was measured to determine if couples not only agreed on whether a type of aggression had ever occurred, but also on how frequently it occurred.

As expected, percent occurrence agreement was quite variable, and highly associated with prevalence, such that the most prevalent behaviors (e.g., psychological aggression) had the highest agreement while the least prevalent (e.g., severe sexual coercion or severe injury) had the lowest agreement. However, the pattern of percent occurrence agreement was fairly consistent with kappa and Yule's *Y*, suggesting that although percent occurrence agreement was affected by prevalence, it was still similar to the other measures of agreement used in this study. Subscale agreement as measured by both kappa and Yule's *Y* was significantly above chance, though within the low-to-moderate range, as defined by Spitzer and Fleiss (1974), who state that kappa values less than .50 indicate poor reliability, values between .50 and .75 indicate fair reliability, and values greater than .75 indicate good reliability. Finally, frequency agreement, measured by tau-b correlations, was also significant, with the exception of severe sexual coercion and husband experience of severe injury, though the correlations were again in the moderate range.

We then assessed agreement for specific items. Percent occurrence agreement at the item level for wife behavior ranged from 0% to 84.44% with a mean of 28.92% (SD = 22.50), while agreement for husband behavior ranged from 0% to 83.20% with a mean of 25.81% (SD = 23.87). Item-level kappa for wife behavior ranged from 0.20 to 0.54 with a mean of 0.38 and standard deviation of 0.10. The item-level kappa for husband behavior ranged from 0.14 to 0.59 with a mean of 0.40 and a standard deviation of 0.11. The item-level Yule's *Y* for wife behavior ranged from 0.20 to 0.66, with a mean of 0.49 and a standard deviation of 0.13. Yule's *Y* for husband behavior ranged from 0.14 to 0.66 with a mean of 0.51 and a standard deviation of 0.13. Finally, the item-level tau-b correlation for wife behavior, assessing agreement on how frequently the wife committed the behavior in question, ranged from 0.00 to 0.66 with a mean of 0.33 and standard deviation of 0.17, while the tau-b correlation for husband behavior ranged from −0.01 to 0.57 with a mean of 0.31 and a standard deviation of 0.18. The item level results are consistent with the subscale level results in that agreement is relatively poor, though there is some variation, with a few items having agreement in the moderate range. These total agreement results are consistent with our hypothesis that agreement would be significantly greater than chance, but at low-to-moderate levels.

Categorical Agreement

Next, because the CTS and CTS-2 are often used to categorize couples as violent or nonviolent, we examined concordance on these categories. Couples were assigned to the "violent" category if the respondent reported any physical aggression, sexual coercion, or injury and to the "nonviolent" category if none of the above were reported. Wives reported that 168 (61.5%), and husbands reported that 174 (63.7%), of couples had experienced some form of violence in the past year. However, although 70 couples agreed that no violence had occurred and 139 agreed that violence had occurred, 64 couples disagreed as to whether or not violence had occurred in the past year. This resulted in a percent occurrence agreement of 68.5%, a kappa of .50 and a Yule's *Y* of .75, consistent with the above findings on item and subscale agreement and our hypothesis of significant, but only moderate, agreement.

Furthermore, as couple therapy researchers and clinicians may base treatment decisions (e.g., whether or not the couple is appropriate for conjoint treatment or should be referred to gender-specific violence programs) on presence or absence of abuse, we examined spousal agreement in this area as well. Although there are no officially agreed upon criteria for defining abuse versus less severe aggression as of yet, the following standard has been used both to exclude couples who are too violent from therapy studies (Christensen et al., 2004) and to include couples in studies of battering (Jacobson & Gottman, 1998). According to this standard, a partner is classified as abusive if he or she has ever committed an act of severe violence during the relationship (e.g., used force to make partner have sex, beat up partner), has committed more than two acts of moderate violence in the past year (e.g., punched or hit partner, slapped partner), or has committed more than six acts of mild violence in the past year (e.g., twisted partner's arm or hair, pushed or shoved partner). Clearly, whether or not these criteria accurately distinguish between abusive and nonabusive partners is arguable, but as they have been used for this purpose, they are quite relevant to the current study. Based on these criteria, 56 couples agreed that the wife had been abusive, 161 agreed that the wife had not been abusive, and 56 couples disagreed about whether or not the wife had been abusive. Similarly, 59 couples agreed that the husband had been abusive, while 138 agreed that the husband had not, while 76 couples disagreed as to whether or not the husband had been abusive. These reports resulted in a percent occurrence agreement of 50.0%, a kappa of .52, and a Yule's *Y* of .58 for wife violence and a percent occurrence agreement of

43.7%, a kappa of .40, and a Yule's *Y* of .44 for husband violence, again falling within the low-to-moderate range of agreement.

