



6-2017

Initial Evidence for the Reliability and Validity of the Sexual Experiences Survey-Short Form Perpetration (SES-SFP) in College Men

RaeAnn E. Anderson
University of North Dakota, raeann.anderson@UND.edu

Shawn P. Cahill

Douglas L. Delahanty

[How does access to this work benefit you? Let us know!](#)

Follow this and additional works at: <https://commons.und.edu/psych-fac>



Part of the [Psychology Commons](#)

Recommended Citation

RaeAnn E. Anderson, Shawn P. Cahill, and Douglas L. Delahanty. "Initial Evidence for the Reliability and Validity of the Sexual Experiences Survey-Short Form Perpetration (SES-SFP) in College Men" (2017). *Psychology Faculty Publications*. 10.
<https://commons.und.edu/psych-fac/10>

This Article is brought to you for free and open access by the Department of Psychology at UND Scholarly Commons. It has been accepted for inclusion in Psychology Faculty Publications by an authorized administrator of UND Scholarly Commons. For more information, please contact und.common@library.und.edu.

Initial Evidence for the Reliability and Validity of the Sexual Experiences Survey-Short Form
Perpetration (SES-SFP) in College Men

Anderson, RaeAnn E*^{1,2}; Cahill, Shawn P²; Delahanty, Douglas L¹

First submitted: August 5th, 2016

Author note

¹ Kent State University ² University of Wisconsin-Milwaukee

Thank you to the participants who shared their experiences with us.

Portions of this data have been presented at the following conferences: International Family Violence Research Conference 2014, Milwaukee Violence Symposium 2014, Wisconsin Psychological Association 2013

*corresponding author: 144 Kent Hall, Kent, OH 44242 / rander52@kent.edu; 330-672-3158

Abstract

Introduction: Estimates of sexual perpetration in college men vary widely, partially due to a lack of reliable and valid measures of sexual perpetration. This study provides psychometric data on the Sexual Experiences Survey – Short Form Perpetration (SES-SFP), a revision of one of the most widely used measures of sexual perpetration, in a sample of college men.

Methods: Participants ($n = 402$) completed a web survey containing the study measures; a subset of 66 participants completed the SES-SFP again two weeks later.

Results: In examining test-retest reliability, most (90.7%) participants were classified correctly using dichotomous scores of sexually aggressive behavior (yes/no). However, test-retest agreement for category scores was poor (0 – 50.0%). Test-retest correlations were largest for lifetime category scores, $r = .69$ and smallest for lifetime dichotomous scores, $r = .59$. Regarding validity, SES-SFP scores were positively associated with measures of partner violence and trait aggression and negatively associated with a measure of rape empathy.

Conclusions: Our results provide initial evidence of internal consistency and convergent evidence of validity for the SES-SFP in college men but question the measure's test-retest reliability. The severity of tactics used to coerce sexually aggressive behaviors was more strongly associated with rape empathy than sexual outcome severity scores, indicating utility of assessing coercive tactics. Additional research is needed regarding the psychometric properties of the SES-SFP and other measures of sexual perpetration in order to accurately assess rates of these behaviors and inform preventive interventions.

Keywords: psychometrics, sexual aggression, rape, college students

Initial Evidence for the Reliability and Validity of the Sexual Experiences Survey-Short Form Perpetration (SES-SFP) and Characteristics of Sexual Perpetration in College Men

To date, there is no gold standard tool for assessing sexual perpetration. Many investigators use modified versions of existing sexual victimization instruments or instruments with theoretical but little empirical support. Yet, the accurate measurement of sexual assault perpetration is critical both to research and to designing interventions to prevent sexual assault. Sexual assault is common on college campuses; approximately 25% of college women experience rape while on campus (Carey, Durney, Shepardson, & Carey, 2015). Despite advances in the identification and treatment of those who experience rape, estimates of sexual perpetration (defined as sexually victimizing another person) are less definite. Although estimates range widely due to differences in definition and measurement, 8 – 31% of college men report engaging in sexual perpetration (Gidycz, Warkentin, & Orchowski, 2007; Kolivas & Gross, 2007). A better understanding of sexual perpetration, including how to accurately measure and reduce this behavior is crucial to reduce the negative impact of sexual assault. The goal of the current study was to provide data on the psychometric properties of one of the most commonly used measures of sexual perpetration (the Sexual Experiences Survey-Short Form Perpetration: SES-SFP: Koss et al., 2007).

