

Emerging Issues in the Measurement of Rape Victimization

Violence Against Women
17(2) 201–218

© The Author(s) 2011

Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>

DOI: 10.1177/1077801210397741

<http://vaw.sagepub.com>



**Sarah L. Cook¹, Christine A. Gidycz²,
Mary P. Koss³, and Megan Murphy²**

Abstract

We provide an overview of emerging directions in the measurement of rape, the most extreme form of sexual victimization. The context for our overview is how operational definitions of rape have evolved, where consensus has emerged, and where it eludes the field. We discuss two approaches to the detection of rape victimization in survey methods, namely behaviorally specific questions and a new, two-stage approach, and how each can be evaluated in terms of validity. We point out promises and pitfalls of the two-stage approach and make suggestions for its implementation and evaluation. We conclude that all empirical research to date supports the use of behaviorally specific compared to broad questions, that a standard definition of rape and its components of act, tactics, and nonconsent is imperative to move the field forward, and that research to systematically validate methods of detecting rape victimization is needed. To that end, we propose an agenda.

Keywords

construct validity, measurement, rape, self-report

The problem of violence against women first became a focus of scientific inquiry during the second wave of the feminist movement in the United States. Consequently, the scientific need to assess perpetration and victimization of unwanted sexual acts became a priority. Early strategies included the Federal Bureau of Investigation Uniform Crime Reports (UCR) and the National Crime Victimization Survey (NCVS; for a review, see Kilpatrick, 2004).

¹Georgia State University, Atlanta

²Ohio University, Athens

³University of Arizona, Tucson

Corresponding Author:

Sarah L. Cook, Georgia State University, Department of Psychology, PO Box 5010, Atlanta, GA 30302

Email: scook@gsu.edu

The national rates of violence against women and sexual victimization obtained by these endeavors drew attention to sexual assault as an endemic problem in U.S. society. Whereas these early efforts were a step in the right direction, the specific intent and methodological limitations of the UCR and NCVRS painted an incomplete picture of women's assault experiences (e.g., Biderman & Lynch, 1991; Koss, 1992; Russell, 1984). As a result, self-report surveys, such as the Sexual Experiences Survey (Koss, Gidycz, & Wisniewski, 1987; Koss & Oros, 1982) were developed as innovative approaches to detection of rape.

The purpose of this article is to provide a brief overview of emerging directions in the measurement of rape. The focus of the article is on the measurement of the most extreme form of sexual victimization, rape and attempted rape, and how it has evolved. Specifically, we discuss two approaches to the detection of rape victimization in survey methods and how these two approaches can be evaluated in terms of validity. To provide context for our discussion, we begin by examining operational definitions of rape and how these, too, have evolved, where consensus has emerged, and where it eludes the field. Although we focus on rape, we believe that as the field moves forward it needs to continue to measure the continuum of sexual violence, including acts that involve psychological coercion. Even though rape by psychological coercion is not always considered criminal, it is prevalent and has potentially traumatic impacts (Koss et al., 2007).

Measurement is a cornerstone in any research enterprise and informs public policy. The purpose of rape measurement is to identify cases accurately. In public health terminology, the identification of cases is known as surveillance. Once identified, cases may be used to estimate the frequency, incidence, or prevalence of rape. Measures designed to identify rape cases may also be used to select participants who have experienced rape for additional data collection on its impact on physical, psychological, social, and economic health individually and on children and families. Finally, a method to measure frequency and severity of victimization is needed to evaluate the outcome of treatment or risk reduction interventions. Accurate identification is essential; without it, systematic tracking and estimates of prevalence and incidence rates, related morbidity and mortality impacts, and the economic costs across the life span are compromised. Furthermore, data on sexual victimization provide an important empirical foundation for the development and evaluation of policy responses in mental health, criminal justice, public health, education, and other arenas as well as the establish of benchmarks for measuring the effectiveness of primary prevention and risk reduction interventions.

Definitions

The measurement of any behavior or experience hinges on its definition. Here we focus on how the World Health Organization (WHO; Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002), the United States Department of Justice (U.S. DOJ; Bureau of Justice Statistics, 2006), and the U.S. Centers for Disease Control and Prevention (U.S. CDC; Basile & Saltzman, 2002) define rape because these world-recognized bodies publish estimates of the nature and scope of rape and sexual assault, set policy, and direct resources accordingly. Sexual assault and rape definitions generally have three components. One component describes the act

(i.e., nature of the unwanted sex that was compelled), a second describes the tactic used (i.e., how the act was compelled, such as physical force or threat of force), and a third specifies the expression of nonconsent. Definitions of the broad category of sexual assault and the specific category of rape have evolved considerably over several decades and a degree of consensus has emerged on the acts that are classified as sexual assault and rape. For example, in earlier decades rape was defined only as penile-vaginal penetration. The U.S. DOJ and the WHO (2002) now advance definitions that include penetration of the vulva or anus (WHO) and mouth (DOJ) by a penis, other body part, or other object. Although the U.S. CDC does not use the legal term “rape,” the agency’s definition clearly aligns with the WHO’s and U.S. DOJ’s definitions. The WHO definition does not include oral penetration.

Less consensus exists for the second component of the definition of rape that involves the tactic used or how the unwanted act is compelled. The WHO employs the terms “*physically forced or otherwise coerced* penetration.” In this phrase, physical force is named as a type of coercion. The U.S. DOJ includes “*forced* sexual intercourse including both *psychological coercion as well as physical force*.” Threatening rape is considered attempted rape. The U.S. CDC does not specify tactics.

Lack of consent, the circumstances that render the inability to consent, or the inability to refuse comprise the third definitional component of rape. The WHO does not define consent or lack thereof, nor does the U.S. DOJ. Clearly, the omission is problematic because critics have focused on language to assess nonconsent in the SES, most particularly in cases of alcohol-facilitated rape (Gilbert, 1993, 2005). Methods of expressing nonconsent or manifesting inability to consent are critical to include in definitions of rape.