Direction of Spousal Disagreement

In order to assess the direction of spousal disagreement, we created a between spouse difference index that was calculated separately for husband and wife behavior. To calculate this index, we subtracted the partner's report of violence from the individual's report of his or her own violence for both husbands and wives (e.g., wife report of own violence minus husband report of wife violence). A positive score indicated that the subject reported that he or she had committed more violence than his or her partner said they did, a score of zero indicates that partners reported the same amount of violence for the individual, and a negative score indicates that the partner reported that the subject committed more violence than the subject self-reported. We then summed the between spouse difference index for each item of each subscale to create difference indices for the subscales. We used single sample *t* tests to determine if there was a significant difference between husband and wife reports of aggression by comparing the mean of the between spouse difference index to 0, the case when the husband and wife reports for a behavior were identical. As hypothesized, husbands reported more overall wife violence than wives reported about themselves and wives reported more overall husband violence than husbands reported about themselves, a statistically significant, though fairly small discrepancy, with small effect sizes ($d = .25$ for wife behavior, $d = .33$ for husband behavior; see Table 4). We then used paired sample *t* tests to determine if spouses disagreed more or less in reporting husband or wife behavior by comparing the between spouse difference index for husbands to that for wives. The analyses revealed no significant differences between spouses for any of the types of violence except for minor sexual coercion, in which husbands reported more wife aggression than wives reported about themselves, but wives reported less husband aggression than husbands reported about themselves, probably a result of husbands reporting more sexual coercion overall than wives (see Table 4). These results provide support for the hypothesis that both spouses would report that their partner had committed more violence than their partner reported about themselves.

Correlates of Agreement

Next, we tested the hypothesis that spouses would agree more on items that were objective, specific, or severe than items that were subjective, vague, or mild. In order to test the hypothesis that couples would be more likely to agree on items that were specific and objective, the combined rating scale was correlated with item percent occurrence agreement, kappa, *Y*, and item tau-b correlations for husband and wife behavior. Partial correlations controlling for item prevalence were used. Although all correlations were in the expected direction, only husband *Y* ($r = .60$, $p < .05$) and husband tau-b correlation ($r = .46$, $p < .05$) were significant, though there was a trend to significance for husband percent occurrence agreement ($r = .32$) and husband kappa ($r = .52$).

Table 4
Between Spouse Difference Index

Subscale	Single sample *t* tests: mean (SD), *t*(272), effect size		Paired sample *t* tests: *t*(272), effect size
	Wife behavior	Husband behavior	
Minor psychological	0.07 (1.28) $t = 0.87$, $d = .08$	−0.07 (1.21) $t = -0.90$, $d = .08$	−0.92, $d = .11$
Severe psychological	−0.20 (1.03) $t = -3.12$**, $d = .27$	−0.13 (0.78) $t = -2.74$**, $d = .24$	0.71, $d = .08$
Psychological	−0.14 (1.78) $t = -1.27$, $d = .11$	−0.20 (1.51) $t = -2.18$*, $d = .19$	−0.36, $d = .04$
Minor physical	−0.14 (1.11) $t = -2.08$*, $d = .18$	−0.27 (1.13) $t = -3.99$***, $d = .34$	−1.17, $d = .12$
Severe physical	−0.10 (0.86) $t = -1.99$*, $d = .16$	−0.17 (0.77) $t = -3.59$***, $d = .31$	−0.77, $d = .09$
Physical	−0.24 (1.64) $t = -2.46$*, $d = .21$	−0.44 (1.63) $t = -4.43$***, $d = .38$	−1.16, $d = .12$
Minor sexual	−0.11 (0.60) $t = -3.02$**, $d = .26$	0.06 (0.82) $t = 1.18$, $d = .10$	2.24*, $d = .24$
Severe sexual	−0.04 (0.23) $t = -2.72$**, $d = .25$	−0.06 (0.43) $t = -2.16$*, $d = .20$	−0.57, $d = .06$
Sexual	−0.14 (0.65) $t = -3.65$***, $d = .30$	0.00 (0.98) $t = 0.00$, $d = 0.00$	1.68, $d = .17$
Minor injury	0.04 (0.46) $t = 1.46$, $d = .12$	0.06 (0.50) $t = 1.95$, $d = .17$	0.30, $d = .04$
Severe injury	0.01 (0.20) $t = 1.12$, $d = .07$	0.02 (0.19) $t = 1.91$, $d = .15$	0.45, $d = .05$
Injury	0.05 (0.54) $t = 1.57$, $d = .13$	0.08 (0.56) $t = 2.48$*, $d = .20$	0.58, $d = .05$
Any violence (not injury)	−0.52 (2.92) $t = -2.97$**, $d = .25$	−0.64 (2.77) $t = -3.81$***, $d = .33$	0.39, $d = .04$

