Self-Reports of Sexual, Physical, and Nonphysical Abuse Perpetration: A Comparison of Three Measures

Sarah L. Cook

Violence Against Women 2002 8: 541
DOI: 10.1177/107780102400388425

Self-Reports of Sexual, Physical, and Nonphysical Abuse Perpetration

A Comparison of Three Measures

SARAH L. COOK
Georgia State University

This study explored self-reports of sexual, physical, and nonphysical abuse perpetration obtained with three measurement instruments. A randomly selected sample of incarcerated men responded to the Sexual Experiences Scale (SES), the Conflict Tactics Scale–Revised (CTS2), and the Severity of Violence Against Women Scales (SVAWS). Results show that the proportion of men who reported sexual aggression, but not physical or nonphysical aggression, differed according to instrument. Findings also show the absence of main effects for social desirability on self-reports obtained with all violence against women instruments. In contrast, a robust effect for social desirability on self-reports of general aggression emerged.

When planning research, investigators face complex choices that critically influence study outcomes. Among the most vexing may be which measure is best suited for a particular research question. Numerous published instruments are available to assess self-reported physical, sexual, or nonphysical (e.g., emotional or psychological) victimization or perpetration. All vary according to characteristics, such as the content and number of items, or how subscales define types of aggression and levels of severity. In

AUTHOR'S NOTE: This study was funded through a dissertation research grant by the Center for Children, Families and the Law at the University of Virginia. Many administrators, staff, and inmates at the Commonwealth of Virginia Department of Corrections facilitated this project, including Mary Reinman, Paula Harpster, Thomas Neumeyer, Barbara Wheeler, Marsha Altizer, Kim Ordille, and Cathy Shearn. Inmates Emmet Shorter, D. C. Mason, and Ahmed Campbell provided feedback on the interview process and assisted with training undergraduate research assistants. University of Virginia undergraduates Clark Whitney, Aristotle Chronis, Jay Wechsler, Tim Leach, Henry Rivera, and David Phillips assisted with interviews and data entry. This study was based on the author's dissertation chaired by N. Dickon Reppucci, Ph.D.

VIOLENCE AGAINST WOMEN, Vol. 8 No. 5, May 2002 541-565

combination, these characteristics convey assumptions about the nature of victimization and perpetration and produce definitions that range from broad to narrow. As a result, comparing findings from studies that use different measures is tenuous.

Empirically comparing measurement approaches can indicate how different instruments lead to disparate findings. Currently, more comparative data are available on methods of assessing victimization than perpetration. For example, a growing body of research demonstrates the superiority of behaviorally specific questions to assess rape and sexual assault (for review, see Koss & Cook, in press; Weaver, 1998). Among other factors, differences in survey context (crime vs. personal safety) and the potential presence of intimate partners, who may be perpetrators, may account for lower incidence rates of physical and sexual assault victimization obtained from the National Crime Victimization Survey (NCVS) than from the National Violence Against Women Survey (Bachman, 2000). Others have found inconsistencies in the stability of reports of sexual assault over time (Scott & Aneshensel, 1997) and within different time frames (Hilton, Harris, & Rice, 1998). Together, these findings begin to elucidate how measurement choice and methodological factors, in addition to sociodemographic (Scott & Aneshensel, 1997) or motivational factors (Hilton et al., 1998), may affect the degree, validity, and reliability of disclosure (for review, see Koss, 1992, 1993; Schwartz, 2000; Spitzberg, 1999).

Aside from studies documenting disagreement between partner reports (e.g., Bohannon, Dosser, & Lindley, 1995; Meyer, Vivian, & O'Leary, 1998; O'Leary et al., 1989), little is known about factors that influence self-reports of perpetration. Two reviews identify methodological factors, such as sample selection, that may confound self-reports of sexual perpetration (Porter & Critelli, 1992; Spitzberg, 1999). One study documented a higher level of disclosure of physical violence through oral versus written self-report methods (O'Leary, Vivian, & Malone, 1992). Hilton and colleagues (1998) found male perpetrators' reports less reliable over three reporting periods (1, 6, and 12 months) than female victims' reports. No studies, however, directly compare self-reports obtained with multiple measures or test the effects of methodological factors within the same sample.

This study examines self-reports of sexual, physical, and nonphysical aggression in a randomly selected sample of 160 incarcerated men. Data were obtained with the Sexual Experiences Scale (SES) (Koss, Gidycz, & Wisniewski, 1987), the Conflict Tactics Scale–Revised (CTS2) (Straus, Hamby, Boney-McCoy, & Sugarman, 1996), and the Severity of Violence Against Women Scales (SVAWS) (Marshall, 1992). In addition to comparing the level of disclosure across instruments, the investigation examines the effect of social desirability on self-reports of violence against women. It also examines whether a methodological factor, defined by measurement conditions of matched or unmatched interviewer and inmate sex, moderates a potential relation between social desirability and self-report. To assess whether social desirability poses a particular threat to self-reports of violence against women, this same model is tested using self-reported general aggression as the dependent variable.

The SES, CTS2, and SVAWS spring from different disciplines (sociology and psychology) and different fields of study (sexual assault and family violence). Not surprisingly, each undertakes a different approach. Below is a brief comparison of these instruments' development, content, structure, and use. To clarify language, this article uses the terms *sexual*, *physical*, and *nonphysical aggression* generally, except when referring to specific subscales in each instrument (e.g., SES sexual coercion or rape subscales).

SEXUAL EXPERIENCES SURVEY (SES)

Although Kanin (1957) documented a high level of sexual aggression among acquaintances as early as the 1950s, most sexual aggression studies through the mid-1980s relied on samples of incarcerated sex offenders. These studies usually measured sexual aggression as a function of criminal history or by physiological responses to sexual stimuli (i.e., plethysomography). As data on sexual aggression between acquaintances in nonincarcerated samples accumulated (e.g., Koss et al., 1987; Malamuth, 1986; Rapaport & Burkhart, 1984), the need for a new approach was apparent. Two types emerged. Malamuth (1989a, 1989b) developed the Attraction to Sexual Aggression Scale (ASA) to measure proclivity to sexually aggress. Others developed behavioral self-

report measures, for example, the Coercive Sexuality Scale (Rapaport & Burkhart, 1984) and the Sexual Experiences Survey (Koss et al., 1987; Koss & Gidycz, 1985; Koss & Oros, 1982).