Experts agree that if the task of creating a clear and concise definition is difficult, operationalizing the definition to create a reliable and valid measurement instrument is even more challenging. In 1987, Smith called rape detection the “biggest methodological challenge” in survey research (p. 185). Among highly personal and sensitive behaviors and experiences, including other forms of interpersonal violence, rape and other forms of sexual violence are probably the most difficult experiences to measure. They are rarely observed and occur in private places. Survey research has been open to criticisms because it is based on self-report data, but it is important to note that health and crime data ultimately rest on victims’ self-report. In fact, the validity of virtually all data on sexual violence is potentially compromised by victims’ decisions to self-disclose or not (Testa, VanZile-Tamsen, Livingston, & Koss, 2004).

The Challenge of Self-Report

Kilpatrick and colleagues (Kilpatrick, Edmonds, & Seymour, 1992) describe self-report as a process that begins when individuals first *perceive* a potentially traumatic experience and *encode* it into emotional, sensory, and narrative memories. Victims’ narratives of their experiences do not necessarily involve remembering an unwanted sexual experience as rape, a phenomenon known as unacknowledged rape (e.g. Fisher, Daigle, Cullen, & Turner, 2003; Kahn & Andreoli Mathie, 2000; Koss, 1985; Layman, Gidycz, & Lynn, 1996). Unacknowledged rape may occur in up to 50% of victims (Fisher, Cullen, & Turner, 2000).

Next, an individual cannot report rape in a study unless the sampling design includes them. Many studies use convenience samples (e.g., Edwards, Desai, Gidycz, & VanWynsberghe, 2009; Edwards, Kearns, Calhoun, & Gidycz, 2009; Turchik, Probst, Irvin, Chau, & Gidycz, 2009). Depending on composition, convenience samples yield rates of rape (often measured as lifetime prevalence or from the age of adolescence on) from 25% to 54% (e.g., Edwards, Desai, et al., 2009; Edwards, Kearns, et al., 2009; Turchik et al., 2009). Large-scale nationally representative studies of college and community women typically report lifetime rates for rape or rates from adolescence on that are comparable to rates found in convenience samples often conducted with college students. Fisher et al. (2000) found that 10.1% of their national sample of college women had experienced a rape prior to their assessment. Similarly, Kilpatrick, Resnick, Rugiero, Consoconti, and McCauley (2007) reported lifetime prevalence rates of rape of 18% for the general population and 11.5% for college student samples. In contrast, the prevalence of rape among at-risk samples of women is often much higher. For example, sexual assault prevalence rates among inmates in correctional facilities approach 100% (see Blackburn, Mullings, & Marquart, 2008; Cook, Smith, Tusher, & Raiford, 2005).

Once a respondent is included in a sample, the third step entails accurately cueing her recall. This step is crucial. Questions must jog recall of the type of experiences the survey aims to identify. At the same time, the selection of questions must fully implement the definition of rape that the researchers have adopted. We return in-depth to this matter later, but here describe some of the many factors beyond the content of the questions themselves that influence the success of questions in leading the participant to remember sexually assaultive incidents. Among these influences are the number of questions asked, phrasing, and the subject matter of the survey and the surrounding questions in which rape screening appears. For example, Abbey, Parkill, and Koss (2005) manipulated whether the frame of reference used at the beginning of the questions influenced rates of self-reported victimization and perpetration. Although both surveys contained the same experiences, the methodology varied the order of phrases in screening items. In one version, the type of unwanted sex act opened the questions, and in the other, the tactic utilized to compel the unwanted act appeared first. For both men and women, having the tactic as the primary clause resulted in higher rates of victimization and perpetration experiences, however, the results were more pronounced for men. When the tactic was listed first, rather than the sex act, an additional 13% of women and 33% of men reported victimization and perpetration, respectively. These results underscore the importance of taking into consideration the effectiveness of questions in terms of cueing appropriate memories from the perspective of the focus of inquiry (Schwarz, 2007).

The fourth step of self-report is that women must be willing to divulge their experiences. Participants control what they disclose. In addition to purposeful decisions, participants may unconsciously fail to disclose because for various reasons they do not remember the assaultive incident. Indeed, research on the cognitive aspects of survey methodology have underscored that once memories have been retrieved, respondents edit them to formulate their response. During the editing process, respondents use criteria such as social desirability, intrusiveness, politeness, and consistency, and threat to their self-image (Ongena & Wil

Dijkstra, 2007). Thus there are many reasons that adult respondents may decide not to disclose: they do not think their experience fits the meaning of any of the questions; they do not feel comfortable with the interviewer; they feel too much shame, especially in a face-to-face setting; they doubt that the survey is really anonymous or confidential and fear consequences that may follow disclosure, including retribution, stigma, blame, disbelief, and minimization of the experience (Testa, Livingston, & VanZile-Tamsen, 2005). Finally, researchers must decide that the response fits the scoring criteria they have created on the basis of the underlying definitions (see above for definitions of sexual violence, sexual assault, and rape). We turn to this issue next.

Approaches to Measuring Rape

Items to identify rape may be phrased in multiple ways. Phrasing flows directly from the survey design and procedures to score and classify the responses.

Broad Questions

The National Crime Victimization Survey (NCVS) initially used a broad item that acted as a gate question to cue recall of rape experiences. An answer of “yes” to the broad question, such as “have you been raped,” set off specific follow-up questions (Koss, 1992). Responses to the follow-up questions determined whether the experience was classified as rape. Critics argued that broad questions led to the underdetection of rape because they were inadequate to accomplish cueing and disclosing rape. If the response to the gate item was “no,” administration of the additional items related to sexual victimization was skipped. Asking a broad question about rape is woefully inadequate for a number of reasons (Koss, 1992, 1996). For example, the persistent stereotype of rape as a brutally violent crime between strangers may lead victims to discount other of their experiences that in fact reflect the three central elements of rape discussed earlier (Kilpatrick, 2004). The gate strategy is not inextricably linked to broad questions, however, as we discuss shortly.

