



Published in final edited form as:

Psychol Violence. 2015 July ; 5(3): 285–293. doi:10.1037/a0037584.

Measuring Sexual Aggression Perpetration in College Men: A Comparison of Two Measures

Maria Testa, Joseph H. Hoffman, Joseph F. Lucke, and Colleen E. Pagnan

University at Buffalo

Abstract

Objective—The present study was designed to provide a comparison of rates of self-reported sexual aggression perpetration obtained using two different measures – a version of the Sexual Experiences Survey (SES; Abbey et al, 2007; Koss, Gidycz, & Wisniewski, 1987) and the Sexual Strategies Scale (SSS, Strang, et al, 2013; Struckman-Johnson, Struckman-Johnson, & Anderson, 2003). We also examined the psychometric structure of each measure using Rasch model item analysis (Rasch, 1966).

Method—Two equivalent cohorts of entering freshman males (N = 994 and N = 1043) from a large northeastern university completed online measures at the end of their first semester.

Results—Identical proportions of men reported using intoxication strategies (3%) and physical force (1%) during the past semester on both measures. However, more men reported verbal strategies on the SSS (7.8%) compared with the SES (3.7%), even when restricting to equivalent items. Rasch analysis suggested that the SSS conformed better to a unidimensional continuum of perpetration severity than the SES; however, Rasch analysis did not provide definitive support for either a tactic - based (SSS) nor a tactic plus outcome- based (SES) hierarchy.

Conclusions—Both measures functioned adequately. However, the SSS may be preferred for its better Rasch properties, better assessment of the less severe tactics, and simpler wording.

Keywords

sexual aggression; perpetration; measurement; psychometrics; college; Rasch model

Behaviorally specific self-report measures of sexual aggression perpetration and victimization are crucial to detecting the prevalence of these hidden and socially proscribed behaviors. However, the large discrepancy between rates of victimization self-reported by women and the much lower rates of perpetration reported by men (Spitzberg, 1999) raises concerns about the ability of existing measures to detect perpetration (Kolivas & Gross, 2007). Although methodological variation may contribute to variations in rates, little systematic research has examined the psychometric properties of these perpetration measures or the impact that variations in item wording, order, or context have on reported

Correspondence regarding this article should be addressed to Maria Testa, Research Institute on Addictions, University at Buffalo, 1021 Main Street, Buffalo, NY 14203. testa@ria.buffalo.edu.

Maria Testa, Research Institute on Addictions, University at Buffalo; Joseph H. Hoffman, Research Institute on Addictions, University at Buffalo; Joseph F. Lucke, Research Institute on Addictions, University at Buffalo.

rates (Kolivas & Gross, 2007; Koss et al, 2007). The present study was designed to compare rates of self-reported sexual aggression perpetration among college freshman males obtained using two different measures – a version of the Sexual Experiences Survey (SES; Abbey et al, 2007; Koss et al, 1987) and the Sexual Strategies Scale (SSS, Strang, et al, 2013; Struckman-Johnson, et al, 2003). Although we cannot validate these self-report measures against a known truth or gold standard, differences in prevalence yielded by these measures provide insight into how features of a measure may influence men’s willingness to report sexual aggression (Abbey, Parkhill, & Koss, 2005; Koss et al, 2007). A second aim of the study was examination of the psychometric properties of the two measures using Rasch analysis (Bond & Fox, 2001). In particular, we examined whether items could be arrayed on a continuum of severity that matches existing theory.

The Sexual Experiences Survey

Measurement of sexual aggression has been dominated for decades by the Sexual Experiences Survey (SES, Koss & Oros, 1982; Koss et al, 1987), which pioneered the use of multiple, behaviorally specific items as a way of detecting higher, and presumably more accurate rates of sexual assault perpetration and victimization (Koss, 1993). Over the years, numerous versions of both the victimization and perpetration forms of the SES have proliferated as researchers modify the measure for their own purposes and seek to address perceived weaknesses in the measure (Koss et al., 2007). Although many variations in wording exist, SES-type measures have maintained the original item structure in which each item specifies both a tactic (e.g., verbal persuasion, physical force, victim intoxication) and an outcome (e.g., contact, attempted intercourse, completed intercourse, oral sex). An advantage is that items provide information about both aspects of sexual aggression; however, a potential disadvantage is that the complex structure increases cognitive burden associated with reading and interpreting items.

Despite its wide usage, methodological and psychometric work on the SES has been lacking (Kolivas & Gross, 2007; Koss et al, 2007; see Cook, 2002; Ross & Allgeier, 1996) for exceptions). In one of the few studies to actually compare different ways of assessing sexual aggression, Abbey et al (2005) examined whether alternate item phrasings influence the likelihood of a positive response. The authors created equivalent forms of the SES for perpetrators and victims that crossed 5 outcomes with 7 tactics and varied whether the first phrase of each item presented the tactic or the outcome. These 35 items included additional items assessing verbal coercion (e.g., “by telling lies or making promises that were untrue” and “by showing displeasure, guilt, or anger”) and intoxication strategies (e.g., “when she was passed out or too intoxicated to give consent or stop what was happening”). Although items assessed apparently identical behaviors, rates of self-reported lifetime perpetration were substantially higher for men who were randomly assigned to complete the tactic-first version compared to the outcome -first version. The authors suggest that the tactic-first phrasing cues relevant experiences in memory more effectively than the outcome-first phrasing. However, the higher rates resulting from the tactics-first version mostly reflected higher endorsement of the less severe, verbal items. Both versions began with less severe and moved toward more severe items. The authors suggest that this ordering may have suppressed reporting of the later items, due to believing that other similar items had already

been reported. Thus, although multiple, behaviorally specific items are preferable to single items, at a certain point, adding more items may not necessarily increase overall reporting.