The original Sexual Experiences Survey (Koss & Gidycz, 1985) is one of the most commonly used measures of sexual victimization and sexual perpetration. The SES was first developed as a measure of sexual victimization of women by men, focusing on behaviorally-specific descriptions of unwanted sexual behavior from the perspective of women as the targets of male aggression (e.g., “Have you ever had sexual intercourse with a man when you didn’t want to because he threatened to use physical force (twisting your arm, holding you down, etc.) if you

didn't cooperate?"). To assess perpetration, a parallel version was created with slight wording changes. Conceptually, this is a problematic way to assess sexual perpetration as items continued to reflect a female perspective of victimization reversed to assess perpetration. An improved method would be to develop a questionnaire focusing on behaviorally specific descriptions of perpetration behaviors, rather than the sexual experience of victims. The SES-SFP was designed to address this limitation and consists of questions from the perspective of the person engaging in coercive behavior; however, the validity of the SES-SFP has not been examined.

Psychometric Properties of the Original SES

Initial research on the psychometric properties of the original SES was promising; internal consistency (Cronbach's $\alpha = .89$) was good. There was also evidence of good test-retest reliability with 73-97% agreement between self-reports and interview responses administered one week apart (Koss & Gidycz, 1985; Ouimette, Shaw, Drozd, & Leader, 2000).

Research on the validity of the SES primarily focused on convergent validity, assessing whether the SES items were interpreted by respondents in the manner intended by researchers. These studies demonstrated significant but relatively low correlations ($r_s = .54-.61$) between men's SES self-reports and interviewer assessments of the same behavior (Koss & Gidycz, 1985; Ouimette et al., 2000). Further, interview studies found that participants often interpreted items differently than intended by researchers (Ross & Allgeier, 1996), calling into question the SES's validity.

Revisions of the Measure

The original SES was considered innovative for its use of behaviorally specific language to measure sexual assault; however, researchers suggested that the SES could be improved in the following areas: the assessment of consent, assessing the consumption of alcohol/other

substances, and improved assessment of the tactics used to coerce sex (for further reading: Cook, Gidycz, Koss & Murphy, 2011; Hamby & Koss, 2003; Koss et al., 2007). As a result, many researchers have modified the SES for their own specific purposes and needs, making it difficult to compare results across studies. Due to acknowledged limitations and the proliferation of modified versions, the SES was revised by a team of expert researchers (Koss et al., 2007). In this revision versions were created that separately measured victimization (the short form victimization: SES-SFV and the long form victimization: SES-LFV) and perpetration (the short form perpetration: SES-SFP and long form perpetration: SES-LFP) in gender neutral terms.

Following, the SES-SFP (and SES-LFP) have a markedly different structure and expanded item content from the original SES. One consequence of creating a measure focusing on perpetration rather than modifying existing victimization items has been a more detailed description of the coercive tactics of perpetration. Each SES-SFP item begins with a description of a sexual outcome as a stem that is followed by five possible tactics that could be used to obtain/coerce the sexual outcome. A focus on tactics can reveal data unique to the aggressor and highlight different targets for intervention. For instance, recent research has found differential predictors of sexual perpetration for men who engaged in verbal versus physical coercion (DeGue, DiLillo & Scalora, 2010). Focusing on tactics increased the number of items from 13 in the original SES to 36 in the SES-SFP. Yet, these changes have received little empirical attention. Abbey, Parkhill, & Koss (2005) found that merely changing the order of the description of the sexual outcome with the tactic used doubled reports of sexual perpetration for some types of sexual assault, indicating that even small changes to the measure can change its psychometric properties. Thus, psychometric testing of the SES-SFP is necessary to ensure its reliability and validity.

In addition, the SES-SFP is now conceptualized as a formative measure rather than a latent measure (Koss et al., 2007). Whereas a latent measurement model assumes that all items are necessarily related to each other as they reflect an underlying latent construct that is the presumed common cause or etiology of each item, in a formative measurement model, items are not *necessarily* related but may be. For the purposes of research that is descriptive, such as studies which assess prevalence rates, a formative measurement model is recommended (Koss et al., 2007). Cronbach's alpha as a measure of internal consistency is not useful for formative measures (Diamantopoulos, Riefler, & Roth, 2008); however it remains unclear what alternative is recommended. However, test-retest reliability is strongly recommended as the best evidence of reliability for formative measures (Diamantopoulos, Riefler, & Roth, 2008).