Behaviorally Specific Questions

Recognizing the limitations of using a term that requires respondents to be familiar with official definitions of rape and to overcome biases in how rape is stereotypically perceived, researchers initiated use of behaviorally specific questions (Koss et al., 1987; Koss & Oros, 1982) such as the Sexual Experiences Survey (SES). The SES included the defining characteristics of rape within the survey items themselves (i.e., the sexual act, the type of coercion or predation, and the lack consent), the questions were administered to all respondents, and thus data to identify rape were directly available for every person. For example, the Revised SES includes the following: “A man put his penis into my vagina”, or “someone inserted fingers or objects without my consent by using me sexually when I was asleep or unconscious *from alcohol*, and when I came to (*regained consciousness*) I could not give consent or stop what was happening” (Koss et al., 2007). A positive response

to this question would directly identify the respondent as having been raped without necessitating follow-up items that were intended solely for case identification although additional items may be used to develop a more detailed understanding of the characteristics and context of the unwanted acts that are reported.

Other researchers have adopted this direct approach when developing measures of victimization (and perpetration) for rape and other forms of sexual assault. At least nine self-report instruments other than the SES follow this model although not all measures consistently include all definitional elements: The Abuse Severity Measure (Lesserman et al., 1997), Aggressive Sexual Behavior Inventory (Mosher & Anderson, 1986), Assessment of Sexual Aggression Scale (Meyer, Muenzenmaier, Cancienne, & Struening, 1996), Coercive Sexuality Scale (Rapaport & Burkhart, 1984), the Conflict Tactics Scale-Revised (Straus, Hamby, Boney-McCoy, & Sugarman, 1996), the Measure of Wife Abuse (Rodenburg & Fantuzzo, 1993), Severity of Violence Against Women Scale (Marshall, 1992), Sexual Coercion in Intimate Relationships Scale (Shackelford & Goetz, 2004), and the Use of Force in Sexual Experience Scale (Petty & Dawson, 1989). Behaviorally specific questions have also been used in large-scale studies including the World Health Organization's Multi-Country Study (Ellsberg, Jansen, Watts, Garcia-Moreno, & the WHO Multi-country Study on Women's Health and Domestic Violence against Women Study Team (2008)), Rape in America (Kilpatrick et al., 1992), and the National Violence Against Women Study (Tjaden & Thonnes, 1998). Furthermore, behaviorally specific questions undergird a large majority of research on rape victimization and perpetration.

Results from a quasi-experimental investigation provide compelling evidence for the superiority of behaviorally specific questions compared to questions utilized in the National Crime Victimization Survey, which are broad and vaguely worded (Fisher, 2009). Specifically, comparison of two methodologically identical national samples of college women's survey responses revealed that behaviorally specific questions resulted in approximately 11 times more disclosures for completed rapes than the NCVS questions. The accumulating and consistent evidence from nationally representative studies (for a review see Cook & Koss, 2005), together with Fisher's (2009) study clearly support the conclusion that broadly worded questions combined with a gate strategy lead to underdetection. This study did not fully cross conditions and systematically examine the outcome of using a behaviorally specific item, but used a variation of a gate approach where rape determination was made at the level of follow-up questions. We examine this combination of item and design strategy shortly.

Popular use of the behaviorally specific approach is not sufficient to advance it as a standard. Establishing the validity of any approach is imperative, and the task is neither simple nor straightforward. The most fundamental matter relates to construct validity: Does a survey and its component items truly measure what it purports to measure? A primary question is whether respondents interpret behaviorally specific questions in the way intended by the instrument's authors (Koss et al., 2007). If the interpretation differs between the respondent and author, the question will not "cue" the respondent to the relevant experiences. This issue quickly becomes complex considering that it is necessary to demonstrate that the questions have equivalent meaning to respondents in groups diverse on factors as ethnicity, socioeconomic status, culture, and age.

Data from a study comparing self-report narratives to SES questions provides some evidence on this issue (Testa et al., 2004). Testa and colleagues obtained reports by telephone on the SES from 1,000 community women, diverse in terms of ethnicity, residing in Buffalo, New York. After completing the SES, researchers asked respondents to describe their most recent experience in their own words and then answer specific questions about the incident including (a) whether penetration had occurred, (b) whether physical force or injury was involved, and (c) what her response was or whether she resisted. After audiotapes were transcribed, a group of postdoctoral fellows in psychology who had no prior experience with sexual aggression served as coders. They read the transcribed descriptions and responses to the specific questions and were instructed to choose the SES item that best fit the transcript. If the coders could not select an SES item, they could determine that the incident was (a) not an incident of sexual assault; (b) unwanted sex, but not one described in an SES item; or (c) unsuited to coding due to insufficient information. For rape, agreement ranged between 81% and 94% and for coercion 86% to 95% agreement was found. A potential limitation to this study is that respondents were primed to describe their experiences in ways consistent with the SES. Regardless, Testa et al.'s study represents the type of examination needed to support a survey's validity.