Sexual Strategies Scale

Several other measures include sexually aggressive tactics similar to those included on the SES (e.g., Camilleri, Quinsey, & Tapscott, 2009; Mosher & Anderson, 1986; Shackelford & Goetz, 2004; Struckman-Johnson et al, 2003). However, in contrast to the SES, these measures assess only tactics or strategies used to obtain sex, without regard to the outcome (e.g., whether intercourse was completed or not). This simpler item structure may make items easier to interpret compared to the compound structure of the SES items. In addition, a criticism of the SES perpetration items is that they are worded from the female victim's perspective and responding positively to an item requires knowledge and understanding that a perpetrator is unlikely to possess (i.e., that she felt overwhelmed, see Kolivas & Gross, 2007). Hence, lower rates of perpetration relative to victimization may reflect differences in the perspectives of male perpetrators and female victims. In contrast, tactic-based measures require only that the man recall behaviors that he used in an attempt to have sex. Finally, assessment of sexually aggressive tactics independent of outcome may yield a purer, more direct means of assessing underlying perpetration severity, since the outcome (e.g., whether rape was completed or attempted) may be influenced by victim resistance or bystander intervention.

In a direct comparison of the two measurement approaches, Strang et al., (2013) compared the performance of the SES to the Sexual Strategies Scale (SSS), modified slightly from Struckman-Johnson et al's (2003) Postrefusal Sexual Persistence Scale. The SSS includes 22 items, presented in non-hierarchical order of severity, that describe a range of tactics used to obtain sex from an unwilling target. Tactics are presumed to represent the following hierarchy: enticement, verbal coercion (originally called emotional manipulation), exploitation of the intoxicated, and physical force (Struckman-Johnson et al 2003). In Study 1, Strang et al compared the SSS to a version of the SES by Peterson, Janssen, and Heiman (2010) and in Study 2, to the Koss et al (2007) SES. In both studies, men completed both measures. Despite similarity in content between the two measures, there were substantial discrepancies in responses, with few reporting aggression on both. The SSS yielded higher rates of all types of sexual aggression (e.g., verbal, intoxication, physical), with more men responding positively to SSS items assessing each type of aggression than to the corresponding SES items. The higher rate using the tactics-focused SSS is consistent with Abbey et al's (2005) suggestion that memory for perpetration is better cued by tactic than by outcome. Higher SSS reports may also reflect the fact that its items are ordered in a non-hierarchical pattern as opposed to the SES practice in which items are ordered according to increasing severity. Non-hierarchical ordering of items has been shown to result in higher reports of physical aggression compared to ordering items according to increasing severity (Ramirez & Straus, 2006).

The present study

We administered a version of the SES (Abbey et al, 2007) to a large cohort of college freshmen males and the SSS (Strang et al, 2013) to a second, equivalent cohort from the same university, using otherwise identical methodology. The between-participant design permitted comparison of rates of sexual perpetration over the first semester of college using two measures that assess similar content but with different items and structure. Both Abbey et al (2005) and Strang et al (2013) caution that the inclusion of too many similar items may result in respondent fatigue and underreporting, as respondents believe that they have already reported on a behavior and fail to respond positively to a subsequent, similar item. Hence, in contrast to Strang et al's (2013) within-participant design, we used a between-participant design to avoid any contamination that might result from completing two largely equivalent measures assessing a socially proscribed behavior. We had no a priori reason to suspect that actual rates of sexual aggression perpetration would differ between the two cohorts of freshmen, thus, differences in self-reported rates are assumed to reflect differences between the two measures in their ability to cue responses and elicit reports. Thus, the present study functioned as a conceptual replication of Strang et al (2013) using a larger, unselected sample and a between-participant design.

We compared rates of any sexual aggression and of three specific types of sexual aggression tactics assessed by both measures: verbal persuasion, victim intoxication, and physical force or threats of force. The SSS also includes enticement tactics, which are non-coercive in the context of an initial sexual advance but become coercive only following refusal (Struckman-Johnson et al, 2003). Although it is unclear whether these behaviors constitute sexual aggression, we speculated that their inclusion might help to normalize the more severe items and increase willingness to report on the SSS. Because the SSS contains more items and more possible tactics than the SES, we expected overall rates of perpetration to be higher on the SSS. However, like Strang et al (2013), we restricted direct comparison between the two scales to items with equivalent meaning. We expected that by presenting items in non-hierarchical order, in the context of less severe items, the SSS would normalize reporting of perpetration, resulting in higher reports of sexual aggression for equivalent items compared with the SES (Hypothesis 1).

A second goal of the paper was to assess the psychometric structures of the two measures. Internal consistency, usually assessed by Cronbach's coefficient α , is a first, but inadequate step toward establishing that the items represent a single underlying construct. However, there has been little psychometric examination of the hierarchy of severity presumed to underlie these measures, a second and important step. Sexual aggression severity may be conceptualized as involving two orthogonal dimensions: tactic (ranging from verbal coercion to physical force) and an outcome (ranging from contact to intercourse, see DeGue & DiLillo, 2005). Following the tradition begun by Koss et al (1987), researchers using the SES have typically assumed a hierarchy of severity that combines tactic and outcome and classified perpetrators according to their most severe act, ranging from unwanted contact (regardless of tactic) to verbally coerced intercourse to attempted rape (using intoxication or force) and rape (using intoxication or force, e.g., White & Smith, 2004). Indirect support for such a hierarchy is provided by studies showing that rapists have the highest levels of

various risk factors compared to non-perpetrators, with men who engaged in contact or coercion (but not rape) falling in between (e.g., Abbey et al, 2007). In contrast to the SES hierarchy which combines tactic and outcome, the SSS items are presumed to represent increasingly severe tactics: enticement, verbal coercion, victim intoxication, and force or threat of force (Struckman-Johnson et al, 2003). However, this presumed hierarchy of severity has not been tested explicitly.

Rasch item analysis (Bond & Fox, 2001) is a statistical measurement model that assesses the extent to which a set of binary items presumed to measure the same construct (in this case, sexual perpetration) fits a single underlying continuum with items varying only by difficulty, or in this case, by perpetration severity. The model yields an estimate of each item's severity based on the assumption that more frequently endorsed items represent less severe manifestations of the construct than items less frequently endorsed. The Rasch model also estimates a subject's intensity on the latent trait (i.e., perpetration), based on the assumption that if a person endorses an item representing a specific level of severity, then that person will probably endorse all items representing lesser severity as well (e.g., a man who has perpetrated forcible rape has also perpetrated less severe behaviors, such as verbal coercion, as well, a pattern reported by DeGue, DiLillo & Scalora, 2010).