* $P < .05$. ** $P < .01$. *** $P < .001$.

Correlations for agreement on wife behavior ranged from .33 to .42, but were all nonsignificant.

Next, we examined impact of item severity on agreement using four by two univariate ANCOVAs. The factors were type of item (i.e., psychological aggression, physical aggression, sexual coercion, or injury) and severity of item (i.e., minor or severe), with the prevalence of aggression as the covariate and the measure of agreement as the dependent variable. Type of item and prevalence of aggression were included to control for possible confounds in assessing severity and agreement. For wife percent occurrence agreement, prevalence, $F(1, 22) = 19.15$, $p < .001$, type of item, $F(3, 22) = 8.31$, $p < .001$, and the item type by severity interaction, $F(3, 22) = 3.54$, $p < .05$, were significantly associated with agreement (adjusted means: minor psychological aggression = 3.96, severe psychological aggression = 26.67, minor physical aggression = 34.83, severe physical aggression = 35.72, minor sexual coercion = 27.69, severe sexual coercion = 22.87, minor injury = 53.66, severe injury = 34.56). In contrast, none of the independent variables were significantly associated with wife kappa or Yule's *Y*, though type of item, $F(3, 18) = 7.34$, $p < .01$, and the type by severity interaction, $F(3, 18) = 3.31$, $p < .05$, were significantly associated with wife tau-b correlation (adjusted means: minor psychological aggression = −.05, severe psychological aggression = .35, minor physical aggression = .43, severe physical aggression = .42, minor sexual coercion = .36, severe sexual coercion = .16, minor injury = .64, severe injury = .51). Similarly, for husband percent occurrence agreement, prevalence, $F(1, 22) = 28.13$, $p < .001$, type of item, $F(3, 22) = 3.80$, $p < .05$, and the item type by severity interaction, $F(3, 22) = 5.03$, $p < .01$, were significantly associated with agreement (adjusted means: minor psychological aggression = −21.49, severe psychological aggression = 32.12, minor physical aggression = 32.34, severe physical aggression = 38.33, minor sexual coercion = 22.24, severe sexual coercion = 28.76, minor injury = 41.02, severe injury = 29.81). And, as with wife agreement, neither husband kappa nor Yule's *Y* were associated with the independent variables, though husband tau-b correlation was significantly associated with prevalence, $F(1, 18) = 7.82$, $p < .05$, type of item, $F(3, 18) = 5.18$, $p < .01$, and the type of item by severity interaction, $F(3, 18) = 5.86$, $p < .01$, (adjusted means: minor psychological aggression = −.34, severe psychological aggression = .47, minor physical aggression = .46, severe physical aggression = .52, minor sexual coercion = .29, severe sexual coercion = .29, minor injury = .54, severe injury = .26). Across the board, we found a consistent pattern of injury and physical aggression items having the highest agreement, followed by sexual coercion, and lastly by psychological aggression items. Although there were no significant differences between minor and severe items overall, we did find a consistent pattern of similar agreement for minor and severe items on the physical aggression and sexual coercion items, but greater agreement for severe than minor psychological aggression items and greater agreement for minor than severe injury items. The patterns that we found for overall percent agreement were also present, though at nonsignificant levels, for husband and wife kappa and Yule's *Y*. Further detailed statistics are not presented here due to space considerations, but are available from the first author upon request. These results did not provide consistent support of our hypothesis that couples would agree more on severe, rather than mild, items.