Two-stage Design

Several recent investigations have used what we will call a two-stage approach (Fisher & Cullen, 2000; Kilpatrick et al., 2007) to correct for potential *overdetection* by instruments such as the SES. These investigations combine behaviorally specific questions with subsequent incident reports (i.e., follow-up questions) that are administered to every respondent. Classification of rape is accomplished in the second stage of self-report. The rationale advanced by Fisher and Cullen (2000) is that

Using *only* behaviorally specific questions . . . assumes that the respondent understands the experience she is being asked about and that these questions are able to cue accurate recall by the respondents (i.e., a rape question cues all rape victims to answer yes; an attempted rape question cues all attempted rape victims to answer yes; and so on). These assumptions are problematic, especially in the absence of followup [*sic*] questions to probe in detail what actually transpired in any given incident. (p. 358)

We note that this assertion conflates the wording of the questions with the manner in which they are administered and scored. Behaviorally specific questions can be long and complex, two characteristics that may contribute to error and reduce validity. In the SES-Revised, for example, each item queries the sexual act, tactic used, and expression of non-consent or reason for an inability to consent. In an effort to be comprehensive and to map onto existing legal definitions of rape and other forms of sexual assault, each question contains multiple components that may lead to ambiguity and complexity for respondents. It is possible that respondents may believe that some parts of the question apply to them but that

others do not. In this case, they may not know how to respond. Furthermore, involved questions may overload a respondent's working memory (Just & Carpenter, 1992), causing respondents to forget parts of the questions and provide partial answers (Tourangeau, Rips, & Rasinski, 2000). Finally, distinctions among strategies (e.g., completed and attempted rape) may be hard for some respondents to comprehend. Two-stage measurement may reduce measurement error if both stages are behaviorally specific and the incident reports (follow-up questions) separately operationalize each component of rape definitions (i.e., act, tactic, and consent). As with the direct approach, the initial question in a two-stage approach would need to cue all potentially relevant experiences (be sensitive) and exclude none (be specific). In a recent article, Fisher (2009) suggests that the two-stage measurement process "appears to be a promising way to address the measurement error typically associated with a single stage process, although it still needs rigorous testing" (p. 144).

We agree that further rigorous testing is required and have concluded that the National College Women Sexual Victimization study (NCWSV) as implemented does not provide even preliminary evidence for the superiority of the two-stage design. It is possible that the two-stage approach reduces error from the behaviorally specific question, but it is also conceivable that it introduces additional error from the incident report questions. In the material that follows, we offer a critique of the implementation of the two-stage design used by Fisher and colleagues. The NCWSV used 20 behaviorally specific questions to detect experiences of sexual assault. The question that screened for rape was

Since school began in the Fall 1996, has anyone made you have sexual intercourse by using force or threatening to harm you or someone close to you? Just so there is no mistake, by intercourse, I mean putting a penis in your vagina.

All respondents regardless of their response to this item were asked four additional follow-up or incident-report questions:

1. Was the sexual contact in this incident threatened, attempted, or completed (at least some sexual contact actually happened?), followed by
2. Tell me which of the following actually occurred to you during this incident. Just say yes or no. Did you experience penis in your vagina, mouth on your genitals, your mouth on someone else's genitals, penis in your anus or rectum, finger in your vagina, finger in your anus or rectum, another object in your vagina, another object in your anus or rectum, or none of these?

If the respondent said "yes" to Question 1 and "yes" to one of the items in Question 2, a gate strategy was then introduced where the following items were administered:

3. Was physical force *actually* [italics added] used against you in this incident? and
4. Were you threatened with physical force in this incident?

Regardless of a respondents' answer to the rape screening item at the first stage, the experience was categorized as rape only at the second stage where answers were "yes" to incident report Items 3 and 4 at the second stage.

In the NCWSV study of over 4,000 participants, only 25% of rapes that were identified in the first stage were confirmed by second stage incident reports; 75% were not. Of the 75%, 50% were "down" classified to another form of victimization, 18% were undetermined, and 6% were deemed invalid either because the experience was out of the reference period or the respondent did not provide enough details. On the other hand, of the 199 rapes used to estimate incidence and prevalence, 55% were not detected at the first stage but were "up" classified at the second stage from their endorsement of nonrape items at the first stage (and denial of the rape item). To understand how three quarters of initial respondents fell down at the second stage, or how others moved up from other behaviorally specific questions utilized in the first stage, we make several observations about the implementation of the two-stage approach described by Fisher and Cullen (2000).

First, the initial behaviorally specific question to detect rape defined sexual intercourse as, "Just so there is no mistake, by intercourse, I mean putting a penis in your vagina." Based on earlier discussion, we note that this is a narrow definition of rape typically confined to criminal justice data collection. The incident report that followed up responses, however, broadened the definition of sexual intercourse from penile-vaginal penetration to include a range of unwanted sexual acts that are more consistent with public health definitions. The change in operational definition of what constitutes sexual intercourse from one stage to another is problematic. Thus a respondent whose experience involved forced fellatio would presumably answer the initial behaviorally specific question with "no" and therefore would not be counted as a rape victim. However, when confronted with the incident follow-up and asked about seven forms of sexual contact other than penile-vaginal intercourse, it would be appropriate to respond "yes" to one of the choices that at the second stage has now been included as rape. We assume, but cannot definitely conclude that any of the forms of rape other than penile-vaginal intercourse would have led to this outcome and explain at least in part how 55% of the 199 identified rape cases ($N = 199$) had inconsistency between their Stage 1 and Stage 2 response.

We further note that the third and fourth follow-up questions in the incident report appear not only to be administered following a gate strategy but they are also broad questions and not behaviorally specific. The third follow-up question reiterates a portion of the behaviorally specific question and adds the word, "actually." It provides no further elaboration of what constitutes physical force. The fourth follow-up question repeats another portion of the behaviorally specific question about threat but eliminates another: "Were you threatened with physical force in this incident?" It does not ask a second time about whether someone close to the respondent was threatened with physical force. If respondents had answered "yes" to the Stage 1 screening question because someone close to them was threatened with harm, they would have answered these last two questions with "no," and would have presumably been downgraded from the rape category.