Rasch analysis would appear to be an appropriate method of testing the assumption that sexual perpetration behaviors represent a unidimensional continuum of severity, ranging from verbal enticement to forcible rape. However, to our knowledge, Rasch analysis has been applied just once to understanding assessment of sexual aggression. Combining male perpetrators' and female victims' responses from the original Koss et al (1987) dataset, Rasch analysis revealed that SES items formed a single dimension and that items generally arrayed in the presumed SES hierarchy from contact to completed rape (Karabatsos, 1997). There was one exception: contact using force appeared more similar in severity to the forced intercourse items than to other contact items, suggesting that tactic may be a more important aspect of severity than outcome. In the current study, we performed Rasch measurement model analysis separately on the SES (Abbey et al 2007) and on the SSS (Strang et al, 2013). The analysis allowed us to determine the extent to which items represent a single, underlying continuum, and to assess whether item severity reflects a tactic-based (SES and SSS) or tactic+outcome-based (SES only) hierarchy.

Method

Participants

Participants included 2037 freshman males, ages 18 and 19, entering a large Northeastern public university in the fall of 2011 (N = 994) or fall of 2012 (N = 1043). Sample composition was 72.6% White, 15.1 % Asian, 5.3 % African American, and 7.0% mixed, other, or unknown. Most (66%) lived on campus. There were no significant differences between cohorts in ethnicity or residence. However, study participants included more White students and fewer campus residents compared with University freshmen as a whole (63% White, 77% live on campus).

Procedure

Data were derived from the baseline survey of a larger longitudinal study. Following an introductory letter, email invitations were sent in mid-November seeking participants for a federally funded study of student behaviors and attitudes. All first-semester freshmen males who resided in the United States, had given permission for their directory information to be used (about 85% of the class), and were 18 or 19 years of age on this date were invited. Students who did not respond to the initial invitation were sent up to 5 email reminders over the next 4 weeks, as well as an additional letter sent over Thanksgiving to home addresses. Response rate was 67.3%. Survey invitations included a link to the secure website where students were asked to enter their student ID number. After providing informed consent, men were directed to the survey. They were compensated \$25 in Campus Cash for completion. All procedures were approved by the (masked) Institutional Review Board.

Measures

The survey, which took 20 to 30 minutes to complete, included measures of personality, attitudes, sexual behavior, alcohol use, and sexual aggression. Assessments were identical for the two cohorts except for the measures of sexual aggression which are described below.

Sexual Experiences Survey (SES)—Cohort 1 participants completed a 16-item version of the Sexual Experiences Survey (SES, Abbey et al., 2007). The measure uses the original SES format of presenting a sexual outcome (e.g., contact, attempted rape; oral sex, intercourse) followed by a tactic used to obtain that outcome (e.g., continual arguments, giving her alcohol). Outcomes are described quite explicitly (e.g., “The following questions are about sexual intercourse. By sexual intercourse we mean penetration of a woman, no matter how slight, by a man’s penis. Ejaculation is not required”). The SES includes 3 contact items (overwhelming with continual arguments and pressure; showing your displeasure; threatening or using force), 3 attempted intercourse items (giving her alcohol/drugs; when she was passed out or too intoxicated to give consent, threatening or using force), 5 oral sex items, and 5 completed intercourse items. The oral sex and intercourse stems are followed by all 5 possible tactics (arguments; displeasure, give her alcohol; passed out, force). Respondents are asked to respond as to the frequency with which they engaged in each behavior over the past semester: 0 (*Never*), 1 (*Once*), or 2 (*Two or more times*). The Flesch-Kincaid reading grade equivalent of the measure was 11.3 (Kincaid, Fishburne, Rogers, & Chissom, 1975).

Sexual Strategies Scale (SSS)—Cohort 2 participants completed the 22-item Sexual Strategies Scale (SSS, Strang et al, 2013), based upon Struckman-Johnson et al. (2003). The introduction states: “Since the semester began, which of the following have you used to convince a woman to have sex (oral, anal, vaginal) with you when she didn’t want to.” This wording was modified slightly from the original which specified that the tactics were used “after the other person had said no,” since this seemed overly restrictive for the more severe tactics and inconsistent with the SES. Respondents indicated whether they had used each of the tactics that followed. Although presented in non-hierarchical order on the scale, tactics represent four increasingly severe types: enticement (3 items), emotional manipulation (equivalent to verbal, 8 items), intoxication (3 items) and force (6 items). SSS tactics overlap

considerably with those presented in the SES with the exception of enticement items (“continue to kiss and touch”, “take off your clothes”, “take off her clothes”), which have no equivalent on the SES. The Flesch-Kincaid reading level (Kincaid et al, 1975) of the measure was 5.3.

Results

Prevalence of Sexual Aggression Perpetration

The SES response options permit respondents to indicate whether each item had occurred never, once, or two or more times. Because most (94%) responded *never* to all items and very few reported 2 or more occasions for any item (0 to 3 men), we collapsed *once* and 2 or more, consistent with Karabatsos (1997). Of the 59 men who responded positively to one or more items, most endorsed 1 item ($n = 20$) or 2 items ($n = 9$); however, 9 men responded positively to all 16 items. Perpetrating every item within a single semester seemed implausible and suggested frivolous responding. Because these 9 men also showed other anomalies in their data (e.g., implausible numbers of sexual partners), they were not included in data analyses or in Table 1. We also removed 2 men who refused to answer all 16 items and 3 others who answered some but not all items (all negatively), leaving 980 cases. Prevalence was 5.1% for any SES perpetration in the past semester. To facilitate comparison with the SSS, which assesses tactics without regard to outcome, prevalence data are presented in Table 1 according to tactic. Using the traditional SES severity scoring method to classify perpetrators (e.g., White & Smith, 2004), 13 (1.3%) reported contact as the most severe outcome, 8 (0.8%) reported verbally coerced intercourse or oral sex, 13 (1.3%) reported attempted rape, and 16 (1.6%) reported rape.