The authors of the SES-SFP also suggested that future research examine temporal discretion as a measure of validity (Koss et al., 2007). Temporal discretion refers to whether participants show discrimination in their answers for two separate time frames (Koss et al., 2007). With respect to the SES-SFP, temporal discretion refers to respondents being able to discriminate between reports of behavior in the past year and since age 14 but excluding the past year. We were unable to find any published research on temporal discretion using the SES, although research has found no difference in reports of the frequency of violence for one vs. twelve months in a sample of teens, suggesting problems with temporal discretion in violence research (Hilton, Harris & Rice, 1998). Given the number of changes in the SES-SFP, the re-conceptualization of the SES-SFP, and research on how even small changes can create different results (Koss & Hamby, 2003; Abbey, Parkhill, & Koss, 2005), the SES-SFP should arguably be treated as an entirely new measure in terms of establishing its psychometric properties.

Psychometric Properties of the SES-SFP

Although the SES-SFP has been available since 2007, little research has examined its psychometric properties. Testa et al., (2015) found evidence of structural validity in a sample of college students. Another study examined the construct validity of the SES-Long Form Perpetration, focusing on whether participant descriptions of their own behavior matched their answers on the SES-LFP (Buday & Peterson, 2015). Results revealed that participants often interpreted the SES-LFP items differently than researchers intended, raising questions about validity. However, convergent evidence of validity of the a modified SES-SFP has been reported; positive relationships between SES-SFP scores and measures of partner violence, and traits associated with violence were found (e.g. rape myths, hostility towards women, impulsivity, and sexual sensation seeking: Davis et al., 2014). Yet, Davis et al., (2014) examined a community sample and did not control for social desirability, an important construct in the assessment of sexual perpetration that has been related to underreporting in past research (Freeman, Schumacher, & Coffey, 2015). Additionally, both Davis et al., (2014) and Testa et al., (2015) used a modified version of the SES and not the SES-SFP (although these versions are similar to the SES-SFP, see Abbey, Parkhill, & Koss, 2005 and Abbey et al., 2007 for further details). In sum, there is no published work (to our knowledge) examining the psychometric properties of the SES-SFP, and none controlling for social desirability, or examining a high risk sample such as college students.

The Nomological Network of Sexual Perpetration

Drawing from the literature, we constructed a nomological network of sexual perpetration via the SES-SFP. As those who engage in one form of violence often engage in other violent acts

as well, we chose to include other forms of violence as an important part of the nomological network (Hamby & Grych, 2013). We expected a correlation of at least $r = .15$ between SES-SFP scores and a measure of partner violence sexual perpetration (Davis et al., 2014). We also hypothesized that SES-SFP scores would be modestly related to trait aggression scores ($r = .10$: Lemmer, Gollwitzer, & Banse, 2015). We also included the construct of rape empathy (feeling empathy for those who experience rape), because of the importance of this construct in models of sexual aggression (Malamuth, 1983) and the emphasis on this construct given by the authors of the SES-SFP (Koss et al., 2007). We expected a negative relationship between rape empathy and sexual perpetration with the strength of correlations between .30 and .46 following previous literature (Osman, 2011; Abbey et al., 2007).

Aims & Analytic Plan

The goal of this study was to evaluate the basic psychometric properties of the Sexual Experiences Scale – Short Form Perpetration (SES-SFP) in a sample of college men using an anonymous web survey. We chose to administer the SES-SFP exactly as published in order to ease interpretation of the psychometric data and provide a baseline for future research. Comparison measures were selected in accordance with the follow basic criteria: a) they measured domains that were relevant but independent constructs from sexual perpetration and b) they had been previously used with college populations.

The first aim of the present study was to assess reliability of the SES-SFP. We hypothesized that we would find evidence of internal consistency and test-retest reliability given the similarity of the content (if not structure) of SES-SFP to the original SES (Koss & Gidycz, 1985). We selected Spearman's rho to measure internal consistency as Spearman's rho assumes a relationship between scores, consistent with prior research on multiple forms of violence

(Hamby & Grych, 2013). A two-week time period was selected as the target timespan between Time 1 and Time 2 in order to allow for a timespan long enough that practice effects would be minimized but short enough that it is unlikely that new sexually aggressive behavior would occur. This is a common test-retest interval in the field of trauma psychology (for example, Foa, Cashman, Jaycox, & Perry, 1997). A second aim of this study was to assess the validity of the SES-SFP. We assessed temporal discretion by testing for differences in the rate of violence reported in the prior year versus since age 14 but not including the prior year. We also assessed convergent validity. As mentioned, we hypothesized that the SES-SFP would be weakly to moderately correlated with convergent measures and that these relationships would be strongest for the most severe categories of sexual perpetration. Finally, we conducted follow-up analyses designed to assess the utility of emphasizing tactics in the SES-SFP by examining the relationship of rape empathy and trait aggression to traditional category scores based on sexual outcome vs. tactic based category scores.