Used in the first national prevalence and incidence study of sexual aggression among higher education students (Koss et al., 1987), the SES employs behaviorally specific definitions of sexual aggression. Subscales include unwanted sexual contact, sexual coercion, and rape. Subsequently, the majority of acquaintance rape research has relied on the SES (Porter & Critelli, 1992). Even so, some critics question the inclusion of items that ask about nonconsensual intercourse (and attempts) obtained through intoxication of the victim, a component of most reformed state statute definitions of rape (Gilbert, 1997; Koss & Cook, in press). However, the SES has been used in subsequent additional large-scale studies as the best available instrument (e.g., Merrill et al., 1998). Item revisions have been published (e.g., Lisak & Roth, 1988), but how self-reports obtained with revised versions compare with original instrument is not clear.

THE CONFLICT TACTICS SCALE (CTS) AND THE CONFLICT TACTICS SCALE–REVISED (CTS2)

Empirical research on physical and psychological violence against women developed separately from rape research, within the study of family violence, pioneered by Gelles and Straus (1979). The most frequently used instrument in studies of family violence across the social sciences is the CTS. Extensive data document the instrument's reliability and validity in diverse populations (e.g., military, college, community, and clinical samples). It is grounded in the theory that conflict, an inevitable part of human relationships, is resolved through rational discussion, verbal and nonverbal attacks, or physical force (Straus, 1990/1995).

A number of critics call attention to several shortcomings in the CTS. A comprehensive review of limitations is beyond the scope of this article, but criticism relevant to this study emphasizes that the CTS measures only conflict-related aggression, defined by a small set of behaviors that range in severity but are considered equally serious (for detailed critiques and responses, see DeKeseredy & Schwartz, 1998; Dobash, Dobash, Wilson, & Daly, 1992; Straus, 1990/1995). Furthermore, it omits sexual aggression

as a tactic. In response, Straus and colleagues (1996) revised the CTS by including new items to assess negotiation and conflict tactics, refining some existing ones for greater specificity, improving the operationalization of minor and severe levels of aggression, and adding subscales for injury and sexual coercion.

SEVERITY OF VIOLENCE AGAINST WOMEN SCALES (SVAWS)

Compared to the SES and CTS, the SVAWS is relatively new. Developed "in response to the need for more sensitive assessment instruments" (Marshall, 1992, p. 103), it includes nine subscales that define symbolic violence; mild, moderate, and serious threats; minor, mild, moderate, and serious physical violence; and sexual violence. Including the SVAWS in this study allows a comparison between the CTS2 and an instrument that uses a greater number of items and subscales that range in seriousness. The SVAWS is also distinctive because items that define nonphysical aggression focus only on threatened, attempted, or completed behaviors that could potentially cause injury or pain (Marshall, 1992). The SVAWS does not purport to measure psychological aggression per se; however, many items in the symbolic violence and threats of mild, moderate, and serious violence subscales overlap with other measures of nonphysical aggression (e.g., Hudson & McIntosh, 1981) and emotional abuse (e.g., Tolman, 1989). In the field, many terms describe similar phenomena (e.g., *controlling behavior*, *emotional abuse*, *psychological maltreatment*).

The impact of these instruments' history, development, and content is evident by comparing operational definitions of sexual, physical, and nonphysical aggression. The instruments define sexual and nonphysical aggression differently but define physical aggression similarly.

SEXUAL AGGRESSION

Most apparent is that these instruments contain different numbers of items and subscales. Ten SES items comprise sexual contact, sexual coercion, and rape/attempted rape subscales that assess sexual aggression against any woman. Seven CTS2 items constitute minor and severe subscales of sexual coercion, and

SVAWS uses six to form one sexual violence subscale. The CTS2 and SVAWS restrict questions to experiences within intimate relationships. One common characteristic is that each measure avoids legal terms, such as *rape* and *sexual assault*, that are poorly understood by the public. Instead, each employs operational definitions that vary along three critical dimensions: (a) type of sexual behavior and definitions, (b) strategy, and (c) specification of nonconsent (Muehlenhard, Powch, Phelps, & Giusti, 1992).

Type of Sexual Behavior

These instruments vary according to the range and definition of sexual behaviors. Only the SES includes attempted and completed behaviors; the CTS2 and the SVAWS ask only about completed acts. Moreover, the SES defines sexual contact as "fondling, kissing, or petting, but not intercourse" and sexual intercourse as "penetration of a woman's vagina, no matter how slight, by a man's penis. Ejaculation is not required" (Koss et al., 1987, p. 167). In contrast, the CTS2 and the SVAWS use the terms *sex*, *anal sex*, *oral sex*, and *sexual intercourse* without defining them. Open to interpretation, these terms may be misunderstood by those more familiar with slang expressions for the same behavior.

Strategy

Definitions of strategies used to obtain nonconsensual sexual activity vary in number and specificity. The SES defines four: verbal coercion ("overwhelmed her with continual arguments and pressure"), misuse of authority ("because you were her boss, teacher, camp counselor, supervisor"), intoxication ("giving her alcohol or drugs"), and threat or force ("threatening or twisting her arm, holding her down") (Koss et al., 1987, p. 167). The CTS2 includes three: verbal coercion ("insisted that my partner"), threats ("used threats to make your partner"), and force ("used force [like hitting, holding down, or using a weapon]"). The CTS2 approach is similar to the SES, but subtle important distinctions exist. For example, the nature of threat in the CTS2 may refer to threats of physical harm or other consequences, whereas in the SES, threats refer to the use of force. This distinction is important in legal contexts. Of the three instruments, the SVAWS uses the most ambiguous language with undefined terms such as

demanded, *made her*, and *physically forced her*. In addition, neither the CTS2 nor the SVAWS include questions about nonconsensual sex obtained when the victim could not consent or through the misuse of authority.

Consent

Specifying nonconsent is critical to defining sexual assault. *Nonconsent* is consistently defined in SES items as "when she didn't want to," whereas in the CTS2, it is variably defined. Some CTS2 items employ the phrase, "when my partner did not want to," but others use "made my partner" or "insisted that my partner." Some SVAWS items use the expression, "against her will." Others imply nonconsent through the use of force, but some do not denote lack of consent (e.g., "used an object on her in a sexual way"), and one is ambiguous ("demanded she have sex whether she wanted to or not"). Some of these phrases are dependent on beliefs about sexual scripts that define "normal" sexual behavior (Porter & Critelli, 1992).