In addition to error that is introduced by introducing a gate design into the incident report, the Stage 1 and Stage 2 questions introduce at least three sources of measurement

error due solely to their construction. Additional error may arise on the part of the respondent. For example, the respondent may have been unwilling to disclose penetration when expressed in eight choices that were far more graphic than the initial question. Sexual behavior is a socially sensitive subject (Lee & Renzetti, 1990; Seiber & Stanley, 1988). At least some error in responding to surveys of sexual behavior is likely due to conscious editing of the answers by the respondent. Second, the respondent may have answered “no” to all choices when the word “actually” was added at the second stage. Follow-up Questions 1 and 3 (Fisher & Cullen, 2000, p. 363) could be perceived as confrontational, indicating disbelief and, as a result, invalidating for the respondent who initially endorsed an item that they were instructed to interpret as referring to penile-vaginal intercourse. The word, “actually” may sound similar to traumatizing reactions the woman may have heard from others when relating her experience, akin to secondary victimization (Campbell & Raja, 1999).

Conclusion

Our discussion of the evolution of rape definitions, rape survey items, and methods of classification in rape victimization in survey research, together with our explication of the complexity involved both in creating new studies and interpreting the results of existing literature leads us to three conclusions. First, accumulating data strongly support the superiority of behaviorally specific questions compared to broad questions. This point needs no further elaboration given the consistency of rape estimates across studies that have used them (Cook & Koss, 2005). Second, the field remains hampered by the lack of a standard definition of rape and its components of act, tactics, and nonconsent. Our third and final conclusion is integrally connected to the second: researchers, all of whom are working toward accurate detection of rape, must turn attention to systematically validating methods of detecting rape victimization with empirical methods. Toward this end we propose an agenda.

Toward a Uniform Definition

No agreed upon definition of sexual violence exists. The use of various definitions of sexual violence across studies renders comparisons of incidence and prevalence rates impossible. Given the costly public health problem that sexual violence presents, assessment that lends itself to systematic national tracking of incidence, prevalence, morbidity and mortality, and costs to society is imperative (Koss, White, & Kazdin, 2010a, 2010b). National and international agencies have put forward definitions of rape, but they have not been integrated in any way. Although the benefits of adopting standardized definitions across investigations are clear, we realize that task is difficult to accomplish. However, we are at a point where work toward this goal is necessary to advance knowledge and policy.

Qualitative and quantitative exploration are needed to guide the development of consensus definitions of all three components of rape: the sexual act, the tactic used to attempt or complete the act, and the expression of nonconsent. As stated earlier, consensus is emerging on the nature of sexual acts that constitute rape. Definitions used in most survey instruments

have broadened to encompass forms of violation beyond penile penetration of the vagina. We caution that the majority of direct approach measures have been developed for use in college student samples. Not surprisingly, they reflect the nature of sexually aggressive behavior typical in this developmental period. It is unclear whether existing instruments capture other ways in which rape may occur in other populations (Cook & Parrott, 2009). For example, the experiences of women with less social, economic, or political privilege than those enrolled in higher education are far less documented. Forms of victimization specific to minority status, defined in terms of race, ethnicity, immigration status, or other dimensions of diversity, may be found in these women's narratives. In spite of challenges present, consensus on the nature of sexual acts that occur in rape appears achievable.

Considerable work is needed to identify and define the range of *tactics* used to attempt or complete the sexual act. Most instruments, and thus definitions, include physical force, but not all include verbal threats, and only some inroads have been made to include alcohol and other drugs as tactics (Kilpatrick et al., 2007; Koss et al., 2007). The nature of threats made during a rape needs further exploration. In the SES-R and the NCWSV, threat is defined as a threat of physical harm to self or someone close, but it is possible to imagine other types of threats with serious life-altering consequences, such as threats to sabotage one's employment, immigration status, or reveal socially stigmatizing information. Alcohol is present in the majority of rapes on college campuses (Kilpatrick et al., 2007; Mohler-Kuo, Dowdall, Koss, & Wechsler, 2004). Definitions must include tactics such as surreptitious administration to incapacitate the victim, pressure to consume to take advantage of lowered inhibitions, and opportunistic preying on victims who are incapacitated or in other ways unable to consent (Koss et al., 2007).

The greatest challenge ahead is to understand and define nonconsent. Existing measures of rape use several phrases such as "when you didn't want to," or "against your will," to signify nonconsent, and others omit any specific mention of consent. Little social and behavioral research explores consent. What exists illuminates an inherent complexity. For example, Hickman and Muelhenhard (1999) asked college men and women to describe how they and their partner communicated his or her consent verbally or nonverbally. Six categories of consent "signals" emerged: direct verbal (e.g., I would like to sleep with you), direct nonverbal (e.g., you don't say anything, you just start having sexual intercourse), indirect verbal (e.g., you ask if he or she has a condom), indirect nonverbal (e.g., you help her or him undress), intoxication (e.g., you say, "I'm feeling a little drunk"), and no response (e.g., you do not resist his or her sexual advances). Only one category of non-consent emerged (direct refusal, e.g., you say, "no."). The most frequently used categories were nonverbal. Note that participants were not directly asked about communicating nonconsent. A parallel study of how young men and women communicate nonconsent is likely to be illuminating.

In contrast, legal scholars offer other categories of consent. For example, Westen (2004) offers three pairs of competing conceptualizations of consent: factual versus legal (can describe the distinction between consensual sex and statutory rape), attitudinal versus expressive (describes the difference between a subjective, internal choice and an objective expression of choice), and actual versus imputed consent (e.g., can distinguish between a woman with attitudinal or expressive consent vs. the law treating a woman as if she consented

due to a prior understanding as in a marital rape exemption). Hickman and Muehlenhard's (1999) categories align most closely with Westen's expressive consent category although Westen does not address nonverbal expressions of consent. It is beyond the scope of this article to describe in greater detail how all of these categories relate to one another, or as Westen argues, are often confused. We point them out, however, to illustrate the gulf between social/behavioral and legal conceptualizations and to suggest that efforts to bridge them may improve the consent component of rape definitions.