Methods

Participants

Participants were 402 college men aged 18 years and older enrolled in psychology courses at a large, urban, Midwestern University who completed the SES-SFP for extra credit. The sample ranged in age from 18-53 ($M = 21.9$, $SD = 5.0$, mode = 19). Participants were mostly heterosexual ($n = 355$, 88.3%) and Caucasian ($n = 311$, 77.4%); 7.2% identified their race as African American, 6.7% as Asian/Asian American, 1.7% as Native American/American Indian, and 7.2% as Hispanic or Latino. The mean number of college years completed was 2.1 ($SD = 2.0$), and one quarter of the sample ($n = 104$) indicated that their major was Psychology. Given the nature of the study topic, the study began as a cross-sectional assessment. Once it became

apparent that online data collection was feasible, we designed methods to allow for anonymous completion of the Time 2 assessment (i.e., self-generated subject numbers that were not linked to participant identifiers); $n = 326$ were invited to complete Time 2. A total of 155 individuals provided adequate SES-SFP data (completed at least one item with no obvious pattern of frivolous responding) at Time 2; 72 of these participants also provided a matching ID and participated within the required 7 – 21 day window (see Procedures for further detail).

Materials

All participants completed the study self-report measures anonymously through the online system Qualtrics.

The SES-SFP. The SES-SFP (Koss et al., 2007) consists of 38 items; items were presented verbatim from the instrument with the exception of the item assessing respondent age and gender (which was eliminated as all participants were male and age was assessed on a demographic questionnaire). The first 35 items are behaviorally specific descriptions of sexual perpetration and are presented in a compound manner. These items begin with a description of a sexual outcome as a stem, and each sexual outcome is followed by five possible tactics (a – verbal pressure, b – verbal criticism, c – incapacitation, d – physical threats, e – physical force) that could be used to obtain the sexual outcome. The same five possible tactics (a-e) are described for each of the seven different sexual outcomes. Participants indicated the number of times (0, 1, 2, 3+) that they had engaged in the specified behaviors in the past twelve months (“one year ago since today”) and, separately, since age 14 but not including the past year (“starting at your fourteenth birthday and ending one year ago today”). Age 14 was selected by the original authors to differentiate the assessed behaviors from childhood sexual abuse. These two time-frames are referred to as the “past year” and “prior years”, respectively. “Lifetime” is

used to refer to estimates that combine the past year and prior years' time frames. Following Davis et al. (2014), separate analyses were computed using frequency scores, category scores (coding the most severe behavior endorsed of the four possible categories of sexual perpetration), and dichotomous scores (0 = no perpetration, 1 \geq perpetration). We also computed tactic category scores by coding the most severe tactic reported as the assigned category.

An additional two items assess the gender of the victim of the sexual perpetration, and the extent to which the respondent acknowledged rape (e.g., "Do you think you may have ever raped someone?") The acknowledgment item is not used to calculate frequency, category, or dichotomous scores but is only used to assess acknowledgment. Acknowledgment is determined by comparing behaviorally specific scores to the acknowledgment item; in other words, do participants who respond affirmatively to behaviorally specific descriptions of rape also acknowledge their behavior as rape?

The SES-SFP defines four mutually exclusive categories of sexual perpetration (none, unwanted sexual contact, sexual coercion, and rape/attempted rape) based on a combination of the tactic used and the outcome obtained. *Unwanted sexual contact* was defined as touching the private areas or sexual organs of another's body or removing clothes without their consent but not attempting sexual penetration. *Sexual coercion* was defined as using verbally coercive tactics to obtain sexual acts. Verbally coercive tactics include using verbal pressure and/or threats such as telling lies or threatening to end the relationship (tactic a), and showing displeasure, criticizing or getting angry (tactic b) to obtain or to attempt to obtain oral sex, anal sex or sexual intercourse. *Rape and attempted rape* was defined as taking advantage of the target's altered consciousness (e.g., being drunk; tactic c), threatening physical harm (tactic d) or using physical

force or a weapon (tactic e) to obtain or to attempt to obtain oral sex, anal sex, or sexual intercourse.