PHYSICAL AND NONPHYSICAL AGGRESSION

The CTS2 and the SVAWS physical and nonphysical subscales mainly differ by number of items and subscales. The SVAWS subscales contain 45% and 50% more items than the CTS physical and nonphysical subscales, respectively. A second distinguishing characteristic is how each instrument measures severity, a dimension of physical violence that predicts criminal justice and mental health outcomes (e.g., seeking protective orders [Fernandez, Iwamoto, & Muscat, 1997], mental health outcomes [Goodman, Dutton, & Harris, 1997]). The CTS2 structures assault severity according to minor and severe levels, whereas the SVAWS specifies four levels for physical (mild, minor, moderate, and serious violence) and for nonphysical (symbolic, mild, moderate, and serious) subscales. However, many items are common to both instruments. The implicit assumption is that more items make a better instrument (i.e., more sensitive and accurate), but whether many items identify more perpetrators than few, and whether discriminant or predictive power is gained from including more than two gradations of severity, are unknown.

Another important difference between nonphysical aggression definitions is that SVAWS subscales include only behaviors that could potentially result in physical harm (Marshall, 1992). CTS2 items are not likely to result in physical harm. In fact, only one describes a threat of physical violence. Thus, this study compares distinct ways to measure nonphysical aggression.

SOCIAL DESIRABILITY AS A THREAT TO VALIDITY

Other methodological factors influence disclosure. Social desirability poses threats to self-report methods in general and is particularly problematic to sensitive questions (Catania, McDermott, & Pollack, 1986). Because the experiences of victimization and perpetration stand in stark contrast, social desirability may operate differently in each case. A meta-analysis of the effect of social desirability on self-reports of physical aggression (mostly based on the CTS) within intimate relationships illustrated that self-reports of perpetration were more vulnerable than reports of victimization (Sugarman & Hotaling, 1997). Yet the average effect size was small ($r = –.22$). Although effect sizes were generally not computed for different levels of severity, the review noted one robust effect when severity was the metric instead of frequency or presence/absence of violence (Saunders, 1991).

There is no comparable meta-analysis on the effect of social desirability on reports of sexual aggression perpetration, but data from samples of college and community men showed that responses to the SES were independent of social desirability (Walker, Rowe, & Quinsey, 1993). Walker and colleagues (1993) theorized that SES items are worded in ways that reduce the emotional consequences men encounter when admitting sexual coercion or rape. Plus, some may argue that the questions reduce desirability because they accurately reflect scripts that describe acceptable dating behavior (Donat & White, 1999; Krahe, 2000). Together, the meta-analysis and Walker and colleagues' findings raise the questions of whether some instruments are more vulnerable to social desirability than others and whether social desirability poses a greater threat to self-reports of violence against women than general violence.

A second important question is whether methodological factors heighten or suppress social desirability. Most victimization

researchers presume that matching sex and ethnicity of interviewers and research participants creates the best climate for disclosure (Koss, 1992, 1993). Factors that create a good climate for disclosing perpetration have not been identified. When planning this study, a number of correctional officers as well as researchers were confident that inmates would disclose significantly fewer reports of perpetration to men than women mainly because prison culture encourages men to be stoic around other men but allows them to reveal intimate experiences to women. Therefore, in a prison context, when inmates report to men, social desirability may decrease reports. In contrast, when they report to women, social desirability may increase reports. In the absence of data on these assumptions, the use of male and female interviewers in this study allowed an exploration of whether measurement conditions, defined by match between interviewer and inmate sex, moderated the relation between social desirability and disclosure.

To summarize, this study examines self-reports of sexual, physical, and nonphysical aggression obtained with three instruments. The following findings are expected. First, because each instrument specifies sexual behavior, strategy, and consent to different degrees, self-reports of sexual aggression will vary across instruments. In particular, because of its specificity, more men will self-report on the SES than on the CTS2 or the SVAWS. In contrast, even though the CTS2 and SVAWS vary in number of items and subscales, equal numbers of men will report physical aggression because each instrument's content is similar. Definitions of nonphysical aggression differ, however, and unequal numbers of men are expected to report on the CTS2 and SVAWS.

Second, social desirability will be negatively related to self-reports on the most severe subscales of physical aggression. Although reports on SES subscales are not expected to relate to social desirability, reports on CTS2 and SVAWS sexual aggression subscales are. Given the high base rate of nonphysical aggression, no relation is expected with social desirability. Where main effects of social desirability exist, measurement conditions may moderate. Inmates high on social desirability and interviewed by men may report fewer aggressive behaviors than inmates low on social desirability and interviewed by men and inmates, regardless of social desirability, who are interviewed by women.

METHOD

SAMPLE AND POWER

Inmates were randomly selected from two maximum-security Commonwealth of Virginia Department of Corrections (DOC) facilities, Dillwyn (DWCC) and Buckingham Correctional Centers (BUCC) between September 1996 and April 1997. At data collection, 48% of inmates were incarcerated for violent crimes; 87% were younger than 40; 67% were African American and 32% were Caucasian (DOC, 1996). Facilities were selected because of institutional support for research and proximity to the investigator's university.

Using a random number table and the prison census, 347 (177 African American, 170 Caucasian) inmates were randomly selected and invited to participate in individual, 1-hour interviews. One hundred and sixty men composed the final sample. Caucasian inmates were oversampled to balance ethnicity (89 African American, 71 Caucasian). Only 20% were incarcerated for sex offenses. Regardless of sexual orientation (98% reported heterosexual), all inmates reported previous relationships with women. The sample size afforded power above .80 to detect medium to large effect sizes for $\alpha = .05$ in the planned analyses.

Seventy-seven inmates were unable to consent because of (a) transfers to another institution ($n = 46$), (b) placement in segregated housing for disruptive behavior ($n = 16$), (c) parole ($n = 7$), (d) previously scheduled dental or medical appointments ($n = 4$), (e) language barriers ($n = 2$), or (f) active psychosis ($n = 1$). These 76 inmates were not included in the calculation of participation rate. Inmates who did not come to their appointment or who did come but subsequently decided not to participate were classified as decliners ($n = 110$). Thus, the participation rate was 59% (160/271). Inmates who participated were compliant and generally completed the entire interview. Interviewers or inmates terminated no interviews. As a result, missing data were rare.[1]

Chi-square analyses indicate nonsignificant differences in participation by ethnicity, $\chi^2(1, n = 347) = 2.53$, $p = .11$; custody status, a classification system based on institutional violence, disciplinary, criminal, and escape histories, parole eligibility, and detainer status, $\chi^2(2, n = 347) = 0.31$, $p = .85$; or current offense, $\chi^2(5, n =$

347) = 9.00, $p = .11$. Male researchers interviewed 45% of the sample. Inmates were more likely to participate if their invitational letter stated that they would be interviewed by a woman rather than a man (56% vs. 34% participation rate, respectively), $\chi^2(1, n = 347) = 10.73$, $p < .01$.