Of 1043 SSS respondents, one man reported all 22 items, suggesting frivolous responding. He also reported identical answers on other questionnaires. After dropping this case, overall prevalence for any type of perpetration on the SSS was 19.5%. As shown in Table 2, the highest rates on the SSS were for use of enticement strategies (16.5%), which are not assessed by the SES. After excluding enticement, 11.8% reported some type of sexual aggression on the SSS. As expected, overall prevalence was higher for SSS than SES; however, this may reflect the greater number of items and tactics in the SSS that have no equivalent in the SES.

We hypothesized that the inclusion of items representing lower severity behaviors on the SSS would normalize reporting of aggression and result in higher rates of self-reported perpetration compared with the SES even on equivalent items (Hypothesis 1). The SSS includes separate items for some tactics (giving her alcohol; slipping her drugs) which are included in a single item on the SES (giving her alcohol or drugs), although the SES asks about the same tactic with different outcomes (e.g., giving her alcohol or drugs to have oral sex versus giving her alcohol drugs for sexual intercourse). We considered the intoxication and force items presented in Tables 1 and 2 to be equivalent across measures, in that they contain the same content.

As shown in the Tables, endorsement of intoxication as a tactic was virtually identical across measures (3.0% reporting this strategy on the SES, 3.4% on the SSS). Similarly,

endorsement of physical threats or force was the same for the SES (1.1%) and the SSS (1.0%). Thus, based on intoxication and force strategies, we failed to find support for the hypothesis that the SSS would yield higher reports. The SSS includes one emotional manipulation tactic (“Telling her lies”) that is not included in the SES verbal items. For equivalence when comparing verbal tactics across measures, we dropped this item from the SSS (seven men endorsed only telling lies and no other item). Even without this item, more men reported verbal tactics on the SSS (7.8%) than on the SES (3.7%), $Z = 3.96, p < .001$, providing support for Hypothesis 1.

The SSS includes enticement tactics and other items that are not included on the SES (shown in italics in Table 2). One of these, “use of older age to convince her”, was endorsed by 31 men (3%), including 7 who endorsed only this item, a surprisingly high rate of endorsement given that participants were only 18 or 19 years old. The SSS item “Using your authority” was included on earlier SES versions but was very infrequently endorsed (Karabatsos, 1997) and subsequently dropped from later versions (e.g., Koss et al, 2007). This SSS item was endorsed only twice and these men also endorsed other SSS items.

Rasch item analysis

A Rasch item analysis yields estimates of item severity and their standard errors. It also yields a likelihood deviance statistic, indicative of global goodness of fit, or the extent to which response patterns across participants conform to the expected single underlying continuum. Rasch analysis also provides infit and outfit statistics which indicate how well each item conforms to a unidimensional model. Large infit and outfit values indicate a tendency to endorse an item while not endorsing less severe ones or vice versa. Infit statistics are more sensitive to misfitting responses to items near the person’s severity level on the latent continuum whereas outfit statistics are more sensitive to misfitting responses to items further away (Piquero, MacIntosh, & Hickman, 2001). Typically, items with infit and outfit values less than 2.0 are considered acceptable. Values exceeding 2.0 suggest that the item does not conform to the underlying ordered continuum and should be eliminated (Wright & Masters, 1982). Items with infit or outfit values between 1.5 and 2.0 or less than .5 are inefficient but do not degrade the scale (Linacre, 2013). Analyses were conducted in R (R Core Team, 2013) and Rstudio (Rstudio Inc, 2013) using the eRM package (Mair, Hatzinger, & Maier, 2012).

Rasch Results for the SES

Based on 980 cases with valid and complete data on the SES, Cronbach’s Coefficient α (Kuder Richardson-20) was .89, ranging from .88 to .90 with individual items excluded, nearly identical to the α of .88 reported by Abbey et al (2007). Because respondents who endorse no items ($n = 930$) are uninformative for Rasch analyses, they were eliminated, leaving 50 cases. Table 3 displays the results of the Rasch model for the SES with items displayed according to decreasing severity, i.e., higher frequency of endorsement. Standard errors are relatively large, reflecting the small sample size, and range of severity was fairly narrow (–1 to +2). Although the severity continuum does not map on exactly to the traditional SES severity scoring (e.g., White & Smith, 2004), the hierarchy makes reasonable sense. Physically forced oral and vaginal sex were at the high end of the severity

continuum and contact using arguments or displeasure at the low end; however, attempted rape with alcohol was found to be the least severe (i.e., most commonly endorsed) item. The likelihood deviance statistic was significant, suggesting poor global fit with an assumed unidimensional continuum, $\chi^2(176) = 212.3, p = 0.032$. As shown in Table 3, item fit statistics were acceptable for all items with the possible exception of item 2 (contact by showing displeasure). Four items (11, 16, 15, and 6) showed outfit statistics less than .5 but their corresponding infit statistics were satisfactory, indicating that these items need not be removed from the scale. In brief, the SES showed less than ideal Rasch properties.

Rasch Results for the SSS

There were 203 cases with complete and valid data on the SSS. Coefficient α (KR-20) for the 22 items was .79, ranging from .76 to .79 for individual items excluded. Because Item 11 (using a weapon) was not endorsed by any respondent, it was uninformative and thus eliminated from the Rasch analysis. Table 4 presents the item response analysis for the 21 SSS items, with items displayed along a continuum of threshold levels ranging from high to low severity. As expected, enticement strategies - continuing to touch her, taking off her clothes, taking off your clothes - emerged at the low end of the severity continuum, whereas physical harm items were at the high end. However, the high severity thresholds associated with some items (e.g., use of authority, threaten to break up) were unexpected. The severity estimates reveal a broader range of severity than the SES, ranging from very mild (continue to kiss and touch) at -4 to quite severe at +2. All of the items showed infit and outfit statistics that were within or just outside the recommended range. Items 18 and 22, however, with small infit and very small outfit values may not be contributing any information regarding the scale. Several items (17, 14, 8, 10, and 5) showed small outfit values, but because their corresponding infit values were satisfactory, they need not be removed from the scale. The likelihood deviance statistic, $\chi^2(231) = 232.8, p = 0.45$, indicated satisfactory global goodness of fit. In sum, the SSS showed satisfactory Rasch properties, reflecting an underlying unidimensional continuum of perpetration severity.