Convergent validity. The Sexual Coercion subscale of the Revised Conflict Tactics Scales (CTS2) was used to assess intimate partner sexual perpetration in the past year (Straus, Hamby, Boney-McCoy, & Sugarman, 1996). Each of the seven items on the CTS2-SC is rated using an 8-point response scale (0,1,2,4,8,15,25,99) with each successive response option representing a different range of frequency including (0=never) to (99=not in the past year but it happened before). The CTS2 has shown evidence of convergent validity and reliability in previous research (Simpson & Christensen, 2005).

The Aggression Questionnaire (AQ) has been used to measure aggression in multiple populations; the AQ consists of 29 items rated 1 (extremely uncharacteristic of me) to 5 (extremely characteristic of me) (Buss & Perry, 1992). In the present sample the mean AQ total score was 63.9 (SD = 17.2, range 29 – 116), Cronbach's alpha = 0.90.

The Rape Empathy Scale (RES: Deitz, Blackwell, Daley, & Bentley, 1982) consists of 19 paired items where each item selected from the pair "I feel that rape is an act that is not provoked..." vs. "I feel that rape is an act that is provoked" is rated from 1(not at all preferred) to 7 (completely preferred). Higher scores indicate greater empathy for rape victims. The mean score for the sample was 101.4 (SD = 19.3, range 24 – 133), Cronbach's alpha = 0.91.

Social desirability. The Marlow Crowne Social Desirability Scale (SDS) has been widely used in the area of sexual violence (Crowne & Marlowe, 1960; Gidycz, Warkentin, & Orchowski, 2007). The SDS consists of 33 true/false items that are socially desirable but unlikely to be universally true to evaluate the tendency to respond in socially desirable ways. The mean SDS score for the sample was 8.40 (SD = 2.71, range = 0 – 21); Cronbach's alpha = 0.77.

Procedures

The following procedures were approved by the first author's University IRB. Participants were recruited using flyers for a study on "Men's Behavior in Relationships". In order to protect participant confidentiality the questionnaires were administered anonymously through a web-based survey. To complete the study anonymously and link participants' data from the separate assessments, a three part algorithm using personal but unidentifiable information was used to allow participants to create their own unique identification codes (study IDs). This algorithm was then provided again at Time 2 so that participants could re-generate their unique study ID. At Time 1, participants completed all study questionnaires in a randomized order. Eleven days later, participants were e-mailed a signup code to remind them to access the SONA experiment management website for the Time 2 survey. At Time 2, participants completed just the SES-SFP; SES-SFP administrations at Time 1 and Time 2 were identical in order to assess whether participant responses remained stable. Reminders were sent on days 13 and 15 to encourage participation at Time 2. In order to further maintain anonymity, researchers were unaware of which participants completed the study at which times and thus, were not able to undertake targeted recruitment/retention efforts. The number of days between Time 1 and Time 2 varied from 0 to 105. The mean number of days between participation at Times 1 and 2 was 17.3 (SD = 14.93); participants completing outside the 7 – 21 day window were excluded from analyses. We tested for characteristics that were related to Time 2 participation in three separate logistic regressions each utilizing a different sexual perpetration variable. Years of college was a significant, positive predictor, OR = 1.21 (95% CI: 1.03 – 1.42).

Results

Data cleaning

There were minimal missing data on study measures ($n < 20$ across variables). For violence variables (SES-SFP, CTS2) missing data were assumed to be the modal value (0). For continuous measures, (AQ, RES, SDS) a cut-off of 20% or more missing was used to exclude participants, otherwise, scores were prorated. Regarding skewness and kurtosis, scale variables were within acceptable limits (-2 to 2), while SES-SFV variables were highly skewed with excess kurtosis (Tabachnick & Fidell, 2007). Following recommendations, we computed results using Spearman's rank correlations to account for skewness/kurtosis; results were highly similar to those computed using Pearson's correlations (Tabachnick & Fidell, 2007). Following, we report Pearson's correlations for consistency with prior research.