PROCEDURE

In conjunction with institutional, regional, and state DOC administrators, the author developed the interview procedure. The research team consisted of the author and six advanced undergraduate research assistants.[2] Individual interviews were conducted confidentially in administrative offices away from dormitories and cellblocks over a 7-month period. Inmates received invitational letters via institutional mail on university stationery 1 to 2 days prior to interviews. Inmates reported unaccompanied to the study location except for 2 weeks when BUCC was partially locked down. During this time, correctional officers escorted inmates. Informed consent procedures stressed that the DOC was not privy to individual inmates' responses, that inmates' responses or participation could not be used in disciplinary or parole hearings, and that a National Institute of Mental Health Certificate of Confidentiality protected study data. Interviews generally lasted 1 hour. To debrief, interviewers told inmates they had participated in a study about men's experiences in relationships, gave them the opportunity to request a study summary, and thanked them for their cooperation. Because inmates who participated returned to the general population, there was no way of preventing them from discussing the study with fellow inmates. The population, however, turned over quickly due to transfers and parole. Few participants reported they had heard details of the study.

INSTRUMENTS AND VARIABLES

In order, the interview protocol included (a) basic demographic questions about education, employment, and relationships with women; (b) the Adult Self-Perception Profile (ASPP) (Messer & Harter, 1986); (c) the CTS2, the SES or the SESR, and the SVAWS; (d) the Aggression Questionnaire (Buss & Perry, 1992); and (e) the

Social Desirability Scale (Crowne & Marlowe, 1960). Participants responded to all instruments, and the order of the CTS2, SES, and SVAWS was counterbalanced. However, approximately one third of the sample received the original SES ($n = 65$) and the remaining two thirds received a revised version of the SES ($n = 95$). Data from the revised SES, the CTS2 injury and negotiation subscales, reports of victimization, and ASPP are not a focus of this study. Thus, the sample size for the CTS2 and the SVAWS is 160; for the SES it is 65.

Self-Reports of Aggressive Behaviors

Self-reports of sexual, physical, and psychological aggression were obtained with the SES, CTS2, and SVAWS as described earlier. Responses to the SES, CTS2, and SVAWS were coded 1 = yes and 0 = no. If an inmate answered yes to any item on a subscale, he received a score of 1 for that subscale. If he answered no to all items, he received a 0. Means calculated from these binary variables equal the proportion of inmates who reported perpetrating each type of aggression. Instructions preceding the SES items asked inmates to answer questions with respect to experiences with any women since age 14. When responding to the CTS2 and SVAWS, interviewers asked inmates to consider experiences with female intimate partners since age 14. The SVAWS item, "Demanded sex whether she wanted to or not," was omitted because it elicited many questions from inmates.

Estimates of internal consistency for the SES are available only for college samples ($\alpha = .89$ in Koss & Gidycz, 1985) although it has been used successfully in community samples (Walker et al., 1993). No published data on the SVAWS are available for comparison purposes. Half of the sexual aggression subscales demonstrated moderate to acceptable reliability assessed with the Kuder-Richardson 20 (K-R 20 = .77 to .89), but estimates for SES contact, SES coercion, CTS2 minor sexual violence, and SVAWS sexual violence were low (K-R 20 < .70) (see Table 1). Physical aggression subscales demonstrated acceptable internal consistency (K-R 20 ≥ .70), except SVAWS minor and serious physical violence (see Table 1). Estimates for CTS2 minor and severe psychological abuse and the SVAWS symbolic subscales were low, but the remaining nonphysical subscales were above .70 (see Table 1).

TABLE 1
Relations Among Self-Reports Obtained From Three Violence Against Women Instruments

			Phi Coefficients				
	n	*K-R 20*	*1*	*2*	*3*	*4*	*5*
1a. Sexual aggression perpetration							
1. SES sexual contact	65	.54					
2. SES sexual coercion	65	.32	*.52*				
3. SES rape	65	.77	*.34*	*.55*			
4. CTS2 minor sexual coercion	158	.50	.07	–.20	.03		
5. CTS2 severe sexual coercion	158	.80	.05	–.09	.07	*.23*	
6. SVAWS sexual violence subscale	155	.66	.14	–.14	.16	*.37*	*.47*
1b. Physical aggression perpetration							
1. CTS2 minor physical assault	160	.71					
2. CTS2 severe physical assault	158	.74	*.41*				
3. SVAWS mild physical violence	158	.79	*.74*	*.41*			
4. SVAWS minor physical violence	159	.40	*.37*	*.64*	*.34*		
5. SVAWS moderate physical violence	159	.75	*.46*	*.49*	*.38*	*.47*	
6. SVAWS serious physical violence	157	.57	*.35*	*.72*	*.37*	*.57*	*.50*
1c. Nonphysical aggression perpetration							
1. CTS2 minor psychological aggression	159	.66					
2. CTS2 severe psychological aggression	159	.50	.16				
3. SVAWS symbolic	158	.56	*.42*	*.29*			
4. SVAWS mild threats of violence	157	.70	*.33*	*.42*	*.25*		
5. SVAWS moderate threats of violence	158	.70	.14	*.53*	*.24*	*.38*	
6. SVAWS serious threats of violence	155	.77	.13	*.35*	*.31*	*.31*	*.36*

NOTE: K-R 20 = Kuder-Richardson 20; SES = Sexual Experiences Survey; CTS2 = Conflict Tactics Scale–Revised; SVAWS = Severity of Violence Against Women Scales. Significant correlations ($p < .05$) are in italics.

General Aggression

The Aggression Questionnaire (Buss & Perry, 1992) was used to measure general aggression. Twenty-nine items make up four subscales of physical aggression (e.g., "If someone hits me, I hit back"), verbal aggression (e.g., "I often find myself disagreeing with people"), anger (e.g., "I flare up quickly but get over it quickly"), and hostility (e.g., "I am sometimes eaten up with jealousy"). Participants responded on a Likert-type scale from 1 = *extremely uncharacteristic of me* to 5 = *extremely characteristic of me*. Internal consistency estimated by Cronbach's alpha for the total scale and subscales in samples of male and female college students range from .72 to .89 (Buss & Perry, 1992). Peer nomination data on college men contribute to a strong foundation for construct validity (Buss & Perry, 1992). Internal consistency of the

total Aggression Questionnaire score in the present sample is high (α = .89).