Finally, we tested the hypothesis that agreement would be associated with marital satisfaction. As before, percent occurrence agreement, kappa, and Yule's *Y* were used to measure occurrence agreement, and tau-b correlations were used to measure frequency agreement. These were calculated at the couple level by comparing husband and wife reports within a couple on all violence items from the CTS-2. Contrary to expectation, marital satisfaction was not significantly associated with agreement (correlations ranged from $r = -.10$ to $r = .10$ across the different measures of agreement), suggesting that among couples seeking therapy moderately distressed couples are no more consistent with each other in their reports of aggression than severely distressed couples. We also conducted several post hoc correlations to examine the relationship between spousal agreement and demographic variables (i.e., age, years of education, and length of marriage), but none of these variables were significantly correlated with agreement.

Discussion

The main purposes of this article were to replicate previous findings on spousal agreement on the original CTS using the newer, more comprehensive CTS-2, and to expand upon current knowledge by examining the patterns of reporting differences and the correlates of agreement. For most items and subscales, partner agreement, as measured by percent occurrence agreement, kappa, Yule's *Y*, and tau-b correlations, was significantly greater than chance, but within the low-to-moderate range. These results are consistent with other studies of agreement regarding relationship aggression that have used the original CTS for a number of populations (Bohannon et al., 1995; Browning & Dutton, 2003; Edleson & Brygger, 1986; Heckert & Gondolf, 2000; Jouriles & O'Leary, 1985; Kwong, Bartholomew, & Dutton, 1999; Moffitt et al., 1997; O'Leary et al., 1992; Russell & Hulson, 1992; Szinovacz, 1983). Our results suggest that the CTS-2, despite its other improvements over the CTS, does not lead to greater spousal agreement.

Interestingly, when agreement for violent/nonviolent and abusive/nonabusive couple categorizations was examined, the statistics were in the same general range as for items and subscales. These results indicate that not only is agreement between partners poor-to-moderate for specific violent acts and subscales, it is also somewhat limited when used to assess broader questions, such as whether or not a couple has ever experienced violence or whether violence can be considered "abusive." This finding is particularly important, as researchers and clinicians frequently rely upon the CTS, CTS-2, and other similar measures to determine if couples are eligible for basic research studies, clinical trials of couple therapy, and couple therapy in the community.

These findings certainly do not preclude the use of the CTS-2 in research or clinical settings, but they do emphasize the need to gather information from both partners whenever possible and to use statistical techniques that allow for differences in reporting on the same information. For example, techniques like multilevel modeling make it possible to examine predictors of couple functioning and track changes over time while accounting for within couple variance (i.e., differences and similarities between husband and wife report on the same measures). In general, researchers and clinicians should remember that it is unlikely that either partner is presenting a truly objective view of the relationship, but rather one

that is likely to be affected by issues of recall, self-presentation motivations, and beliefs and feelings about the relationship. As such, there may be no real "truth" in assessing for domestic violence through self-report measures, but only individual perceptions and experiences that may or may not agree. If the goal is to obtain "true" estimates of the frequency and prevalence of relationship violence, independent measures, such as medical or legal records, may be used alone or in combination with self- and spousal-report.

The results further revealed that, as expected, both husbands and wives reported that their partner had committed more acts of aggression than the partners reported about themselves, a significant, but relatively small, effect. This pattern of disagreement may be a result of stronger recall of partner negative events or a desire to present oneself as the victim, rather than the aggressor, when reporting relationship violence. This is probably also affected by issues of social desirability, in that few people are likely to wish to portray themselves as violent, especially toward a loved one.