Next, we tested whether social desirability scores were predictors of any study variables using regression. Social desirability was predictive of trait aggression scores, $p = .05$; following, social desirability was controlled for in the tests of convergent validity using partial

correlations. Descriptive Information on Sexual Perpetration

Nearly one quarter ($n = 98$, 24.4%) of the sample reported engaging in some type of sexual perpetration on the SES-SFP since age 14. When considering coding the most severe type of sexual perpetration reported by participants, rape was the most frequent type of sexual perpetration reported since age 14 ($n = 52$, 12.9% of the sample) followed by unwanted sexual contact (7.0%) and sexual coercion (4.5%). Regarding the past year only, 13.4% of the sample reported some type of sexual perpetration with rape as the most frequency type of sexual perpetration reported (7.5% of the sample) followed by sexual coercion (3.2%) and unwanted sexual contact (2.7%). Regarding frequency, the estimated frequency of sexual perpetration ranged from 1 – 70 for both time periods examined. Only six respondents answered “yes” to the item, “Do you think you may have ever raped someone?” The rate of acknowledged perpetrators

(i.e., those who endorsed behaviors consistent with rape and answered affirmatively to the aforementioned item), was 3.8%.

Reliability

Internal consistency. We calculated Spearman's rho correlations for each tactic as a potential measure of internal consistency, see Table 1. Frequency scores for lifetime tactics a (telling lies, verbal pressure) and b (showing anger, criticism) were highly correlated, $\rho(402) = .67$, $p < .001$; whereas tactics c (taking advantage-when drunk), d (physical threat), and e (physical force), were moderately correlated with tactic a.

*Test-retest reliability*¹. Partial correlations (controlling for social desirability, deleting listwise) comparing SES-SFP scores (frequency, category, and dichotomous) at Time 1 and Time 2 were calculated using the Time 2 sample ($n = 60$). Correlations for the same score from Time 1 to Time 2 were all were significant at $p < .001$. Lifetime category and prior years' frequency scores were correlated the strongest, partial $r(60) = .69$. Lifetime dichotomous scores correlated the most weakly, partial $r(60) = .59$; past year frequency scores were similar, partial $r(60) = .61$.

Next, we investigated percent agreement in the lifetime SES-SFP category scores across the two assessments, see Table 2. There was agreement in 54 of 66 cases (81.8%), unweighted kappa = .45, quadratic weighted kappa = .61. Five participants who reported "none" at Time 1 reported some sexual perpetration at Time 2 while four participants reported some type of sexual perpetration at Time 1 but "none" at Time 2.

Validity

Temporal discretion. In order to test temporal discretion we examined the frequency of SES-SFP items for past year versus prior years' time frames and computed chi-squares. As

¹ Analyses were repeated using all Time 2 participants, including those who we were unable to match conclusively, $n = 168$. Results were similar.

shown in Table 3, participants appeared to discriminate between past years and prior years' in making responses. In other words, the rates of endorsement in these two time periods (scored dichotomously) are not identical numerically or statistically, $\chi^2(402) = 117.93, p < .001$.

Convergent validity. Validity analyses used the entire sample, $n = 402$ to ascertain relationships between SES-SFP scores and partner violence, trait aggression, and rape empathy scores. All correlations were modest in magnitude (partial r values between .14 - .31), with similar values in the two timeframes and between scoring approaches, see Table 4. The scores between CTS2-SC, AQ, and SES-SFP scores were consistent with hypotheses while the relationship of SES-SFP scores to RES scores was weaker than expected (Osman, 2011; Abbey et al., 2007).

Follow-up Analyses

Rape empathy and trait aggression. Levels of rape empathy and trait aggression by SES-SFP category scores were also examined. The distribution of lifetime category scores were: none, 75.6%; sexual contact, 7.0%; sexual coercion, 4.5%; rape/attempted rape, 12.9%. An ANOVA revealed that rape empathy varied by lifetime category score, $F(3,391) = 6.05, p < .001$, Cohen's $d = .43$; post-hoc analysis (Dunnett T3) indicated that rape empathy was lower in the rape/attempted rape group compared to the no perpetration group, see Table 5. The same was true for trait aggression, $F(3,397) = 6.87, p < .001$, Cohen's $d = .45$.

We also examined rape empathy scores using category scores based on tactics rather than outcome. In addition to separating variance by specific tactic, this also separates intoxication vs. physical coercion rape cases which had been grouped in outcome based category scores. Distribution of tactic category scores were: none, 75.6 %; tactic a (telling lies, verbal pressure), 3.0%; tactic b (showing anger, criticism), 3.7%; tactic c (intoxication), 11.9%; tactic d (physical

threats), 0.5%; tactic e (physical force), 5.2%. Rape empathy declined as scores rose, in other words, the more severe tactics that participants reported, the lower rape empathy scores reported, $F(5,388) = 5.57, p < .001$, Cohen's $d = .54$. Tukey's post-hoc tests indicated that RES scores were lower in the physical force group compared to the intoxication, verbal coercion (telling lies, verbal pressure), and no perpetration groups. Utilizing the same strategy to assess trait aggression, AQ scores were higher in the verbal coercion (getting angry, criticism) and intoxication groups than the no perpetration, group, $F(3,395) = 4.60, p < .001$, Cohen's $d =$

