J Fam Viol (2007) 22:703–708
DOI 10.1007/s10896-007-9118-7

ORIGINAL ARTICLE

Test–Retest Reliability of the Revised Conflict Tactics Scales (CTS2)

Edward M. Vega · K. Daniel O'Leary

Published online: 31 July 2007

Abstract The Revised Conflict Tactics Scales (CTS2) are widely used to measure intimate partner aggression. Various aspects of the CTS2 and its subscales have been examined, e.g. factor structure, internal consistency, and validity, but little or no evidence exists regarding the stability of the self-report of aggression on the CTS2 This study examines the stability of reports of aggression against a partner on the CTS2 among 82 men court-mandated to a batterer intervention program. At both testing times men reported on aggression occurring in the year prior to beginning treatment. Using variety scores, stability of report was strong for psychological aggression (r=0.69), physical assault (r=0.76), injury (r=0.70), and negotiation (r=0.60), but weaker for sexual coercion (r=0.30). Implications for the use of the CTS2 in court-mandated populations are discussed.

Keywords Test–retest reliability · Conflict Tactics Scale · Partner violence · Reliability · Aggression · Measurement

Professionals have an increasing recognition that physical aggression against intimate partners occurs in a significant percentage of the population (Schafer et al. 1998; Straus and Gelles 1990; Tjaden and Thoennes 1998). Further, the deleterious effects of such aggression have been documented in both community and clinical samples (Byrne et al. 1999; Cascardi et al. 1992; O'Leary and Arias 1988b; Sedlak et al. 1985). Moreover, participants in numerous research studies have shown a willingness to report acts of physical aggression by themselves and their partners. As a result, greater emphasis has been placed on the accuracy of reports of physical aggression by an individual and his/her partner, while recognizing that reliance on self-reports may result in a consistent under-reporting of such aggression (Archer 1999; Arias and Beach 1987; Gelles 1979; Riggs et al. 1989).

Despite some obvious problems associated with self-reports of negative behaviors, such measures are commonly used for assessing the prevalence of intimate partner aggression. The Conflict Tactics Scales (CTS; Straus 1979) are the most commonly used self-report measures of marital and courtship violence. They were designed to assess the various types of behaviors used to solve conflicts between intimate partners. A modified form of the CTS, the Revised Conflict Tactics Scales (CTS2; Straus et al. 1996) now consists of 39 item pairs, asking respondents to report how many times in a given time period the respondent or their partner has engaged in a list of behaviors that sometimes occur during relationship conflicts, ranging from showing concern and respect to severe physical assaults. This revised version includes four subscales that assess different types of tactics used (negotiation, psychological aggression, physical assault, and sexual coercion), as well as the injury subscale that addresses the impact of violence.

Several criticisms of the CTS have arisen, both on the basis of conceptual/ theoretical concerns (see summary in Straus 1990) and concerns about the psychometric properties of the instrument (Archer 1999; Schafer 1996). For example, because the CTS2 itself may demonstrate different psychometric properties depending on the population in which it is used, Straus (1993) recommended the gathering of information about the CTS2 among specific populations.

E. M. Vega (✉)
Atlanta Veterans Affairs Medical Center,
1670 Clairmont Road,
Decatur, GA 30033, USA
e-mail: Edward.Vega@va.gov

K. D. O'Leary
Psychology Department, Stony Brook University,
Stony Brook, NY 11794-2500, USA
e-mail: K.D.OLeary@sunysb.edu

Specifically, it seemed desirable to collect data on populations which may be characterized by higher levels of aggression than the population of college women among whom it was originally evaluated. Indeed the factor structure of the CTS2 has been found to differ in populations characterized by high levels of aggression (e.g., Jones et al. 2002), and according to gender (Schafer 1996).