Discussion

The present study, by comparing the performance of two self-report measures, adds to our limited understanding of the impact of measurement on our ability to obtain accurate estimates of sexual aggression perpetration. We hypothesized that the SSS (Strang et al, 2013) would result in higher rates of self-reported perpetration than the SES (Abbey et al, 2007) due to its focus on tactics and its greater number of more simply phrased items, presented in non-hierarchical order of presumed severity. We found mixed support for this hypothesis. Endorsement of less severe tactics, and hence, overall rate of self-reported perpetration, was higher for the SSS, even when we limited comparison to tactics that were assessed by both measures. However, these higher rates of perpetration reflected greater willingness to endorse the less severe, verbal tactics. The simplest explanation for this advantage is that the SSS separated tactics that are combined in a single SES item (e.g., “by showing your displeasure - sulking, making her feel guilty, swearing, getting angry or threatening to end the relationship”) into separate items, allowing respondents to indicate that they had engaged in more than one strategy. Despite substantial differences in item

wording and presentation between the two scales, the proportion who endorsed use of intoxication and force tactics on the SES and the SSS was virtually identical. Thus, we failed to find support for the hypothesis that including more items of lesser severity, in non-hierarchical order, would normalize reporting and yield higher rates of endorsement for more severe types of aggression. However, the consistency in the rates of self-reported force and intoxication strategies (1% and 3%, respectively) across very different measures is reassuring, and suggests that these rates, while low, reflect the true rates of these types of perpetration rather than measurement effects.

The study also provides unique psychometric information on the two measures. Although an underlying hierarchy of increasingly severe sexual aggression has been presumed, dating back to Koss et al (1987), the current study provides unique evidence in support of such a hierarchy. That is, men who perpetrate more severe types of sexual aggression have perpetrated less severe types as well. Traditional severity classification using the SES is based on a combination of tactic and outcome: contact (regardless of tactic) is least severe, followed by verbally coerced intercourse. Rasch analysis provides some support for this hierarchy in that use of force and taking advantage of incapacitation to obtain vaginal or oral sex emerged as the most severe items. However, replicating the findings of Karabatsos' (1997) Rasch analysis, we also found that contact using force was more severe than contact and attempted rape using verbal or intoxication tactics, providing some support for a hierarchy based on tactic rather than outcome. Attempted intercourse by giving alcohol or drugs emerged as the least severe item, although based on tactic (intoxication) or outcome (attempted rape), this behavior should fall in the middle of the hierarchy. It seems plausible, as suggested by Ouimette et al (2000), that this item may be interpreted more broadly than intended by the scale developers (e.g., as using alcohol as a means of encouraging a woman to consent).

Although Rasch analysis provided some empirical evidence for a unidimensional structure with reasonable severity hierarchy for both scales, the SSS conformed better than the SES to a unidimensional structure, with better global fit and no ill-fitting items. In support of the tactics-based hierarchy that is presumed by the SSS, physical force items clustered toward the severe end of the severity hierarchy. However, verbal/emotional manipulation items were disbursed within the hierarchy, with some of the items at the severe end seeming misplaced. Rasch modeling uses frequency of endorsement to determine severity and it appears that some items that were infrequently endorsed (e.g., threaten to break up, use authority) are not necessarily severe but rather are just rarely used by college men as tactics to obtain sex. Importantly, Rasch analysis also provided support for the expansion of the perpetration dimension to include less severe items. Enticement tactics, as intended, represented the low end of severity. Yet, even the item "continue to kiss and touch", which had the lowest severity threshold of any SSS item (endorsed by 16.5% of men), fit on the same dimension as more severe forms of perpetration. This finding corroborates results of Camilleri et al (2009), who found that a similar item loaded on a coercion factor rather than on a more benign coaxing factor.

Is There an Advantage for One Measure Over the Other?

Both the SES and SSS function reasonably well and our results do not provide clear support for one measure over the other over this short timeframe. However, we believe that our findings suggest an advantage for the SSS. First, the Rasch analysis revealed better global fit for the SSS compared with the SES and no ill-fitting items. Moreover, the SES does a poorer job in assessing the less severe end of the continuum relative to the SSS. Significantly fewer men endorsed verbal tactics on the SES compared with the SSS. Inclusion of additional verbal tactic items (e.g., telling lies) or separating tactics into separate items would be a reasonable way of improving the SES. Finally, there was more frivolous responding and more missing data on the SES compared with the SSS. Nine men reported that they had engaged in all 16 forms of perpetration in the past semester. Although this number is not high, it represented a significant proportion of all positive responses on the SES and including such invalid data could conceivably distort research conclusions (see Cornell, Klein, Konold, & Huang, 2012). Researchers considering the SES should consider the potential for poorer quality responses, particularly when assessing sexual aggression in the context of a web-based survey that includes several other measures. Finally, as discussed above, SES items are considerably longer and more complex in their vocabulary and structure, resulting in increased Flesch-Kincaid reading level and cognitive burden. Hence, the SSS may be preferred particularly for non-college samples with lower reading levels.

Limitations

Although we began with a large sample of college freshmen, we assessed sexual aggression for the first semester of college only, resulting in low rates of perpetration. Consequently, the modest number of cases included in the Rasch analyses may have contributed to the less than conclusive results regarding the best way to characterize dimension severity. It is important that analyses be replicated using larger samples and perpetration over a longer period of time. The two measures were given separately to entire cohorts rather than randomly assigning individuals within the two cohorts to the SES or SSS. Although we found no significant demographic differences between cohorts and have no reason to believe that systematic differences would have accounted for the pattern of findings, we cannot rule out effects of history. Moreover, findings may be specific to college freshmen and should not be generalized to other samples of men. The use, and hence the Rasch severity of some tactics, may be shaped by the environment or age of the sample. For example, alcohol may be a common sexual coercion tactic due to its availability and ubiquity in the college culture, whereas college freshmen are unlikely to use positions of authority. It is important to replicate findings with other samples of perpetrators.