Social Desirability

The Marlowe-Crowne Social Desirability Scale contains 33 items that ask about everyday desirable but rare events (Crowne & Marlow, 1960) and has been used extensively over its 40-year existence. Participants responded on a Likert-type scale from 1 = *extremely unlike me* to 5 = *extremely like me*. Four items were omitted because incarcerated men do not have the opportunity to engage in the events (e.g., the item, "I never make a long trip without checking the safety of my car,") and 3 others because they impaired internal consistency. Cronbach's alpha for the resulting scale was .78. Mean social desirability scores computed with the remaining 26 items were normally distributed with a mean of 3.2 and a range of 2.20 to 4.57. Low scores indicate less social desirability.

PLAN OF ANALYSIS

To test whether the SES, CTS2, and SVAWS assess equivalent numbers of sexual aggression reports, 95% confidence intervals (CIs) were computed for subscale means. For binary data, the mean is equal to the proportion of men who responded yes (yes = 1, no = 0). This approach demonstrates the lower and upper bounds in which the true population mean has a 95% chance of falling. For example, if the mean of subscale A falls within the CIs of subscale B's mean, the two means do not differ. Tests of the relation between social desirability and self-reports were assessed with correlation analyses.

RESULTS

COMPARING SELF-REPORTS ACROSS INSTRUMENTS

Self-Reports of Sexual Aggression

Means and 95% CIs for sexual aggression subscales are presented in Figure 1a. Three comparisons are particularly salient. First, SES coercion and CTS2 minor sexual coercion subscales

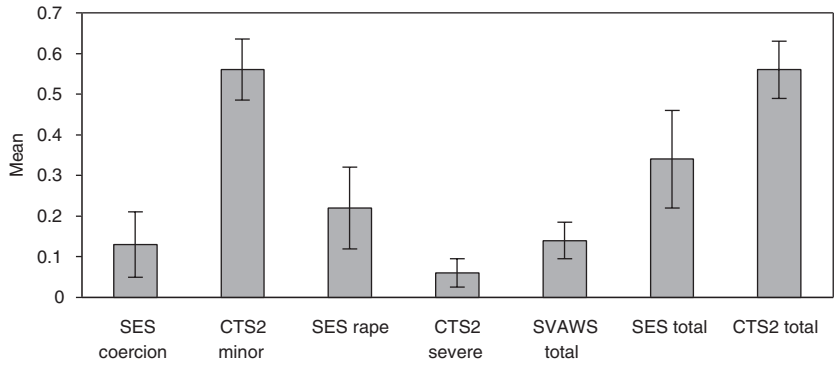

Figure 1a: Proportion of Men Who Reported Sexual Aggression Against Women on Three Measurement Instruments

NOTE: Mean is equal to the proportion of men who reported yes to any subscale item (95% confidence intervals indicated). $n = 65$ for Sexual Experiences Survey (SES) subscales; $n = 158$ for Conflict Tactics Scale–Revised (CTS2) and $n = 155$ for Severity of Violence Against Women Scales (SVAWS).

inquire about nonconsensual sex obtained without force, but the number of men who reported these behaviors on each subscale was significantly different. Based on point estimates, the number of men who reported minor sexual coercion obtained with three broad CTS2 items was 4 times higher than the number obtained with two relatively specific SES coercion items. The difference remains when the CTS2 item "refused to wear a condom" is removed. Second, the number of men who reported rape on the SES is significantly higher than the number who reported yes to the CTS2 severe sexual coercion subscale. The SES rape subscale elicited approximately 3.5 times more reports. Third, the SVAWS subscale can only be compared to the number detected by total SES and CTS2 scale scores. When these broad definitions are used, each instrument detects significantly different numbers of men (see Figure 1a).

Phi coefficients among these subscales assess the degree to which subscales elicit reports from the same men. Reports on the SES subscales (contact, coercion, and rape, respectively) were positively but minimally related to CTS2 sexual coercion and SVAWS sexual violence subscales (see Table 2). However, self-reports of CTS2 minor and CTS2 severe sexual coercion were positively and moderately related to self-reports of SVAWS sexual violence.

TABLE 2
Relation Between Self-Reports Obtained From Three Violence Against Women Instruments and Social Desirability

Self-Report on Violence Against Women Instrument Subscales	*Point Biserial Correlation With Social Desirability*
SES sexual contact	–.20
SES sexual coercion	–.14
SES rape	–.14
CTS2 minor sexual coercion	–.04
CTS2 severe sexual coercion	.05
SVAWS sexual violence subscale	.05
CTS2 minor physical assault	.07
CTS2 severe physical assault	.08
SVAWS mild physical violence	.05
SVAWS minor physical violence	.12
SVAWS moderate physical violence	.10
SVAWS serious physical violence	.12
CTS2 minor psychological aggression	–.10
CTS2 severe psychological aggression	–.01
SVAWS symbolic	–.05
SVAWS mild threats of violence	–.09
SVAWS moderate threats of violence	–.02
SVAWS serious threats of violence	*.17*

NOTE: $n = 64$ for SES subscales; $n = 142$ for all other subscales. SES = Sexual Experiences Survey; CTS2 = Conflict Tactics Scale–Revised; SVAWS = Severity of Violence Against Women Scales. Significant correlations ($p < .05$) are in italics.

Self-Reports of Physical Aggression

Means and 95% CIs for physical aggression subscales are presented in Figure 1b. The number of men who reported minor physical violence (CTS2) was significantly different from those who reported mild physical aggression (SVAWS). In contrast, the number of men who reported minor physical assault (CTS2) and minor physical violence (SVAWS) was almost identical. Likewise, means for CTS2 severe, SVAWS moderate, and SVAWS serious violence fell within CIs for each subscale. Self-reports of physical aggression on the total CTS2 and SVAWS were similar.

CTS2 minor physical assault reports correlated only moderately with SVAWS minor physical violence reports ($r_b = .37$), which were strongly related to CTS2 severe physical assault reports ($r_b = .64$). Reports of CTS2 severe physical violence and SVAWS serious physical violence reports were also strongly related ($r_b = .72$).