This pattern of disagreement may also be affected by the structure of the CTS-2 in which subjects are asked to report first whether they have committed an act and then whether their partner has. It is possible that as distressed subjects complete the measure, their feeling that their partner is the one to blame for the problems in the relationship affects their reporting and they automatically feel that, no matter how many times they have committed a violent act, their partner must have committed it more frequently. Ordering the items on the CTS-2 so that each item of individual violence is not immediately followed by the same item of partner violence may be useful in examining whether this pattern is a result of the structure of the measure as opposed to issues such as social desirability.

Finally, we assessed whether item specificity/objectivity, item severity, and the marital satisfaction of the couple was associated with the degree of partner agreement. As expected, couples were more likely to agree on the occurrence of more objective/specific items, though these effects were not large and only for some cases. These findings are consistent with the results of Christensen et al. (1983) and Jacobson and Moore (1981), who found that couples were more likely to agree on items that required less inference (i.e., were more objective and specific). There was no significant relationship between item severity and agreement, however an interaction between severity and item type was found, with couples agreeing more on severe than minor psychological aggression items and more on minor than severe injury items. This is probably a result of two factors: the severe psychological aggression items tend to be somewhat more specific and less common than the minor items, while the severe injury items, such as having a broken bone, were extremely rare. In contrast to the findings on item characteristics, the primary couple characteristic that was examined, marital satisfaction, was unrelated to agreement. However, it is possible that the relatively restricted range of this variable may have affected the significance of the associations and that a relationship may be revealed with a broader sample. The current sample consisted of moderately to severely distressed couples that were seeking marital therapy; if a control group of happier couples had been available, we might have found agreement to be greater among satisfied couples. Our results do support the previous research that unhappy couples as a group tend not to agree in reporting marital events, especially negative acts, such as aggression, and exhibit a tendency for spouses to report their partner to have committed more negative acts (i.e., aggression) than they self-report.

Limitations

The main limitation of this study is the nature of the sample. Although the results of this study are consistent with those of a number of studies drawing from several different populations of couples, it is limited by a reliance on heterosexual, married, primarily Caucasian, and primarily middle-class couples. This limitation characterizes much of the research on relationships, which would benefit from a more diverse sample of participants. Also, the sample consists of moderately to severely distressed couples with a negatively skewed range of aggression (i.e., most couples reported relatively low frequency of aggression, with a minority reporting more severe violence). Although the results are consistent with other studies of agreement across different populations, the restricted range in these areas may have limited the finer tuned analyses we attempted in trying to explain why agreement is only low to moderate. Several other factors may have affected the restricted range as well. Although couples were assured of the confidentiality of their responses, it is possible that some respondents may have feared legal repercussions and underreported aggression as a result. In addition, couples in this study were seeking to participate in a therapy program; although there was no indication to the participants that moderate-to-severe violence was an exclusionary criterion for therapy, some may have suspected this and underreported as a result. Finally, we instructed couples to complete the take-home measures separately, but there was no way to guarantee that all participants followed these guidelines. As a result, it is possible that some participants restricted the openness of their partner's responses on the CTS-2.

A second limitation of this study is the reliance on the CTS-2 alone. Although this study was provided important psychometric information for this specific measure, the larger issue of spousal agreement in reporting domestic violence has not been fully addressed. Future research addressing this with different types of measures, both interview and self-report, and exploring the reasons why spouses do not agree, would be useful. Our results did not reveal couple-level characteristics that were related to agreement, but more fine-tuned analysis of the purpose and context of aggression may be of use in further explaining the disparity in couple reports. For example, it is possible that spouses with a history of low-level violence do not recall occasional incidents of aggression because they do not perceive them as important, while spouses with a history of severe violence may deliberately not report certain acts for fear of retribution or legal action. Further research is needed in understanding both the context and function of aggression in relationships, but also in developing measures that provide us the best possible picture of violence in any given relationship.