Additionally, due to the reality that the actual level of aggression in a given couple is not known, there is no "gold standard" against which to test the actual precision of the Conflict Tactics Scales. One consequence of the absence of a gold standard of aggression is an increased focus on the reliability of the CTS, and specifically the common reliance on inter-rater agreement as an estimate of the reliability of the instrument. CTS rates of spousal agreement have been found to be low or moderate with the CTS1 (Browning and Dutton 1986; Jouriles and O'Leary 1985; O'Leary and Arias 1988a; Moffitt et al. 1997) as well as with the CTS2 (O'Leary and Williams 2006). For example, Browning and Dutton (1986) found partner agreement in the form of Pearson correlations ranging from 0.32 to 0.57 across items (excluding items involving use of a weapon). Jouriles and O'Leary (1985) found mean inter-spousal report Pearson correlations for clinic and community couples of 0.59 and 0.41 respectively; kappa coefficients were slightly lower, at 0.43 and 0.40 for report of husband violent behavior (for the clinic and community sample, respectively). Data from Moffitt et al. (1997) are closely related. The study utilized items from the CTS, but did not use entirely and exclusively CTS items. Using variety scores, or number of different aggressive behaviors reported, they obtained the following reliabilities (Pearson product–moment correlations): 0.58 for physical aggression by men to women and 0.53 for psychological aggression of men to women. With the CTS2, in a sample of 453 randomly recruited parents of 3- to 7-year-old children, Kappa for physical aggression by men was 0.47 and by women 0.45 (O'Leary and Williams 2006). Across studies, depending upon the measurement strategy used to calculate reliability, inter-partner agreement ranges from near medium to large using the Cohen and Cohen (1983, p. 61) guide for correlations of low (0.10), medium (0.30), and large (0.50).

As compared to assessments of other types of behavior, measurement of aggression presents additional difficulties in terms of estimating reliability. First, because the assessment is of actual behavior (which may change over time, or may have a low base rate, in the case of extreme behaviors such as severe physical assault) rather than attitudes or personality characteristics (which, often by definition should remain stable), stability potentially takes on a different meaning. Specifically, one may examine stability of the *behavior itself*, or stability of the *report of the behavior* (holding constant the behavior itself, by assessing the same time period twice). Second, because the assessment is of actual behavior, the occurrence of the target event may not be normally distributed; rather, base rates may be extremely high or extremely low. Both cases can present statistical difficulties. Additionally, inter-rater correlations may not be a fair estimate of reliability, due to different motivations affecting self-report versus report of partner behavior (e.g., social desirability) and fear of (or desire for) negative consequences.

The current study focused on assessing the test–retest reliability of the self-report of aggression on the Revised Conflict Tactics Scales, as one means of furthering the field of evidence on the overall reliability of the scale. Men were asked at two different occasions to report on the aggression during the same time period. Therefore, this analysis is of the stability of self-report of aggression, rather than of the stability of aggression itself.

Method

Participants

A total of 87 men who were court-mandated to a psychoeducational intervention following arrest for a domestic violence offense were evaluated twice, with an interval of approximately nine weeks. The men were participants in an 18-week group batterer intervention program, based on the Duluth Model (Pence and Paymar 1993). All program participants who were able to read English were initially included in the study. Eighteen of the original participants (from a total of 105 men who completed the first evaluation) did not complete the second assessment. Due to the anonymous data collection procedure it could not be determined why particular men did not complete the second assessment. Possible reasons include client dropout from treatment, transfer to another treatment group, or expulsion from the program. Of the 87 men who completed both assessments, data from four individuals were removed due to missing data. Data from one additional individual was removed as an outlier, with a pattern of responding that appeared to be intentionally and spuriously aggressive. Consequently, all analyses involved a total of 82 participants. Thus, the sample herein is based on 82 of the 105 men who were asked to participate (79%).

Fifty-seven percent of the participating men were Caucasian (n=47), 22% were of African or Caribbean-American descent (n=18), 16% were Latin American (n=13), and 5% were of other ethnic backgrounds (n=4). The participants ranged in age from 17 to 61 (median=36 years of age). Median level of income was $35,000 (ranging from $6,000 to $155,000). Seventy-six percent reported being employed full time; 13% reported being unemployed. Median level of education completed was completion of high school or a GED.