Research Implications

Improving measures of sexual aggression perpetration and their psychometric properties requires additional research to understand how respondents interpret the items and what types of experiences they are cueing (e.g., Ouimette et al, 2000; Ross & Allgeier, 1996; Struckman-Johnson et al, 2003). For example, although a potential advantage of the SES is that it includes information about outcomes, it has not yet been established whether men are sensitive to the different outcomes contained in the SES questions. For example, the

phrasing of the SES contact items implies that sexual contact is the goal; however, we suspect that these items are actually measuring unsuccessful attempts at intercourse. In previous research we found that independent coders had a difficult time distinguishing victims' descriptions of SES contact experiences from descriptions of SES attempted rape experiences (Testa, VanZile-Tamsen, Livingston, & Koss, 2004). Similar research using men's descriptions would provide insight into how men are interpreting the items (e.g., Ross & Allgeier, 1996).

Perhaps more important in our quest to improve measurement of this elusive construct is more careful consideration of what it is we seek to measure. The concern that perpetration is underreported by men relative to reports of victimization by women (Spitzberg, 1999) has led to the assumption that higher rates of self-reported perpetration are the ideal, reflecting more truthful reporting. Yet, equal rates of self-reported perpetration and victimization may be an unattainable and perhaps not entirely desirable goal given the inherent differences in the perspectives of perpetrators and victims (see Kolivas & Gross, 2007). Expanding the range of lower level items as indicators of sexual aggression perpetration may be warranted from psychometric standpoint and helpful for research applications in which assessment of a broad tendency toward sexual aggression is desired. Behavioral validation of self-report measures (e.g., examining whether men who report perpetration display behavior consistent with sexual aggression in a laboratory analog situation, Bernat, Calhoun, & Adams, 1999) may help to bolster convergent and hence construct validity. However, although psychometric and empirical research can provide insight in improving measurement, it is not a substitute for conceptual and definitional clarity, which has been lacking in the area of sexual aggression assessment (see DeGue & DiLillo, 2005).

Clinical and Prevention Implications

Despite years of research and prevention efforts, sexual victimization remains a prevalent (Krebs, Lindquist, Warner, Fisher, & Martin, 2009) and significant (Griffin, Wardell, & Read, 2013) problem among college women. Research on sexual aggression perpetration, and particularly measurement of perpetration, has lagged behind research on sexual victimization. Better measurement of sexual aggression perpetration is critical to these efforts since only with valid measurement of sexual aggression perpetration can we determine whether sexual aggression prevention efforts are effective (Schewe & O'Donohue, 1993). Rasch analysis of the SSS suggests that even enticement behaviors fit on the same dimension as more severe forms of perpetration. Although more research is needed to understand this novel finding, the ability to identify men with a tendency toward perpetration before they progress to more serious types of offenses has potentially important clinical implications.

Acknowledgments

This research was supported by R01AA19478 from the National Institute on Alcohol Abuse and Alcoholism to Maria Testa.

References

- Abbey A, Parkhill MR, Clinton-Sherrod AM, Zawacki T. A comparison of men who committed different types of sexual assault in a community sample. *Journal of Interpersonal Violence*. 2007; 22:1567–1580.10.1177/0886260507306489 [PubMed: 17993642]
- Abbey A, Parkhill MR, Koss MP. The effects of frame of reference on responses to questions about sexual assault victimization and perpetration. *Psychology of Women Quarterly*. 2005; 29:364–373.10.1111/j.1471-6402.2005.00236.x
- Bernat JA, Calhoun KS, Adams HE. Sexually aggressive and nonaggressive men: Sexual arousal and judgments in response to acquaintance rape and consensual analogues. *Journal of Abnormal Psychology*. 1999; 108:662–673.10.1037/0021-843X.108.4.662 [PubMed: 10609430]
- Bond, TG.; Fox, CM. *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum; 2001.
- Camilleri JA, Quinsey VL, Tapscott JL. Assessing the propensity for sexual coaxing and coercion in relationships: Factor structure, reliability, and validity of the Tactics to Obtain Sex Scale. *Archives of Sexual Behavior*. 2009; 38:959–973.10.1007/s10508-008-9377-2 [PubMed: 18607709]
- Cook SL. Self-reports of sexual, physical, and nonphysical abuse perpetration. *Violence Against Women*. 2002; 8:541–565.10.1177/107780102400388425
- Cornell D, Klein J, Konold T, Huang F. Effects of validity screening items on adolescent survey data. *Psychological Assessment*. 2012; 24:21–35.10.1037/a0024824 [PubMed: 21823800]
- DeGue S, DiLillo D. ‘You would if you loved me’: Toward an improved conceptual and etiological understanding of nonphysical male sexual coercion. *Aggression and Violent Behavior*. 2005; 10:513–532.10.1016/j.avb.2004.09.001
- DeGue S, DiLillo D, Scalora M. Are all perpetrators alike? Comparing risk factors for sexual coercion and aggression. *Sexual Abuse: A Journal of Research and Treatment*. 2010; 22:402–426.10.1177/1079063210372140 [PubMed: 20693517]
- Griffin MJ, Wardell JD, Read JP. Recent sexual victimization and drinking behavior in newly matriculated college students: A Latent Growth Analysis. *Psychology of Addictive Behaviors*. 2013; 27(4):966–973.10.1037/a0031831 [PubMed: 23528195]
- Karabatsos G. The Sexual Experiences Survey: Interpretation and validity. *Journal of Outcome Measurement*. 1997; 1:305–328. [PubMed: 9661726]
- Kincaid, JP.; Fishburne, RP.; Rogers, RL.; Chissom, BS. Research Branch Report 8–75. Chief of Naval Technical Training; Naval Air Station, Memphis, TN: 1975. Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula) for Navy Enlisted Personnel.
- Kolivas ED, Gross AM. Assessing sexual aggression: Addressing the gap between rape victimization and perpetration prevalence rates. *Aggression and Violent Behavior*. 2007; 12:315–328.10.1016/j.avb.2006.10.002
- Koss MP. Detecting the scope of rape: A review of prevalence research methods. *Journal of Interpersonal Violence*. 1993; 8:198–222.10.1177/088626093008002004
- Koss MP, Abbey A, Campbell R, Cook S, Norris J, Testa M, et al. Revising the SES: A collaborative process to improve assessment of sexual aggression and victimization. *Psychology of Women Quarterly*. 2007; 31:357–370.10.1111/j.1471-6402.2007.00385.x
- Koss MP, Gidycz CA, Wisniewski N. 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*. 1987; 55:162–170.10.1037//0022-006x.55.2.162 [PubMed: 3494755]
- Koss MP, Oros CJ. Sexual Experiences Survey: A research instrument investigating sexual aggression and victimization. *Journal of Consulting and Clinical Psychology*. 1982; 50:455–457.10.1037//0022-006x.50.3.455 [PubMed: 7096751]
- Krebs CP, Lindquist CH, Warner TD, Fisher BS, Martin SL. College women’s experiences with physically forced, alcohol- or other drug-enabled, and drug-facilitated sexual assault before and since entering college. *Journal of American College Health*. 2009; 57:639–647.10.3200/jach.57.6.639-649 [PubMed: 19433402]