Establishing validity. The construct validity of any measurement instrument develops over time and through multiple qualitative and quantitative investigations (Cronbach & Meehl, 1955). Here we suggest a few of many possible avenues. Once components of a consensus rape definition are conceptualized, researchers must turn the concepts into questions and ensure that diverse groups of research participants receive questions as researchers intend (i.e., face and content validity). Fisher and Cullen (2000), for example, suggest focus group participants could describe what the components of rape definitions mean to them when placed in the context of a victimization survey. Likewise, focus groups could highlight how trends in language and sexual behavior norms influence meaning. Focus groups of survivors exploring whether the questions map onto their experiences may identify missing elements in questions. An example of this type of work is Hamby and Koss' (2003) exploration of terms in five groups divergent with respect to ethnicity, gender, religious beliefs, education, and geographic location. Two of several themes from these groups are cause for caution. Researchers often use several words to convey a single idea, such as unwanted, forced, or nonvoluntary, but these three terms did not hold the same, single meaning to participants. In addition, comments suggested that while questions should be clear and specific, they should not be unnecessarily graphic.

Social and behavioral scientists should collaborate with legal scholars and practicing attorneys to understand when questions reflect statutory language and when they do not. For example, Gylys and McNamara (1996) asked 310 prosecutors to compare items from the SES (Koss & Oros, 1982) to statutes from the Ohio Revised Code. Data indicated that the SES items meant to identify rape and attempted rape were commensurate with legal definitions of sexual offenses of rape and attempted rape with at least 88% of their sample confirming each item. We are not suggesting that social and behavior surveys should always parallel the law. Indeed, rape reform efforts could be buttressed by empirical data that illustrate the limits of existing laws to criminalize coercive and nonconsensual sexual experiences as rape.

With evidence for face and content validity secured, research must then focus on criteria-related validity, such as predictive, concurrent, convergent, and discriminant validity.

Reducing error. In the direct approach, respondents are identified as victims or nonvictims based on responses to a series of questions that include specific definitions of each component of rape definitions. A critique of this approach is that it allows no opportunity to verify responses to the questions. In the two-stage approach as implemented in the NCWSV study, follow-up questions are intended to verify answers. We demonstrate, however, that the two-stage approach is open to the same criticism because as presently studied, second-stage responses are selected as the "gold standard," when in fact, they are just a different version

of self-report questions. In lieu of reducing error, we argue that the follow-up questions introduce additional sources of error. Fisher and Cullen (2000) allude to this

For example, in our NCWSV study, nearly 1 in 5 incidents that were initially classified as rape by the responses to the screen questions were then classified as “undetermined” because the respondents either refused to answer or answered “don’t know” to one or more questions in the incident report used to categorize incidents as a rape. It is possible, of course, that the incident report “did its job” by diverting from our count of rape victimization those incidents that did not meet all of the legal criteria for rape (i.e., criteria measured by the questions in the incident report). It is also possible, however, that some “real” rape victims were not counted because they did not understand questions in the incident report or wearied at having to answer a second round of questions about a potentially painful event in their lives. In the absence of further research, discussions of possible measurement error associated with the use of incident reports will remain speculative. (pp. 379)

We agree, noting the intrusive nature of the questions. Fisher and Cullen (2000) suggest that experimental designs, possibly using vignettes, could be utilized to better understand how the range and wording of questions affect how respondents report victimizations. The two-stage approach may indeed produce different results from the direct approach if it were implemented as we suggest. Additional studies would be needed to identify reasons for the discrepant data, only one of which may be that the two-stage method is less prone to error. Nevertheless, the goal of the two-stage approach, to reduce error, deserves full attention. Other methods of reducing error—for example, reducing the cognitive burden of complex questions but not changing the intended meaning of the questions—could be developed and evaluated. Within-subject designs, complemented with interview techniques, may help to identify underlying reasons for discrepancies across methods of questions.

In conclusion, in this article we have attempted to identify some critical measurement and assessment issues to which researchers of sexual violence need to attend. It is likely that regardless of whether a one- or two-stage approach is utilized that researchers’ goals converge in that they are working toward accurate detection of rape cases. By focusing our efforts on validating our methods, which we acknowledge is not an easy task, we hope to see progress in the assessment of rape and other forms of sexual assault. Progress in assessment ultimately sets the foundation for policy decisions that will ultimately lead to better treatment for rape survivors and improved methods for prevention.

Authors’ Note

Portions of this article are based on Measurement of Sexual Victimization, presented by Mary P. Koss as the *Allen Edwards Memorial Lecture*, University of Washington, Seattle, WA on February 3, 2009.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

The authors received no financial support for the research and/or authorship of this article.