Procedure

Participants completed questionnaires as part of their participation in a batterer intervention program consisting of 18 sessions, held once per week. They were asked to complete the written questionnaires during their first session in the program, and again approximately nine weeks later. Due to the treatment groups utilizing an "open group" model, participants' first and tenth sessions did not all coincide with each other; participants completed questionnaires apart from the group, typically in a separate room. The data were collected in an anonymous manner; data sets were matched on the basis of identification numbers chosen by the participants themselves. The choice of nine weeks as an interval of assessment for test–retest reliability was chosen for the purpose of minimizing the subjects' recall of their precise questionnaire answers, while also minimizing any possible decrement in memory for the actual events being queried. There was a possibility that being in a batterer intervention program would result in reports of fewer acts of physical aggression following treatment even though the time period in question was before the program began, namely reports of aggression in the year prior to the initial assessment. However, as will be seen in the means on the various scales at the initial and second assessment, reports of physical aggression did not change from the first report of the aggression to the second report.[1]

Measures

Revised Conflict Tactics Scales The Revised Conflict Tactics Scales (CTS2; Straus et al. 1996) consist of 39 item pairs, assessing both positive and negative relationship behaviors that may occur in the context of relationship conflict. The paired items ask respondents to report acts that they have committed towards a partner (perpetration), as well as acts committed by a partner towards them (victimization), during a 1-year period. The items on the CTS2 can be separated into five scales: psychological aggression, physical assault, sexual coercion, injury, and negotiation items. Based on suggestions by the authors of the CTS2, each question is rated on a scale of 0 to 7 (has never happened, happened 1 time, 2 times, 3–5 times, 6–10 times, 11–20 times, more than 20 times, or has happened, but not in the time period in question). The items of each of the five scales on the CTS2 appear interspersed throughout the measure, rather than appearing in sets in accord with the various scales.

In this study, men completed *both* administrations of the CTS2 in reference to behaviors occurring in the 1-year period prior to the beginning of treatment.

[1] Details regarding the lack of change from the initial to the second assessment can be obtained from the authors.

Results

To determine the prevalence and chronicity of aggression in this sample (for the purpose of demonstrating the comparability of this data to other samples), the Revised Conflict Tactics Scales was scored using the midpoint value method described by Straus et al. (1996). Following Straus et al. (1996), prevalence was calculated as the percentage of men who reported the occurrence of any behavior of a given scale within the test period (i.e., reporting the occurrence, within the year prior to treatment, of any of the items on the Physical Assault scale would indicate a positive score for that scale). Also, as described by Straus et al. (1996), chronicity was calculated only from among those participants who reported at least one act on a given scale; it refers to the sum total of all reported occurrences of all acts from that scale.

Although all items were administered, one item, "forced sex without a condom," was excluded from analyses of the Sexual Coercion subscale. This item has been found to load as well on the Negotiation and Psychological Aggression subscales as it does on Sexual Coercion subscale (Lucente et al. 2001), and has previously been removed from some analyses of CTS2 data because of alleged confusion of respondents about the item (Jones et al. 2002; O'Leary and Williams 2006). In this study several subjects seemed to misunderstand the question, and either asked questions, left it blank, or wrote comments to the extent that they routinely did not use condoms with their partner (indicating a misunderstanding of the intent of the question). Thus, the Sexual Coercion subscale was tabulated without the inclusion of this item. Table 1 shows the prevalence and chronicity of each of the CTS2 scales regarding men's self-report, at Time 1 and at Time 2. Table 2 shows men's reports (at Time 1 and Time 2) of their partner's aggression.

Reliability of self-reports and reports of partner aggression were calculated using both frequency scores and variety scores. Frequency scores are derived using the midpoint substitution method, which provides a means of allowing a respondent to estimate not simply the types of different aggressive behaviors engaged in but to also provide an estimate of the frequency of their occurrence (Straus and Gelles 1990; Straus et al. 1996). Specifically, scores for each subscale consist of the total number of acts reported, using the midpoint of each range as the total for that item. For example, if an individual reported they had committed each act on a six-item subscale 6–10 (the midpoint being 8) times in the past year, the total score for that subscale would be 48 (i.e. 8 occurrences for each of 6 items totaling 48 occurrences). The variety score method calculates the sum of the different types of aggressive behaviors engaged in by the self and partner, without regard to the number of occurrences of each type of behavior (i.e. each item is scored

Table 1 Prevalence and chronicity of self-reported male aggression (CTS2)

Scale	Time 1	Time 2
Negotiation		
Prevalence	96.3%	96.3%
Chronicity		
Mean	76.18	75.81
SD	40.04	38.26
Psychological Aggression		
Prevalence	89.0%	93.9%
Chronicity		
Mean	35.22	33.42
SD	27.26	32.99
Physical Assault		
Prevalence	70.7%	67.1%
Chronicity		
Mean	11.10	9.55
SD	15.34	15.49
Sexual Coercion		
Prevalence	24.4%	24.4%
Chronicity		
Mean	5.25	9.00
SD	6.30	14.08
Injury		
Prevalence	48.8%	46.3%
Chronicity		
Mean	6.73	4.24
SD	9.98	7.24

$N=82$

dichotomously for occurrence or non-occurrence, so in the above example, the total variety score would be 6 for the subscale).