Conclusions and Implications

The current study provides important additions to the literature on agreement in several areas. First, this study replicates previous findings on spousal agreement on the CTS with the newer, more comprehensive, CTS-2, demonstrating the consistency of poor-to-

moderate interspousal agreement in this area across different versions of the measure. It may be that this is an inherent limitation of any self-report measure of relationship violence, such that spouses are unlikely to agree on the frequency of any list of behaviors in the past year, much less behaviors as emotionally loaded as aggression and abuse. Regardless of whether this is a problem specific to CTS-2 or one that is more general, the results suggest that the most commonly used self-report questionnaire of violence provides two differing viewpoints on the frequency and prevalence of aggression in a relationship. In fact, given the difficulties subjects have with reporting on events in the past 24 hours (Christensen et al., 1983), it may be impossible to obtain high spousal agreement for violence using retrospective self-report methods. Options for improving agreement in violence assessment might include diary methods, in which subjects would report on violent events as they occur, or detailed interviews that use memory prompts. Such efforts have already begun, as can be seen in the development of interviews for domestic violence (Fals-Stewart, 2003; Fals-Stewart, Birchler, & Kelley, 2003; Fals-Stewart, Golden, & Schumacher, 2003; Fals-Stewart, Lucente, & Birchler, 2002; Heyman, Feldbau-Kohn, Ehrensaft, Langhinrichsen-Rohling, & O'Leary, 2001), which focus on reinstating contextual events to improve the subject's memory. In fact, Fals-Stewart and colleagues have demonstrated improvement in agreement through training respondents in diary methods and the Timeline Followback method (Fals-Stewart, Birchler, & Kelley, 2003).

It should be noted, however, that though spousal agreement is an issue when reporting prevalence and frequency statistics for a given population, it may not be as important in other contexts. Many researchers and clinicians are primarily interested in differentiating between abuse and mutual aggression and assessing for danger in a relationship, which goes beyond estimates of prevalence and frequency of aggression. It may be time to develop measures and interviews that examine the context, intentions, perceptions, and effects of violence in a relationship rather than focus primarily on the occurrence of specific behaviors (Dobash et al., 1992).

By examining both item level and couple level correlates of agreement, this study provides new information as to how to improve existing measures of violence. Subjects revealed higher agreement when reporting on items that were more specific and objective, probably because it is easier to recall what events might be included under such items. In addition, such items may be less affected by factors such as social desirability, because it is harder for an individual to convince him/herself that such an event did not really occur when the event is defined clearly. For example, it may be easy for respondents to convince themselves that they have not been spiteful toward toward a partner as often as the partner has been spiteful toward them because we generally like to see ourselves in a positive light. It is easy to believe that *I* am not a spiteful person, but my partner, who I am unhappy with, is spiteful. In contrast, items like "I slapped my partner" delineate specific, objective acts that, if respondents are answering honestly, are hard to convince themselves did not actually happen. This suggests that, in addition to assessing context and purpose of aggression, continued efforts to make items as specific and objective as possible, one of the strengths of the CTS and CTS-2, would probably be of use.

In contrast to the study of individual behavior, the study of relationships provides the researcher with two windows into any events that take place. These windows provide different, sometimes complementary, and sometime contradictory views of those events. Researchers face the continuing challenge of how to integrate those views into a compelling and coherent account of the behavior and dynamics of relationships.

References

American Psychological Association (1996). *APA presidential task force on violence and the family report.* Washington, DC: Author.

Archer, J. (1999). Assessment of the reliability of the conflict tactics scales: A meta-analytic review. *Journal of Interpersonal Violence, 14,* 1263–1289.

Arias, I., & Beach, S. R. H. (1987). Validity of self-reports of marital violence. *Journal of Family Violence, 2,* 139–149.

Barling, J., O'Leary, K. D., Jouriles, E. N., Vivian, D., & MacEwen, K. E. (1987). Factor similarity of the Conflict Tactics Scales across samples, spouses, and sites: Issues and implications. *Journal of Family Violence, 2,* 37–54.

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin, 91,* 3–26.

Bograd, M., & Mederos, F. (1999). Battering and couples therapy: Universal screening and selection of treatment modality. *Journal of Marital and Family Therapy, 25,* 291–312.

Bohannon, J. R., Dosser, D. A., & Lindley, S. E. (1995). Using couple data to determine domestic violence rates: An attempt to replicate previous work. *Violence and Victims, 10,* 133–141.

Browning, J., & Dutton, D. (2003). Assessment of wife assault with the Conflict Tactics Scale: Using couple data to quantify the differential reporting effect. *Journal of Marriage and the Family, 48,* 375–379.