Note

1. The vast majority of rape victims are female. Thus we will use gendered pronouns hereafter.

References

- Abbey, A., Parkhill, M. R., & Koss, M. P. (2005). The effects of frame of reference on responses to questions about sexual assault victimization and perpetration. *Psychology of Women Quarterly, 29*, 364-373.
- Basile K. C., & Saltzman L. E. (2002). *Sexual violence surveillance: Uniform definitions and recommended data elements*. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.
- Biderman, A. D., & Lynch, J. P. (1991). *Understanding crime incidence statistics: Why the UCR diverges from the NCS*. New York, NY: Springer-Verlag.
- Blackburn, A. G., Mullings, J. L., & Marquart, J. W. (2008). Sexual assault in prison and beyond: Toward an understanding of lifetime sexual assault among incarcerated women. *The Prison Journal, 88*, 351-377.
- Bureau of Justice Statistics. (2006). *Criminal victimization in the United States, 2006, statistical tables*. Washington, DC: US Department of Justice.
- Campbell, R., & Raja, S. (1999). Secondary victimization or rape victims: Insights from mental health professionals who treat survivors of violence. *Violence and Victims, 14*, 261-175.
- Cook, S. L., & Koss, M. P. (2005). More data have accumulated supporting date and acquaintance rape as significant problems for women. In D. L. Loske, R. J. Gelles, & M. M. Cavanaugh (Eds.), *Current controversies on family violence* (2nd ed., pp. 97-112). Thousand Oaks, CA: SAGE.
- Cook, S. L., & Parrott, D. J. (2009). Exploring a taxonomy for violence against women: Can it aid conceptual clarity? *Aggressive Behavior, 35*, 463-476.
- Cook, S. L., Smith, S. G., Tusher, C. P., & Raiford, J. (2005). Self-reports of traumatic events in a random sample of incarcerated women. *Women & Criminal Justice, 16*, 107-126.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests *Psychological Bulletin, 52*, 281-302.
- Edwards, K. M., Desai, A. D., Gidycz, C. A., & VanWynsberghe, A. (2009). College women's aggression in relationships: The role of childhood and adolescent victimization. *Psychology of Women Quarterly, 33*, 255-265.
- Edwards, K. M., Kearns, M. E., Calhoun, K. S., & Gidycz, C. A. (2009). College women's reactions to sexual assault research participation: Is it distressing? *Psychology of Women Quarterly, 33*, 225-234.
- Ellsberg, M. , Jansen, A.F.M., Heise, L, Watts, C.H., Garcia-Moreno, C. & the WHO Multi-country Study on Women's Health and Domestic Violence against Women Study Team (2008). Intimate partner violence and women's physical and mental health in the WHO multi-country study on women's health and domestic violence: an observational study. *The Lancet, 371*, 1165-1172. doi:10.1016/S0140-6736(08)60522-X

- Fisher, B. S. (2009). The effects of survey question wording on rape estimates: Evidence from a quasi-experimental design. *Violence Against Women, 15*, 133-147.
- Fisher, B. S., & Cullen, F. T. (2000). Measuring the sexual victimization of women: Evolution, current controversies and future research. *Measurement and Analysis of Crime and Justice, 4*, 317-390.
- Fisher, B. S., Cullen, F. T., & Turner, M. G. (2000). *The sexual victimization of college women* (NCJ 182369). Washington, DC: National Institute of Justice.
- Fisher, B. S., Daigle, L. E., Cullen, F. T., & Turner, M. G. (2003). Reporting sexual victimization to police and others: Results from a national-level study of college women. *Criminal Justice and Behavior, 30*, 6-38.
- Gilbert, N. (1993). Examining the facts: Advocacy research overstates the incidence of date and acquaintance rape. In R. J. Gelles & D. R. Loseke (Eds.), *Current controversies on family violence* (pp. 120-132). Newbury Park, CA: SAGE.
- Gilbert, N. (2005). Advocacy research overstates the incidence of date and acquaintance rape. In D. R. Loseke, R. J. Gelles, & M. M. Cavanaugh (Eds.) *Current controversies on family violence* (2nd ed., pp. 117-129). Thousand Oaks, CA: SAGE.
- Gyls, J. A., & McNamara, J. R. (1996). A further examination of the validity for the Sexual Experiences Survey. *Behavioral Sciences and the Law, 14*, 245-260.
- Hamby, S. L., & Koss, M. P. (2003). Shades of gray: A qualitative study of terms used in the measurement of sexual victimization. *Psychology of Women Quarterly, 27*, 243-255.
- Hickman, S. E., & Muehlenhard, C. L. (1999). "By the semi-mystical appearance of a condom": How young women and men communicate sexual consent in heterosexual situations. *Journal of Sex Research, 36*, 258-272.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension. *Psychological Review, 99*, 122-149.
- Kahn, A. S., & Andreoli Mathie, V. A. (2000). Understanding the unacknowledged rape victim. In C. B. Travis & J. W. White (Eds.), *Sexuality, society, and feminism: Psychological perspectives on women* (pp. 377-403). Washington, DC: American Psychological Association.
- Kilpatrick, D. G. (2004). What is violence against women? Defining and measuring the problem. *Journal of Interpersonal Violence, 19*, 1209-1234.
- Kilpatrick, D. G., Edmunds, C. M., & Seymour, A. K. (1992). *Rape in America: A report to the nation*. Arlington, VA: National Victim Center and the Crime Victims Research and Treatment Center.
- Kilpatrick, D. G., Resnick, H. S., Rugiero, K. J., Conoscenti, L. M., & McCauley, J. (2007). *Drug-facilitated, incapacitated, and forcible rape: A national study. Final report*. Retrieved Oct 30, 2009 from <http://www.ncjrs.gov/pdffiles1/nij/grants/219181.pdf>
- Koss, M. P. (1985). The hidden rape victim: Personality, attitudinal, and situational characteristics. *Psychology of Women Quarterly, 9*, 193-212.
- Koss, M. P. (1992). The underdetection of rape: Methodological choices influence incidence estimates. *Journal of Social Issues, 48*, 61-75.
- Koss, M. P. (1996). The measurement of rape victimization in crime surveys. *Criminal Justice and Behavior, 23*, 55-69.