Variety scores have been found to be less skewed than frequency scores, more balanced in rating aggressive acts (compared to frequency scores which give additional weight to less serious acts that occur more often), and more reliable than frequency scores, especially among individuals who are involved in a large number of aggressive acts (Elliott and Huizinga 1989; Hirshi et al. 1980). However, with batterers who engage in relatively high rates of partner aggression, the midpoint substitution scoring method may also be of interest since one may wish to know if their physical aggression decreases, even if it does not cease entirely, and such a decrease might be more detectable using a frequency measure. Thus, the two methods may each be useful, depending on specific populations and contexts. Consequently, we present data using both methods.

As is typical of aggression data, several of the distributions violated assumptions of normality (e.g., Physical Assault, Sexual Coercion, and Injury were highly skewed and kurtotic, while Negotiation behaviors were normally distributed). As such, differences were examined with parametric as well as nonparametric statistics. Consistent with findings from previous research, participants tended to report that they themselves committed more positive behaviors and less negative behaviors compared to their partners (Browning and Dutton 1986). At Time 1 men's self-reports, based on midpoint substituted data (using the Wilcoxon T test), were significantly higher than their reports of partner behavior on the Negotiation scale ($T=-3.56$, $p<0.001$); men's self-reported aggression towards partner was lower than their reports of their partner's aggression towards them on both the Psychological Aggression scale ($T=-3.78$, $p<0.001$) and the Physical Assault scale ($T=-2.60$, $p<0.01$). Men's self-report and report of partner on the Sexual Coercion scale and Injury scale did not differ significantly. The same pattern emerged at Time 2, with significant differences in Psychological Aggression ($T=-3.15$, $p=0.002$), Physical Assault Scale ($T=-3.95$, $p<0.001$), and Negotiation ($T=-2.17$, $p=0.030$). Similar results were obtained with parametric tests; however, because they represent the more conservative test, only the nonparametric Wilcoxon T tests are presented.

Test–retest reliability of the CTS2 was assessed using Pearson product–moment (PPM) correlations. Although the data for most of the CTS2 subscales were non-normally distributed, the distributions for each subscale were

Table 2 Prevalence and chronicity of report of partner aggression (CTS2)

Scale	Time 1	Time 2
Negotiation		
Prevalence	95.1%	97.6%
Chronicity		
Mean	64.54	68.51
SD	38.64	40.11
Psychological Aggression		
Prevalence	90.2%	93.9%
Chronicity		
Mean	44.43	38.92
SD	32.12	33.90
Physical Assault		
Prevalence	74.4%	69.5%
Chronicity		
Mean	21.23	19.56
SD	36.18	28.74
Sexual Coercion		
Prevalence	24.4%	25.6%
Chronicity		
Mean	7.35	11.24
SD	10.93	16.25
Injury		
Prevalence	45.1%	39.0%
Chronicity		
Mean	6.14	7.50
SD	7.77	16.78

$N=82$

Table 3 Test–retest reliability of men's self-report of aggression, and report of partner aggression (Pearson product–moment), based on frequency scores and variety scores

Scale	Frequency scores		Variety scores	
	Self-report	Partner-report	Self-report	Partner-report
Negotiation	0.486**	0.617**	0.602**	0.672**
Psychological aggression	0.716**	0.650**	0.693**	0.708**
Physical assault	0.677**	0.863**	0.762**	0.776**
Sexual coercion	0.666**	0.798**	0.303*	0.538**
Injury	0.794**	0.531**	0.699**	0.718**

N=82
**p<0.001
*p<0.005

relatively similar at Time 1 and Time 2, permitting valid comparisons to be made (Nunnally 1978). Standards have been suggested for magnitudes of correlations that may be meaningful, independent of the level of statistical significance, e.g. Cohen's guidelines that view correlations below 0.30 as low, correlations between 0.30 and 0.50 as moderate, and correlations above 0.50 as high (Cohen and Cohen 1983). Thus, although significance levels (p values) are reported here, the primary focus is on the magnitude of the correlations. As shown in Table 3, test–retest reliability (based on frequency data) was high for the four aggression subscales for both self-report of aggression and report of partner aggression. Negotiation (0.49) was very close to the Cohen and Cohen (1983) high criterion, namely 0.50. Further, using variety scores, reliability of reports of the four aggression scales as well as negotiation were high (Table 3). A similar pattern of findings occurred based on Spearman correlations, although correlations based on rank-ordered data were somewhat attenuated.