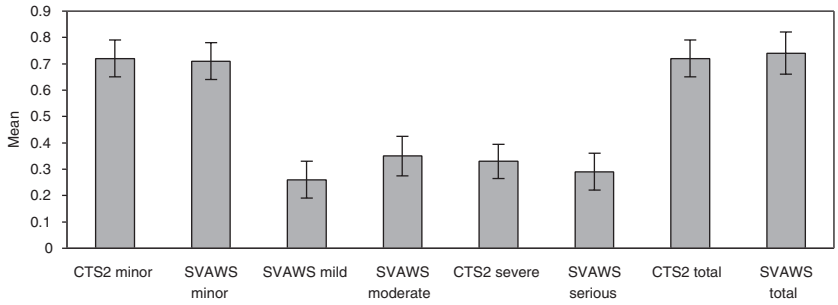

Figure 1b: Proportion of Men Who Reported Physical Aggression Against Women on Three Measurement Instruments

NOTE: Mean is equal to the proportion of men who reported yes to any subscale item (95% confidence intervals indicated). *n*s ranged from 157 to 160 for Conflict Tactics Scale–Revised (CTS2) and Severity of Violence Against Women Scales (SVAWS).

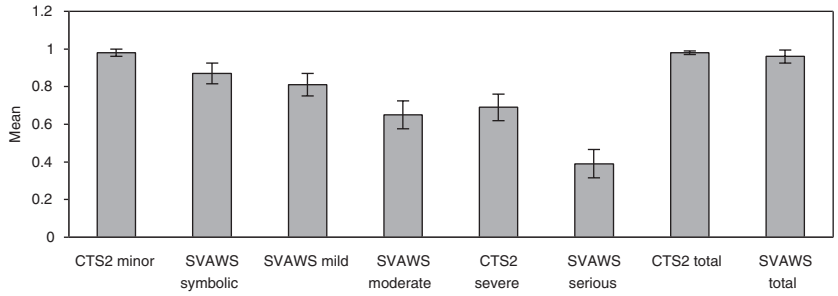

Figure 1c: Proportion of Men Who Reported Nonphysical Abuse Against Women on Three Measurement Instruments

NOTE: Mean is equal to the proportion of men who reported yes to any subscale item (95% confidence intervals indicated). *n*s ranged from 155 to 159 for Conflict Tactics Scale–Revised (CTS2) and Severity of Violence Against Women Scales (SVAWS).

Self-Reports of Nonphysical Aggression

A slightly different pattern emerged for nonphysical aggression subscales (see Figure 1c for means and 95% CIs). The number of men who reported minor psychological aggression (CTS2) was significantly different from the number of men who reported symbolic violence (SVAWS) and mild threats (SVAWS). CIs for the CTS2 severe subscale overlapped with the SVAWS moderate subscale but not with SVAWS serious threats of violence.

Reports of CTS2 minor psychological aggression and of SVAWS symbolic and mild threats of violence were moderately related. Reports of CTS2 severe psychological aggression were

most highly related to SVAWS moderate threats (r_b = .53) (see Figure 1c).

THE EFFECT OF SOCIAL DESIRABILITY ON SELF-REPORTS OF AGGRESSION AGAINST WOMEN

Point biserial correlations between social desirability and self-reports of sexual, physical, and nonphysical aggression were weak across all 18 instrument subscales and nonsignificant except for SVAWS serious threats (r = .17) (see Table 2). Because this finding could have occurred by chance, it was not pursued further.

Conversely, the Pearson coefficient for social desirability and general aggression was moderate (r = –.23). To explore whether measurement condition moderated this relation, a hierarchical regression with social desirability, sex of interviewer, and a centered interaction term was conducted. With the constant entered, the main effect for social desirability was significant (β = –.23, p < .05), but the main effect for sex of interviewer was not (β = –.14, p < .09). This model accounted for 26% of the variance in social desirability. Adding the interaction term (Social Desirability × Sex of Interviewer) did not significantly increase the explanatory power of the model (ΔR^2 = .07), and the regression coefficient was not significant (β = .53, p < .43).

DISCUSSION

This study highlights how measurement choice influences study findings. It is the first to compare self-reports of perpetration of violence against women obtained with three instruments in the same sample. Findings illustrate that varying approaches to measuring sexual, physical, and nonphysical aggression influence disclosure and identify different groups of men as perpetrators. It also demonstrates that social desirability does not appear to pose a significant threat to self-reports of violence against women.

Three instruments compared in this study obtained different estimates of men who had perpetrated sexual aggression, whether self-reports were measured by subscale or a total scale score. Because the SES, CTS2, and SVAWS differ on many dimensions, this study could not determine conclusively which aspects

of the instruments' operational definitions influenced disclosure. However, the data do suggest that the relationship between definition and disclosure is not simple. Somewhat contrary to expectations, when total scale scores were used, the SES obtained more than twice the number of men who disclosed than the SVAWS but a third less than the CTS2.

Self-reports on subscales also showed variation across instruments. Three broad CTS2 minor sexual coercion items identified more than 4 times the number identified by two detailed SES sexual coercion items. This finding suggests that broad items elicit high rates of disclosure. Yet five SES rape items highly specific in terms of consent, behavior, and strategy detected 3.5 times the number of men who reported rape on three less specific CTS2 severe sexual coercion items. These differences are larger than expected if disclosure of sexual aggression varied only by the number of items. Broad, nonspecific items may lead individuals to include or exclude relevant experiences, depending on the severity of the behavior questioned. Future research that blends qualitative and quantitative self-report methods and cognitive psychological theories may shed light on the process by which study participants consider their experiences when responding to broad or specific questions.

In addition to differences in number of men identified, patterns of relations between instrument subscales show that the subscales detect different groups of men as sexually aggressive. The CTS2 and SVAWS subscales identify some, but not the majority, of the same men. Minimal correlations indicate that there is almost no overlap between groups of men detected by the SES and the other two measures. Differing operational definitions undoubtedly play a critical role in the lack of overlap, but low internal consistency estimates for SES subscales should not be discounted as a contributing factor.

Discrepancies in the measurement of physical and nonphysical aggression are not as pronounced. With the exception of differences in disclosure on the CTS minor and SVAWS mild physical aggression subscales, the SVAWS and the CTS2 elicited the same number of men who reported across remaining subscales and total scales. The lack of difference between self-reports on the 12-item CTS2 and the 21-item SVAWS total scale scores demonstrates that a small number of items does not necessarily lead to

underidentifying physically aggressive men. Groups of men who reported minor and serious or severe physical aggression (CTS2 minor and SVAWS minor, CTS2 severe and SVAWS serious) in intimate relationships overlapped considerably, but this is not surprising given their similar content.