- Linacre, JM. User's guide to Winsteps and Ministep Rasch-Model computer programs. 2013. Available at <http://www.winsteps.com/winman/index.htm?copyright.htm>
- Mair P, Hatzinger R, Maier MJ. eRm: Extended Rasch Modeling, R package version 0.15–1. 2012
- Mosher DL, Anderson RD. Macho personality, sexual aggression, and reactions to guided imagery of realistic rape. *Journal of Research in Personality*. 1986; 20:77–94.10.1016/0092-6566(86)90111-x
- Quimette PC, Shaw J, Drozd JF, Leader J. Consistency of reports of rape behaviors among nonincarcerated men. *Psychology of Men and Masculinity*. 2000; 1:133–139.10.1037/1524-9220.1.2.133
- Peterson ZD, Janssen E, Heiman JR. The association between sexual aggression and HIV risk behavior in heterosexual men. *Journal of Interpersonal Violence*. 2010; 25:538–556.10.1177/0886260509334414 [PubMed: 19474034]
- Piquero AR, MacIntosh R, Kickman M. Applying Rasch modeling to the validity of a control balance scale. *Journal of Criminal Justice*. 2001; 29:493–505.10.1016/S0047-2352(01)00112-X
- R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. Available at <http://www.r-project.org/>
- Ramirez IL, Straus MA. The effect of question order on disclosure of intimate partner violence: An experimental test using the Conflict Tactics Scales. *Journal of Family Violence*. 2006; 21:1–9.10.1007/s10896-005-9000-4
- Rasch, G. An individualistic approach to item analysis. In: Lazarsfeld, PF.; Henry, NW., editors. *Readings in mathematical social science*. Chicago, IL: Chicago Science Research Associates; 1966. p. 89-108.
- Ross RR, Allgeier ER. Behind the pencil/paper measurement of sexual coercion: Interview-based clarification of men's interpretations of Sexual Experiences Survey items. *Journal of Applied Social Psychology*. 1996; 26:1587–1616.10.1111/j.1559-1816.1996.tb00087.x
- RStudio, Inc. RStudio; Boston, MA: 2013. Available at <http://www.rstudio.com/>
- Schewe P, O'Donohue W. Rape prevention: Methodological problems and new directions. *Clinical Psychology Review*. 1993; 13:667–682.10.1016/0272-7358(93)90032-h
- Shackelford TK, Goetz AT. Men's sexual coercion in intimate relationships: Development and initial validation of the sexual coercion in intimate relationships scale. *Violence and Victims*. 2004; 19:541–556.10.1891/vivi.19.5.541.63681 [PubMed: 15844724]
- Spitzberg BH. An analysis of empirical estimates of sexual aggression victimization and perpetration. *Violence and Victims*. 1999; 14:241–260. [PubMed: 10606432]
- Strang ET, Peterson ZD, Hill YN, Heiman JR. Discrepant responding across self-report measures of men's coercive and aggressive sexual strategies. *Journal of Sex Research*. 2013; 50:458–469.10.1080/00224499.2011.646393 [PubMed: 22329465]
- Struckman-Johnson C, Struckman-Johnson DL, Anderson PB. Tactics of sexual coercion: When men and women won't take no for an answer. *Journal of Sex Research*. 2003; 40:76–86.10.1080/00224490309552168 [PubMed: 12806533]
- Testa M, VanZile-Tamsen C, Livingston JA, Koss MP. Assessing women's experiences of sexual aggression using the Sexual Experiences Survey: Evidence for validity and implications for research. *Psychology of Women Quarterly*. 2004; 28:256–265.10.1111/j.1471-6402.2004.00143.x
- White JW, Smith PH. Sexual assault perpetration and reperpetration: From adolescence to young adulthood. *Criminal Justice and Behavior*. 2004; 31:182–202.10.1177/0093854803261342
- Wright, BD.; Masters, GN. Rating scale analysis. Chicago, IL: Mesa Press; 1982.

Table 1

Prevalence of Perpetration of Sexual Aggression in the Past Semester Based on the Sexual Experiences Survey, Perpetrator Version (SES), Cohort 1, T1 (N = 980)

Sexual Experiences Survey (SES) Item	<i>n</i>	%
<i>Over the past semester, how many times (never, once, two or more times) have you:</i>		
Attempted sexual intercourse by with a woman giving her alcohol or drugs to make her intoxicated	21	2.1
Had oral sex by giving her alcohol or drugs to make her intoxicated	7	0.7
Had sexual intercourse by giving her alcohol or drugs to make her intox	11	1.1
Attempted sexual intercourse with a woman who was passed out or too intoxicated to give consent or stop what was happening	12	1.2
Had oral sex when she was passed out/too intoxicated to give consent	7	0.7
Had intercourse when she was passed out/too intoxicated to give consent	6	0.6
Total reporting use of intoxication as a strategy for sexual perpetration	29	3.0
Had sexual contact by threatening or using some degree of physical force	8	0.8
Attempted sexual intercourse by threatening or using some degree of physical force	7	0.7
Had oral sex by threatening or using some degree of physical force	2	0.2
Had sexual intercourse by threatening or using some degree of physical force	5	0.5
Total reporting use of physical force as a strategy for sexual perpetration	11	1.1
Had sexual contact by overwhelming her with continual arguments/pressure	18	1.8
Had sexual contact by showing your displeasure	21	2.1
Had oral sex by overwhelming her with continual arguments and pressure	9	0.9
Had oral sex by showing your displeasure	12	1.2
Had sexual intercourse by overwhelming her continual arguments/pressure	8	0.8
Had sexual intercourse by showing your displeasure	12	1.2
Total reporting use of verbal coercion as a strategy for sexual perpetration	36	3.7
TOTAL REPORTING ANY SEXUAL PERPETRATION	50	5.1