Discussion

Excellent test–retest reliability existed for most of the CTS2 scales over approximately a 2-month interval (using Cohen's guide for low correlations less than 0.30; moderate correlations 0.30–0.50, and high correlations 0.50 and greater; Cohen and Cohen 1983). It is important to remember that at the two testing sessions participants were asked to complete the CTS2 in reference to the exact same time period—the full year prior to their beginning the treatment program. Thus, the analysis is precisely of men's *reports* of aggression, with no chance that the events themselves have changed.

The 2-month stability of men's reports of their own aggression and their partners' aggression indicates strong test–retest reliability. Other measures conceptually related to the CTS2 include the Novaco Anger Scale, which demonstrated 1-month test–retest reliability values ranging from 0.78 to 0.91 (Mills et al. 1998), and the Response to Conflict Scale, which showed 2-week retest reliability coefficients of 0.78 and 0.81 (Birchler and Fals-Stewart 1994).

Comparisons of the frequency scores and variety scores did not indicate that one had higher reliabilities than the other. In fact, of the ten comparisons, (five measures compared at Time 1 and Time 2 for both variety scores and frequency scores), variety scores showed numerically higher reliabilities for five comparisons, and frequency scores showed numerically higher reliabilities for the other five comparisons. In examination of the ten corresponding Spearman correlations (of midpoint substituted and variety scores), three were numerically higher based on midpoint substituted scores as well as variety scores. However, the differences between the correlations were not significant. Thus, the overall findings were robust across various methods of analysis.

One would expect that all the men in a sample of men referred by the court to a program because of their aggressive behavior or threats against their partner would have, in the last year, committed at least one act classified as physical aggression, or, at minimum, one act of psychological aggression on the CTS2. In fact, only 72.4% of the men reported any physically assaultive behavior at Time 1, and only 89% reported that they engaged in any acts of psychological aggression against their partners. Thus, these figures presumably suggest some degree of underreporting in these self-reports of aggression. However, the current findings present evidence of excellent stability for the self-report of aggression by men in a court-mandated population on the aggression subscales of the CTS2: Psychological Aggression, Physical Assault, Sexual Coercion, and Injury. Additionally, the Negotiation subscale showed good stability in this high-risk sample.

Our data show that men mandated to treatment are quite consistent in their reports of their physical aggression against partners. Moreover, despite the fact that these men were in a court mandated program for batterers, with the exception of about 10% of men who denied engaging in any physical aggression, the men admitted to a relatively large number of physically aggressive acts per year on the anonymous reports (10–11 physically aggressive acts against their partners). In short, using a method of retrospective reporting, there was high stability of reports of psychological aggression, physical aggression, and injury for frequency scores. For variety scores, with the exception of sexual coercion, stability of reports was high. These data provide information about stability of reports of aggression that may be deemed essential in treatment outcome studies with samples of batterers.

It is noteworthy that this sample of men was mandated to a batterer intervention program that clearly articulates the

socially undesirable nature of physical aggression, and emphasizes that intimate partner aggression is illegal. As such, it seems plausible that men might be motivated to report less aggression over time. Correlations of rank-ordered test–retest data might still be high in such a case. However, as noted, in the current study the men did not decrease the amounts of physical aggression they reported for the time frame in question. It would be helpful for additional research to evaluate the test–retest stability of self-reports aggression for different time periods, such as 4–6 weeks apart, as well as longer time frames. Similarly, obtaining stability data from both men and women would be helpful in knowing more about perceptions of aggression by males and females, especially in populations where reports of physical aggression are quite high. It would be very helpful if other projects continue to use a similar methodology, in terms of assessing the same time period at both assessment points, namely the full year prior to the beginning of treatment. If such research yielded stability data similar to those herein, one could place even greater confidence that men can provide stable reports of their use of physical aggression by themselves and their partners. Similar data from partners may help elucidate the extent to which men in treatment programs may underestimate their own aggression while overestimating the aggression of their partners.