Burger, J. M., & Rodman, J. L. (1983). Attributions of responsibility for group tasks: The egocentric bias and the actor-observer difference. *Journal of Personality and Social Psychology, 45,* 1232–1242.

Cascardi, M., Avery-Leaf, S., O'Leary, K. D., & Slep, A. M. S. (1999). Factor structure and convergent validity of the Conflict Tactics Scale in high school students. *Psychological Assessment, 11,* 546–555.

Christensen, A., Atkins, D., Berns, S., Wheeler, J., Baucom, D., & Simpson, L. E. (2004). Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples. *Journal of Consulting and Clinical Psychology, 72,* 176–191.

Christensen, A., Sullaway, M., & King, C. E. (1983). Systematic error in behavioral reports of dyadic interaction: Egocentric bias and content effects. *Behavioral Assessment, 5,* 129–140.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 10,* 37–46.

Dobash, R. P., Dobash, R. E., Wilson, M., & Daly, M. (1992). The myth of sexual symmetry in marital violence. *Social Problems, 39,* 71–91.

Edleson, J. L., & Brygger, M. P. (1986). Gender differences in reporting of battering incidences. *Family Relations, 35,* 377–382.

Ehrensaft, M. K., & Vivian, D. (1996). Spouses' reasons for not reporting existing marital aggression as a marital problem. *Journal of Family Psychology, 10,* 443–453.

Fals-Stewart, W. (2003). The occurrence of partner physical aggression on days of alcohol consumption: A longitudinal diary study. *Journal of Consulting and Clinical Psychology, 71,* 41–52.

Fals-Stewart, W., Birchler, G. R., & Kelley, M. L. (2003). The timeline followback spousal violence interview to assess physical aggression between intimate partners: Reliability and validity. *Journal of Family Violence, 18,* 131–142.

Fals-Stewart, W., Golden, J., & Schumacher, J. A. (2003). Intimate partner violence and substance use: A longitudinal day-to-day examination. *Addictive Behaviors, 28,* 1555–1574.

Fals-Stewart, W., Lucente, S. W., & Birchler, G. B. (2002). The relationship between the amount of face-to-face contact and partners' reports of domestic violence frequency. *Assessment, 9,* 123–130.

Fincham, F. D., Bradbury, T. N., Arias, I., Byrne, C. A., & Karney, B. R. (1997). Marital violence, marital distress, and attributions. *Journal of Family Psychology, 11,* 367–372.

Gauthier, L. M., & Levendosky, A. A. (1996). Assessment and treatment of couples with abusive male partners: Guidelines for therapists. *Psychotherapy, 33,* 403–417.

Heckert, D. A., & Gondolf, E. W. (2000). Assessing assault self-reports by batterer program participants and their partners. *Journal of Family Violence, 15,* 181–197.

Heyman, R. E., Feldbau-Kohn, S. R., Ehrensaft, M. K., Langhinrichsen-Rohling, J., & O'Leary, K. D. (2001). Can questionnaire reports correctly classify relationship distress and partner physical abuse? *Journal of Family Psychology, 15,* 334–346.

Heyman, R. E., & Schlee, K. A. (1997). Toward a better estimate of the prevalence of partner abuse: Adjusting rates based on the sensitivity of the Conflict Tactics Scale. *Journal of Family Psychology, 11,* 332–338.

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

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

Johnson, M. P. (1995). Patriarchal terrorism and common couple violence: Two forms of violence against women. *Journal of Marriage and the Family, 57,* 283–294.

Johnson, M. P., & Ferraro, K. J. (2000). Research on domestic violence in the 1990's: Making distinctions. *Journal of Marriage and the Family, 62,* 948–963.

Jones, N. T., Ji, P., Beck, M., & Beck, N. (2002). The reliability and validity of the Revised Conflict Tactics Scale (CTS-2) in a female incarcerated population. *Journal of Family Issues, 23,* 441–457.

Jouriles, E. N., & O'Leary, K. D. (1985). Interspousal reliability of reports of marital violence. *Journal of Consulting and Clinical Psychology, 53,* 419–421.