Similarly, the CTS2 psychological aggression subscales contain about half as many items as the SVAWS threat subscales, and when total scores are used, these instruments detected equivalent numbers of men who disclosed nonphysical aggression. However, disclosure varied across subscales, notably between those that measure the same level of severity. Relations between these instrument subscales also suggest variability in groups identified. Recall that the CTS2 defines psychological aggression generally and that the SVAWS measures only threats that potentially lead to physical harm.

Practical implications of these findings are easy to enumerate. Comparable disclosure rates and high relations between physical aggression subscales may guide a researcher's choice of instruments when the goal is to detect the presence or absence of physical aggression, or when resources, such as time or literacy, are limited. Choices about the measurement of sexual and nonphysical aggression are more complex. This study shows that the use of the SES, CTS2, and SVAWS sexual aggression subscales likely lead to contradictory incidence and prevalence estimates, identify discrepant comparison groups within studies, and produce different samples when used to select or screen out potential study participants. Whether the CTS2 or the SVAWS is used to assess nonphysical or physical aggression may not have an impact on incidence or prevalence estimates, but to the degree that these subscales are not highly related, they may assess different aspects of nonphysical aggression, and instrument selection may influence relations between nonphysical aggression and other key variables under study.

Underlying these practical implications are theoretical questions that can be empirically evaluated. For example, which dimensions of sexually aggressive behavior need to be included for valid assessment? How do these dimensions influence critical relations in causal models? Is strategy or sexual behavior the more influential component? Future research should elucidate the effects of including and excluding these and other dimensions of

sexual aggression in causal models of sexual aggression. A first step may be to expand the SES by including questions about all sexual behaviors obtained through each strategy and testing whether causal factors such as attitudes, hostility, or personality variables differentially predict strategy or behavior. Similarly, many questions remain about the nature of nonphysical violence, although considerable work has already advanced understanding (see O'Leary, 1999).

In the face of such complexity, it is reassuring that in contrast to reports of general aggression, social desirability does not appear to pose a threat to self-reports of violence against women. The effects between social desirability and self-reports of sexual, physical, and nonphysical aggression were generally weak and nonsignificant. In addition, social desirability and interviewer sex did not interact to create distinct measurement conditions that influenced willingness to disclose. This finding is similar to the meta-analysis by Sugarman and Hotaling (1997). This study, however, extends their findings because effects for social desirability were examined across levels of severity. One reason for the lack of effects could be that self-report was coded as a binary variable and thus restricted in range; however, exploratory analysis with frequency of aggression as the dependent variable (and thus a greater range) produced the same results. Contrary to correctional staff predictions, inmates made similar reports of aggression against women and general aggression to female and male interviewers.

This study has several limitations related to conceptualization and design. First, it is important to point out that SES, CTS2 and SVAWS authors have not claimed that these instruments measure sexual, physical, and nonphysical aggression equivalently. Yet violence against women research is conducted within social science research, a discipline that depends on the accumulation of comparable evidence to develop knowledge. Thus, it is reasonable to compare these measures to achieve some understanding of how they may contribute to convergent or divergent findings.

This study's sample does limit generalizability and conclusions. When planning this study, limited fiscal and human resources necessitated the selection of a sample that was likely to have experienced the behavior or related risk factors frequently. An incarcerated sample fits this criterion, but findings cannot be extended to other populations (Dutton & Hart, 1992). The same

questions, however, are important to answer in other samples of men.

Despite these limitations, data on incarcerated men's perpetration of sexual, physical, and psychological aggression are rare. Few studies of incarcerated men, not selected for criminal history, are available. Outside of criminal justice and correctional research, this ever-increasing and changing population is understudied. In light of the high number of men who reported sexual, physical, and nonphysical aggression against a woman, yet who were not detected as sexual or domestic violence offenders by the criminal justice system, it is defensible to assert that incarcerated men are a group at risk for further aggression when released from prison. They should be a group of interest to researchers studying violence against women.

CONCLUSION AND FUTURE DIRECTIONS

This study supports calls to improve the state of measurement in violence against women research (Crowell & Burgess, 1996). Although the field has made appreciable gains in instrumentation, additional research specifically focused on establishing the internal and external validity of all forms of violence against women, and particularly sexual aggression, is needed. A simple but powerful strategy for advancing measurement development and validation is the consistent use of multiple measures within studies to allow direct comparisons.

NOTES

1. Missing data, particularly on the social desirability scale, were generally due to institutional constraints that necessitated some inmates' returning to dormitories or cellblocks for counts before completion of the interview.

2. Undergraduates completed tours of the institutions and security clearances, participated in institutional orientation, and practiced interview skills extensively over a 3-month period, culminating in role-play interviews with inmates at each institution.

REFERENCES

Bachman, R. (2000). A comparison of annual incidence rates and contextual characteristics of intimate-partner violence against women from the National Crime Victimization Survey (NCVS) and the National Violence Against Women Survey (NVAWS). *Violence Against Women*, *6*, 839-867.

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.

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, *63*, 452-459.

Catania, J. A., McDermott, L. J., & Pollack, L. M. (1986). Questionnaire response bias and face-to-face interview sample bias in sexuality research. *Journal of Sex Research*, *22*, 52-72.

Commonwealth of Virginia Department of Corrections. (1996). *Inmate statistical summary*. Richmond, VA: Author.

Crowell, N. A. & Burgess, A. W. (1996). *Understanding violence against women*. Washington, DC: National Academy of Sciences.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting and Clinical Psychology*, *24*, 349-354.

DeKeseredy, W., & Schwartz, M. (1998, February 1). *Measuring the extent of woman abuse in intimate heterosexual relationships: A critique of the Conflict Tactics Scales*. Available from http://www.vawnet.org

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

Donat, P.L.N., & White, J. (1999). Re-examining the issue of nonconsent in acquaintance rape. In C. B. Travis & J. W. White (Eds.), *Sexuality, society, and feminism* (pp. 355-376). Washington, DC: American Psychological Association.

Dutton, D. G., & Hart, S. D. (1992). Risk markers for family violence in a federally incarcerated population. *International Journal of Law and Psychiatry*, *15*, 101-112.

Fernandez, M., Iwamoto, K., & Muscat, B. (1997). Dependence and severity of abuse: Impact on women's persistence in utilizing the court system as protection against domestic violence. *Women and Criminal Justice*, *9*, 63.