References

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.

Birchler, G. R., & Fals-Stewart, W. (1994). The Response to Conflict Scale: Psychometric properties. *Assessment, 1*(4), 335–344.

Browning, J., & Dutton, D. (1986). 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.

Byme, C. A., Resnick, H. S., Kilpatrick, D. G., Best, C. L., & Saunders, B. E. (1999). The socioeconomic impact of interpersonal violence on women. *Journal of Consulting and Clinical Psychology, 67*, 362–366.

Cascardi, M., Langhinrichsen, J., & Vivian, D. (1992). Marital aggression, impact, injury, and health correlates for husbands and wives. *Archives of Internal Medicine, 152*, 1178–1184.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/ correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Elliott, D. S., & Huizinga, D. (1989). Improving self-reported measures of delinquency. In M. W. Klein (Ed.), *Cross-national Research in Self-reported Crime and Delinquency* (pp. 155–186). Dordrecht, The Netherlands: Kluwer.

Gelles, R. J. (1979). The myth of battered husbands and new facts about family violence. *Ms., 66*, 71, October.

Hirschi, T., Hindelang, M. J., & Weis, J. G. (1980). The status of self-report measures. In M. W. Klein & K. S. Teilmann (Eds.), *Handbook of Criminal Justice Evaluation* (pp. 473–488). Beverly Hills, CA: Sage.

Jones, N. T., Ji, P., Beck, M., & Beck, N. (2002). The reliability and validity of the Revised Conflict Tactics Scale (CTS2) 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.

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.

Mills, J. F., Kroner, D. G., & Forth, A. E. (1998). Novaco Anger Scale: reliability and validity within an adult criminal sample. *Assessment, 5*, 237–248.

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

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York, NY: McGraw-Hill.

O'Leary, K. D., & Arias, I. (1988a). Assessing agreement of reports of spouse abuse. In G. T. Hotaling, D. Finkelhor, et al. (Eds.), *Family Abuse and its Consequences: New Directions in Research* (pp. 218–227). Newbury Park, CA: Sage.

O'Leary, K. D., & Arias, I. (1988b). Prevalence, correlates and development of spouse abuse. In R. D. Peters & R. J. McMahon (Eds), *Social Learning and Systems Approaches to Marriage and the Family* (pp. 104–127). New York: Brunner/Mazel.

O'Leary, K. D., & Williams, M. C. (2006). Agreement about acts of aggression in marriage. *Family Psychology, 20*, 656–662.

Pence, E., & Paymar, M. (1993). *Education Groups for Men who Batter: The Duluth Model*. New York: Springer.

Riggs, D. S., Murphy, C. M., & O'Leary, K. D. (1989). Intentional falsification in reports of interpartner aggression. *Journal of Interpersonal Violence, 4*, 220–232.

Schafer, J. (1996). Measuring spousal violence with the Conflict Tactics Scale: Notes on reliability and validity issues. *Journal of Interpersonal Violence, 11*(4), 572–585.

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

Sedlak, A. J., Bowen, G. L., Hotaling, G. T. & Sugarman, D. B. (1985). *Domestic violence risk factors and consequences: Current state of knowledge*, Paper presented at the 93rd Annual Convention of the American Psychological Association, Los Angeles, CA.

Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family, 41*, 75–86.

Straus, M. A. (1990). The conflict tactics scales and its critics: An evaluation and new data on validity and reliability. In M. A. Straus & R. J. Gelles (Eds.), *Physical Violence in American Families: Risk Factors and Adaptations to Violence in 8,145 Families* (pp. 49–73). New Brunswick, NJ: Transaction.

Straus, M. A. (1993). Identifying offenders in criminal justice research on domestic assault. *American Behavioral Scientist, 36*, 587–600.

Straus, M. A., & Gelles, R. J. (Eds.) (1990). *Physical Violence in American Families: Risk Factors and Adaptations to Violence in 8,145 Families.* New Brunswick, NJ: Transaction.

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.

Tjaden, P., & Thoennes, N. (1998). Prevalence, incidence, and consequences of violence against women: Findings from the national violence against women survey. In National Institute of Justice Centers for Disease Control and Prevention, *Research in Brief*, November 1998.