.48. Discussion

The goals of this study were to evaluate the psychometric properties of the SES-SFP in a sample of college men in order to facilitate rigorous empirical research on sexual perpetration and sexual assault intervention. Currently, there is no gold-standard, psychometrically supported, behaviorally specific assessment of sexual perpetration; instead, many investigators have used modified versions of the original SES or alternative instruments. Thus, presenting evidence for the reliability and validity of the SES-SFP is critical for further research examining sexual perpetration to develop both assessments and effective interventions.

Data from this investigation provides good evidence for internal consistency as measured by Spearman's rho but questionable test-retest reliability for the SES-SFP. Test-retest correlations were very similar in this study to the validation of the original SES (*partial* $r = .59 - .69$ vs. $r = .61$). Percentage agreement in scores between the two time points was similar to the original validation for dichotomous scores (90.7%) but poor for category scores (0 - 50.0%). This indicates that the SES-SFP is most reliable when used with a dichotomous (yes/no for perpetration) scoring system to identify perpetration. For dichotomous scores our study found a 10% error rate in identifying cases, which is within the recommended margin of error; however,

for category scores, which assess the type of perpetration, agreement was poor, outside the acceptable rate of error (McHugh, 2012). Notably, our small sample size is only adequate for assessing the dichotomous scores (Sim & Wright, 2005); but we find our results concerning nonetheless. We recommend that the SES-SFP be used with a dichotomous scoring system until further test-retest reliability data are available.

We were unable to find test-retest reliability on other measures of sexual perpetration to compare the present results to; therefore, although our data are limited we consider this an important baseline for comparison in future reliability research. It is unclear whether agreement dropped due to low numbers of positive cases, regression to the mean, or participant memory errors. While it is unclear what the exact standard for reliability should be in the case of a behavioral measure, we see no reason to recommend a lower standard than that currently suggested for biomedical research, which is 80% (McHugh, 2012). Indeed, given the seriousness of sexual perpetration, we are inclined to recommend a higher standard. A prospective design with multiple time periods may better assess whether change in scores is related to new episodes of behavior. We strongly recommend further test-retest reliability research to more conclusively address this important scientific standard.

The present data provided evidence of validity. We found that participants responded differentially regarding their behavior in the two SES-SFP time frames, indicating adequate temporal discretion. This is important in demonstrating that the SES-SFP items can be used to assess sexual perpetration behavior in the recent as well as distant past. Consistent with Davis et al. (2014) and hypotheses we found correlations with convergent measures were consistent regardless of the SES-SFP scoring system used. Of particular note, SES-SFP scores were negatively related to rape empathy scores regardless of scoring system used and the degree of

this relationship differed by the severity of the perpetration behavior. We also examined this relationship for trait aggression; trait aggression was higher in participants with a history of rape. We found a larger effect size for perpetration on rape empathy when we considered tactics rather than outcomes; the same was true for trait aggression although to a lesser degree. Further, rape empathy was lowest for those who reported using physical force rather than incapacitation; indicating there may be clinical differences between these types of perpetration. An emphasis on assessing tactics places the focus of measurement on the behavior that is both the putative target of intervention and the behavior about which the aggressor can most reliably describe their own intentions, motivations, etc. An emphasis on tactics is also useful conceptually, relying on the *means* of coercion, which are unique to the person acting aggressively, rather than the *ends* of coercion, which are an interaction of the person acting aggressively and their target. Overall, this indicates that emphasis on tactics in the SES-SFP is a useful change and suggests the potential need for specialized intervention programs based on tactic. We recommend future research continue to explore the utility of tactic based scores.

We found that most perpetrators endorsed 1-2 tactics; this runs counter to the lay stereotype of sexual perpetrators as predators who will do whatever they can to coerce any kind of sex. Rather, perpetrators may specialize, consciously or not, in specific tactics and seek environments that facilitate them. These results suggest differing prevention/intervention approaches for individuals who engage in different tactics, and that a one-size-fits-all intervention would be less effective. For example, targeting rape empathy and other constructs related to rigid gender/masculinity norms may be most relevant for those who engage in physical tactics whereas alcohol focused interventions may be most relevant for those who utilize

incapacitation. However, it should be noted that interventions which attempted to change perpetration behavior through attitudes alone have been unsuccessful (Breitenbecher, 2000).