Table 2

Prevalence of Perpetration of Sexual Aggression in the Past Semester Based on the Sexual Strategies Scale (SSS), Cohort 2, T1 (N = 1,042)

Sexual Strategies Scale (SSS) Item	<i>n</i>	%
<i>Since the semester began which of the following have you used to convince a woman to have sex (oral/anal/vaginal) with you when she didn't want to:</i>		
Getting her drunk/high	33	3.2
Slipping her drugs so that you can take advantage of her	2	0.2
Taking advantage of her being drunk/high and unable to stop what is happening	7	0.7
Total reporting use of intoxication	35	3.4
Blocking her if she tries to leave the room	5	0.5
Threatening to harm her physically if she doesn't have sex	2	0.2
Using a weapon to frighten her into having sex	0	0.0
Using physical restraint	4	0.4
Harming her physically	2	0.2
Tying her up	4	0.4
Total reporting use of physical force	10	1.0
<i>Telling her lies (e.g. saying "I love you" when you don't)</i>	38	3.6
Asking her repeatedly to have sex	34	3.3
Threatening to break up with her if she doesn't have sex	2	0.2
Threatening to harm yourself if she doesn't have sex	2	0.2
Threatening to tell others a secret or lie about her if she doesn't have sex	3	0.3
Questioning her sexuality (e.g. calling her a lesbian)	8	0.8
Questioning her commitment to the relationship	15	1.4
Accusing her of "leading you on" or being "a tease"	47	4.5
Total reporting use of verbal coercion	103	9.9
Total reporting verbal coercion using SES equivalent items	81	7.8
<i>Continuing to touch and kiss her in the hopes that she will give in to sex</i>	163	15.6
<i>Taking off her clothes in the hopes that she will give into sex</i>	42	4.0
<i>Taking off your clothes in the hopes that she will give in to sex</i>	44	4.2
Total reporting use of arousal/enticement for sexual perpetration	172	16.5
<i>Using your older age to convince her</i>	31	3.0
<i>Using your authority to convince her (e.g. if you were her boss, etc.)</i>	2	0.2
Total reporting use of older age/authority for sexual perpetration	32	3.1
TOTAL REPORTING ANY SEXUAL PERPETRATION	203	19.5
TOTAL REPORTING ANY SEXUAL PERPETRATION USING SES EQUIVALENT ITEMS	93	8.9

Note. Italicized items have no equivalent on the SES.

Table 3

Results of Rasch Modeling for the SES (Cohort 1, N = 50)

Item no.	Item content	% endorsed	Item severity	SE	Outfit	Infit
11	Oral force	0.2	2.10	0.73	0.11	0.56
16	Intercourse force	0.5	0.95	0.51	0.21	0.54
15	Intercourse incapacitated	0.6	0.69	0.48	0.31	0.70
09	Oral alcohol	0.7	0.46	0.45	1.06	1.35
06	Attempt force	0.7	0.46	0.45	0.49	0.87
10	Oral incapacitated	0.7	0.46	0.45	0.87	0.64
03	Contact force	0.8	0.25	0.43	0.56	0.88
12	Intercourse arguments	0.8	0.25	0.43	0.71	0.80
07	Oral arguments	0.9	0.06	0.41	0.84	1.06
14	Intercourse alcohol	1.1	-0.27	0.39	1.18	1.27
05	Attempt incapacitated	1.2	-0.43	0.38	0.59	0.74
08	Oral displeasure	1.2	-0.43	0.38	0.86	1.07
13	Intercourse displeasure	1.2	-0.43	0.38	0.61	0.83
01	Contact arguments	1.8	-1.17	0.33	1.05	1.03
02	Contact displeasure	2.1	-1.47	0.32	2.11	1.68
04	Attempt alcohol	2.1	-1.47	0.32	1.18	1.21

Table 4

Results of Rasch Modeling for the SSS (Cohort 2, N = 203)

Item no.	Item content	% endorsed	Item severity	SE	Outfit	Infit
17	Use authority	0.2	1.99	0.73	0.19	1.08
18	Harm physically	0.2	1.99	0.73	0.03	0.48
15	Threaten to break up	0.2	1.99	0.73	0.19	1.08
22	Give her drugs	0.2	1.99	0.73	0.03	0.48
08	Threaten harm	0.2	1.99	0.73	0.37	0.68
10	Threaten to harm self	0.2	1.99	0.73	0.26	0.96
05	Tell secret/lie	0.3	1.51	0.61	0.09	0.57
19	Tie her up	0.4	1.17	0.53	1.54	0.63
14	Physical restraint	0.4	1.17	0.53	0.19	0.76
07	Block her	0.5	0.90	0.48	1.50	0.88
09	Incapacitated	0.7	0.49	0.42	0.51	0.84
16	Question sexuality	0.8	0.32	0.39	0.73	1.09
20	Question commitment	1.4	-0.48	0.30	0.93	0.82
03	Older age	3.0	-1.46	0.24	1.17	1.05
04	Get her drunk	3.2	-1.55	0.23	0.79	0.85
06	Ask repeatedly	3.3	-1.60	0.23	0.86	1.09
02	Tell her lies	3.6	-1.76	0.23	1.15	1.08
12	Remove her clothes	4.0	-1.91	0.22	0.70	0.86
13	Remove your clothes	4.2	-1.98	0.22	0.61	0.80
21	Accuse of being a tease	4.5	-2.08	0.21	1.18	1.00
01	Continue to touch/kiss	15.6	-4.65	0.21	0.91	0.91