- Koss, M. P., Abbey, A., Campbell, R., Cook, S., Norris, J., Testa, M., Ullman, S., West, C., & White, J. (2007). The revised Sexual Experiences Survey. *Psychology of Women Quarterly, 31*, 357-370.
- Koss, M. P., Gidycz, C. A., & Wisniewski, N. (1987). The scope of rape: Incidence and prevalence of sexual aggression and victimization among a national sample of students in higher education. *Journal of Consulting and Clinical Psychology, 55*, 162-170.
- Koss, M. P., & Oros, C. J. (1982). The Sexual Experiences Survey: An empirical instrument investigating sexual aggression and victimization. *Journal of Consulting and Clinical Psychology, 50*, 455-457.
- Koss, M. P., White, J., & Kazdin, A. E. (2010a). Violence against women and children: Perspectives and next steps. In M. Koss & J. White (Eds.), *Violence against women and children: Navigating Solutions* (Vol. 2). Washington DC: American Psychological Association.
- Koss, M. P., White, J., & Kazdin, A. E. (Eds.). (2010b). *Violence against women and children: Navigating Solutions* (Vol. 2). Washington, DC: American Psychological Association.
- Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). *World report on violence and health*. Geneva, Switzerland: World Health Organization.
- Layman, M. J., Gidycz, C. A., & Lynn, S. J. (1996). Unacknowledged versus acknowledged rape victims: Situational factors and posttraumatic stress. *Journal of Abnormal Psychology, 105*, 124-131.
- Lee, R. M., & Renzetti, C. M. (1990). The problem of researching sensitive topics: An overview and introduction. *American Behavioral Scientist, 33*, 510-528.
- Lesserman, J., Li, Z., Drossman, D. A., Toomey, T. C., Nachman, G., & Glogau, L. (1997). Impact of sexual and abuse measure. *Psychosomatic Medicine, 59*, 152-160.
- Marshall, L. L. (1992). Development of the Severity of Violence Against Women Scales. *Journal of Family Violence, 7*, 103-121.
- Meyer, I. H., Muenzenmaier, K., Cancienne, J., & Struening, E. (1996). Reliability and validity of a measure of sexual and physical abuse histories among women with serious mental illness. *Child Abuse and Neglect, 20*, 213-219.
- Mohler-Kuo, M., Dowdall, G. W., Koss, M. P., & Wechsler, H. (2004). Correlates of rape while intoxicated in a national sample of college women. *Journal of Studies on Alcoholism, 65*, 37-53.
- Mosher, D. L., & Anderson, R. D. (1986). Macho personality, sexual aggression, and reactions to guided imagery of realistic rape. *Journal of Research in Personality, 20*, 77-94.
- Ongena, Y. P., & Dijkstra, W. (2007). A model of cognitive processes and conversational principles in survey interview interaction. *Applied Cognitive Psychology, 21*, 145-163.
- Petty, G. M., & Dawson, B. (1989). Sexual aggression in normal men: Incidence, beliefs, and personality characteristics. *Personality and Individual Differences, 10*, 355-362.
- Rapaport, K., & Burkhart, B. R. (1984). Personality and attitudinal characteristics of sexually coercive males. *Journal of Abnormal Psychology, 93*, 216-221.
- Rodenburg, F. A., & Fantuzzo, J. W. (1993). The measure of wife abuse: Steps toward the development of a comprehensive assessment technique. *Journal of Family Violence, 8*, 203-228.
- Russell, D. E. H. (1984). *Sexual exploitation: Rape, child sexual abuse, and workplace harassment*. Beverly Hills, CA: SAGE.

- Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology, 21*, 277-287.
- Seiber, J., & Stanley, B. (1988). Ethical and professional dimensions of socially sensitive research. *American Psychologist, 43*, 49-55.
- Shackelford, T. K., & Goetz, A. T. (2004). Men's sexual coercion in intimate relationships: Development and initial validation of the sexual coercion in intimate relationships scale. *Violence and Victims, 19*, 541-556.
- Smith, M. D. (1987). Changes in the victimization of women: Is there a "new female victim"? *Journal of Research in Crime and Delinquency, 24*, 291-301.
- 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.
- Testa, M., Livingston, J. A., & VanZile-Tamsen, C. (2005). The impact of questionnaire administration mode on response rate and reporting of consensual and nonconsensual sexual behavior. *Psychology of Women Quarterly, 29*, 345-352.
- Testa, M., VanZile-Tamsen, C., Livingston, J. A., & Koss, M. P. (2004). Assessing women's experiences of sexual aggression using the Sexual Experiences Survey: Evidence for validity and implications for research. *Psychology of Women Quarterly, 29*, 345-352.
- Tjaden, P., & Thonnes, N. (1998). *Stalking in America: Findings from the National Violence Against Women Survey* (Research in Brief [NJC 169592]). Washington, DC: U.S. Department of Justice.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.
- Turchik, J. A., Probst, D. R., Irvin, C. R., Chau, M., & Gidycz, C. A. (2009). Prediction of sexual assault experiences in college women based on rape scripts: A prospective analysis. *Journal of Consulting and Clinical Psychology, 77*, 361-366.
- Westen, P. K. (2004). Some common confusions about consent in rape cases. *Ohio State Journal of Criminal Law, 2*, 332-359.

Bios

Sarah L. Cook, PhD, professor of psychology at Georgia State University, focuses on measurement of victimization and perpetration, preventive interventions, and ethical issues in violence and trauma research. She has confronted the problem of aggression against women as a peer educator, rape crisis advocate, child protection social worker, and consultant to local and state advocacy organizations. Currently she serves as chair of the board of directors for the Georgia Network to End Sexual Assault.

Christine A. Gidycz, PhD, professor of psychology at Ohio University, directs the Laboratory for the Study and Prevention of Sexual Assault. She has conducted investigations of sexual assault correlates and risk factors, including those that contribute to sexual revictimization. She has also studied the co-occurrence of various forms of violence. She developed the Ohio University Sexual Assault Risk Reduction Program, designed to decrease women's risk for sexual victimization.

Mary P. Koss is a Regents' Professor in the Mel and Enid Zuckerman Arizona College of Public Health at the University of Arizona. She is coeditor of a forthcoming two-book series for the American Psychological Association, *Violence Against Women and Children: Mapping the Terrain* (Vol. 1) and *Violence Against Women and Children: Navigating Solutions* (Vol 2).

Megan Murphy is a clinical psychology PhD candidate at Ohio University. She works in the Laboratory for the Study and Prevention of Sexual Assault directed by Dr. Christine Gidycz. Her research focuses on the physical and psychological effects of sexual assault with a specific focus on the effect of sexual assault on women's body esteem. She completed her undergraduate work at Saint Michael's College in Vermont.