Kwong, M. J., Bartholomew, K., & Dutton, D. G. (1999). Gender differences in patterns of relationship violence in Alberta. *Canadian Journal of Behavioural Sciences, 31,* 150–160.

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

Lucente, S. W., Fals-Stewart, W., Richards, H. J., & Goscha, J. (2001). Factor structure and reliability of the Revised Conflict Tactics Scales for incarcerated female substance abusers. *Journal of Family Violence, 16,* 437–450.

Meyer, S., Vivian, D., & O'Leary, K. D. (1998). Men's sexual aggression in marriage: Couples' reports. *Violence Against Women, 4,* 415–435.

Moffitt, T. E., Caspi, A., Krueger, R. F., Magdol, L., Margolin, G., Silva, P. A., & Sydney, R. (1997). Do partners agree about abuse in their relationship? A psychometric evaluation of interpartner agreement. *Psychological Assessment, 9,* 47–56.

Newton, R. R., Connelly, C. D., & Landsverk, J. A. (2001). An examination of measurement characteristics and factorial validity of the revised conflict tactics scale. *Educational and Psychological Measurement, 61,* 317–335.

O'Farrell, T. J., Van Hutton, V., & Murphy, C. M. (1999). Domestic violence before and after alcoholism treatment: A two-year longitudinal study. *Journal of Studies on Alcohol, 60,* 317–321.

O'Leary, K. D., Heyman, R. E., & Neidig, P. H. (1999). Treatment of wife abuse: A comparison of gender-specific and conjoint approaches. *Behavior Therapy, 30,* 475–505.

O'Leary, K. D., Vivian, D., & Malone, J. (1992). Assessment of physical aggression against women in marriage: The need for multimodal assessment. *Behavioral Assessment, 14,* 5–14.

Pan, H. S., Neidig, P. H., & O'Leary, K. D. (1994). Male-female and aggressor-victim differences in the factor structure of the modified Conflict Tactics Scale. *Journal of Interpersonal Violence, 9,* 366–382.

Russell, R. J. H., & Hulson, B. (1992). Physical and psychological abuse of heterosexual partners. *Personality and Individual Differences, 13,* 457–473.

Schafer, J. (1996). Measuring spousal violence with the conflict tactics scale: Notes on reliability and validity issues. *Journal of Interpersonal Violence, 11,* 572–585.

Schafer, J., Caetano, R., & Clark, C. L. (1998). Rates of intimate partner violence in the United States. *American Journal of Public Health, 88,* 1702–1704.

Schafer, J., Caetano, R., & Clark, C. L. (2002). Agreement about violence in U. S. couples. *Journal of Interpersonal Violence, 17,* 457–470.

Simpson, L. E., Doss, B. D., Wheeler, J., & Christensen, A. (2005). *Physical and verbal aggression among couples seeking therapy: Common couple violence or battering?* Manuscript submitted for publication.

Snyder, D. K. (1997). *Marital Satisfaction Inventory, Revised (MSI-R): Manual.* Los Angeles: Western Psychological Services.

Spitzer, R. L., & Fleiss, J. L. (1974). A re-analysis of the reliability of psychiatric diagnosis. *British Journal of Psychiatry, 125,* 341–347.

Spitznagel, E. L., & Helzer, J. E. (1985). A proposed solution to the base rate problem in the kappa statistics. *Archives of General Psychiatry, 42,* 725–728.

Straus, M. A. (1979). Measuring intrafamily conflict and violence: The conflict tactics (CT) scales. *Journal of Marriage and the Family, 41,* 75–88.

Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised conflict tactics scales (CTS2): Development and preliminary psychometric data. *Journal of Family Issues, 17,* 283–316.

Szinovacz, M. E. (1983). Using couple data as a methodological tool: The case of marital violence. *Journal of Marriage and the Family, 45,* 633–644.

Thompson, S. C., & Kelley, H. H. (1981). Judgments of responsibility for activities in close relationships. *Journal of Personality and Social Psychology, 41,* 469–477.

Walker, L. E. (1999). Psychology and domestic violence around the world. *American Psychologist, 54,* 21–29.

Received March 17, 2003
Revision received March 21, 2005
Accepted April 4, 2005 ■