Gelles, R. J., & Straus, M. S. (1979). Violence in the American family. *Journal of Social Issues*, *35*, 15-39.

Gilbert, N. (1997). Advocacy research exaggerates rape statistics. In M. R. Walsh (Ed.), *Women, men and gender: Ongoing debates* (pp. 236-242). New Haven, CT: Yale University Press.

Goodman, L. A., Dutton, M. A., & Harris, M. (1997). The relationship between violence dimensions and symptom severity among homeless, mentally ill women. *Journal of Traumatic Stress Studies*, *10*, 51-70.

Hilton, N. Z., Harris, G. T., & Rice, M. E. (1998). On the validity of self-reported rates of interpersonal violence. *Journal of Interpersonal Violence*, *13*, 58-72.

Hudson, W. W., & McIntosh, S. R. (1981). The assessment of spouse abuse: Two quantifiable dimensions. *Journal of Marriage and the Family*, *43*, 873-885.

Kanin, E. J. (1957). Male aggression in dating-courtship relations. *American Journal of Sociology*, *63*, 197-204.

Koss, M. P. (1992). The underdetection of rape: Methodological choices that influence the magnitude of incidence estimates. *Journal of Social Issues*, *48*, 61-75.

Koss, M. P. (1993). Detecting the scope of rape: A review of prevalence research methods. *Journal of Interpersonal Violence*, *8*, 198-222.

Koss, M. P., & Cook, S. L. (in press). More data have accumulated supporting date and acquaintance rape as significant problems for women. In R. J. Gelles & D. R. Loseke (Eds.), *Current controversies on family violence* (2nd ed.).Thousand Oaks, CA: Sage.

Koss, M. P., & Gidycz, C. A. (1985). Sexual Experiences Survey: Reliability and validity. *Journal of Consulting and Clinical Psychology*, *53*, 422-423.

Koss, M. P., Gidycz, C. A., & Wisniewski, N. (1987). The scope of rape: Incidence and prevalence of sexual aggression and victimization in a national sample of higher education students. *Journal of Consulting and Clinical Psychology*, *55*, 162-170.

Koss, M. P., & Oros, C. (1982). The Sexual Experiences Survey: A research instrument investigating sexual aggression and victimization. *Journal of Consulting and Clinical Psychology*, *50*, 455-457.

Krahe, B. (2000). Sexual scripts and heterosexual aggression. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 273-292). Mahwah, NJ: Lawrence Erlbaum.

Lisak, D., & Roth, S. (1988). Motivational factors in nonincarcerated sexually aggressive men. *Journal of Personality and Social Psychology*, *55*, 795-802.

Malamuth, N. M. (1986). Predictors of naturalistic aggression. *Journal of Personality and Social Psychology*, *50*, 953-962.

Malamuth, N. M. (1989a). The Attraction to Sexual Aggression Scale: Part one. *Journal of Sex Research*, *26*, 26-49.

Malamuth, N. M. (1989b). The Attraction to Sexual Aggression Scale: Part two. *Journal of Sex Research*, *26*, 324-354.

Marshall, L. L. (1992). Development of the severity of violence against women scales. *Journal of Family Violence*, *7*, 103-121.

Merrill, L. L., Newell, C. E., Milner, J. S., Koss, M. P., Hervig, L. K., Gold, S. R., et al. (1998). Prevalence of premilitary adult sexual victimization and aggression in a navy recruit sample. *Military Medicine*, *163*, 209-212.

Messer, B., & Harter, S. (1986). *Manual for the Adult Self-Perception Profile*. Denver, CO: University of Denver.

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

Muehlenhard, C. L., Powch, I. G., Phelps, J. L., & Giusti, L. M. (1992). Definitions of rape: Scientific and political implications. *Journal of Social Issues*, *48*, 23-44.

O'Leary, K. D. (1999). Psychological abuse: A variable deserving attention in domestic violence. *Violence and Victims*, *14*, 3-23.

O'Leary, K. D., Barling, J., Arias, I., Rosenbaum, A., Malone, J., & Tyree, A. (1989). Prevalence and stability of aggression between spouses. *Journal of Consulting and Clinical Psychology*, *57*, 263-268.

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.

Porter, J. F., & Critelli, J. W. (1992). Measurement of sexual aggression in college men: A methodological analysis. *Archives of Sexual Behavior*, *21*, 525-542.

Rapaport, K., & Burkhart, B. R. (1984). Personality and attitudinal characteristics of sexually coercive college males. *Journal of Abnormal Psychology*, *93*, 216-221.

Saunders, D. G. (1991). Adjusting for social desirability bias. *Journal of Interpersonal Violence*, *6*, 336-344.

Schwartz, M. D. (2000). Methodological issues in the use of survey data for measuring and characterizing violence against women. *Violence Against Women*, *6*, 815-838.

Scott, K. D., & Aneshensel, C. S. (1997). An examination of the reliability of sexual assault reports. *Journal of Interpersonal Violence*, *12*, 361-374.

Spitzberg, B. H. (1999). An analysis of empirical estimates of sexual aggression victimization and perpetration. *Violence and Victims*, *14*, 241-260.
Straus, M. A. (1995). 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 Publishing. (Original work published 1990)
Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The Revised Conflict Tactics Scales: Development and preliminary psychometric data. *Journal of Family Issues*, *17*, 283-316.
Sugarman, D. B., & Hotaling, G. T. (1997). Intimate violence and social desirability: A meta-analytic review. *Journal of Interpersonal Violence*, *12*, 275-290.
Tolman, R. M. (1989). The development of a measure of psychological maltreatment of women by their male partners. *Violence and Victims*, *4*, 159-177.
Walker, W. D., Rowe, R. C., & Quinsey, V. L. (1993). Authoritarianism and sexual aggression. *Journal of Personality and Social Psychology*, *65*, 1036-1045.
Weaver, T. L. (1998). Method variance and sensitivity of screening for traumatic stressors. *Journal of Traumatic Stress*, *11*, 181-185.

Sarah L. Cook, Ph.D., (University of Virginia, 1998) is an assistant professor of community psychology in the Department of Psychology at Georgia State University. Her research interests include measurement and methodological challenges in violence research and the interface between public policy and psychological research on violence against women. She is currently involved in developing a method to assess women's appraisals of abuse experiences through the Women's Life Experiences Project funded by the National Institute of Justice. Before beginning an academic career, she addressed the problem of violence against women as an educator, advocate, and social worker.