Limitations

The test-retest reliability data were limited in this study by the small sample size and difficulty of matching participants from Time 1 to Time 2. Our sample was also one of convenience but highly relevant to the problem of campus sexual assault. Following the power analysis suggestions of Sim & Wright (2005), future studies should recruit larger samples to specifically examine the reliability of the category scores. This study provides initial support for the internal consistency evidence of reliability and convergent evidence of validity of the SES-SFP in one relatively homogenous sample. However, additional research is needed to fully establish the construct validity of the SES-SFP across diverse populations, including diversity in age, gender, sexual orientation, racial background, and ethnic identity. Similarly, rape and attempted rape are treated as a combined category; this is based on clinical research finding that the experience of attempted rape is highly traumatic (Becker, 1982). Yet, in perpetration, attempted rape may represent something different than it does in victimization (for example, interference from environmental factors) or the distinction may not be very meaningful in terms of the etiology of perpetration (the tactic occurred but was not successful).

This study was also unable to examine the context of sexual perpetration reported by participants; the context of sexual perpetration as well as the relationships to targets, locations, antecedents, and consequences are important data for intervention programs. Future research which includes either an interview or questionnaire on the specific characteristics of incidents (ala Krahe et al., 2016 or Buday & Peterson, 2015) is recommended to answer these questions (although careful consideration of research ethics would be necessary for these studies). We

strongly recommend interview research with tools such as the Timeline Follow Back Interview (Carey, Carey, Maisto, Gordon, & Weinhardt, 2001) to answer these questions about context as well as to examine whether differences in test-retest scores are related to memory errors, differences in item interpretation, social desirability, and/or carelessness.

Conclusions

This study found initial evidence for the internal consistency reliability for the SES-SFP, but weak evidence of test-retest reliability. There was good agreement for dichotomous scores but not category scores. We also found convergent evidence of validity; SES-SFP scores were positively related to partner violence and trait aggression, and negatively related to rape empathy. This study found a high rate of sexual perpetration among college men, with approximately 25% of participants reporting any type of sexual perpetration behavior and one fifth reporting frequency estimates of sexual perpetration greater than one. The results from this study indicate that much more research is needed on the psychometric properties of the SES-SFP and related instruments in order to better understand the nature of these measures and implications of their use. We recommend only using the dichotomous scoring system of the SES-SFP until further test-retest evidence is available. Finally, we found an emphasis on the tactics employed (rather than the outcome) most useful in detecting differences in levels of rape empathy.

References

Abbey, A., Parkhill, M. R., Clinton-Sherrod, A. M., & Zawacki, T. (2007). A Comparison of

men who committed different types of sexual assault in a community sample. *Journal of Interpersonal Violence*, 22 (12), 1567–1580. doi:10.1177/0886260507306489

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(4), 364–373. JOUR. <http://doi.org/10.1111/j.1471-6402.2005.00236.x>

Becker, J. V. (1982). The effects of sexual assault on rape and attempted rape victims. *Victimology*, 7(1-4), 106–113.

Breitenbecher, K. H. (2000). Sexual assault on college campuses: Is an ounce of prevention enough? *Applied and Preventive Psychology*, 9(1), 23–52. [http://doi.org/10.1016/S0962-1849\(05\)80036-8](http://doi.org/10.1016/S0962-1849(05)80036-8)

Buday, S. K., & Peterson, Z. D. (2015). Men's and Women's Interpretation and Endorsement of Items Measuring Self-Reported Heterosexual Aggression. *Journal of Sex Research*, 52(9), 1042–1053. article. <http://doi.org/10.1080/00224499.2014.967373>

Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology*, 63(3), 452–459. doi:10.1037/0022-3514.63.3.452

Carey, M. P., Carey, K. B., Maisto, S. A., Gordon, C. M., & Weinhardt, L. S. (2001). Assessing sexual risk behaviour with the Timeline Followback (TLFB) approach: continued development and psychometric evaluation with psychiatric outpatients. *International Journal of STD & AIDS*, 12(6), 365–75. <http://doi.org/10.1258/0956462011923309>

Carey, K. B., Durney, S. E., Shepardson, R. L., & Carey, M. P. (2015). Incapacitated and

forcible rape of college women: Prevalence across the first year. *Journal of Adolescent Health, 56*(6), 678–680. <http://doi.org/10.1016/j.jadohealth.2015.02.018>

Cook, S. L., Gidycz, C. A., Koss, M. P., & Murphy, M. (2011). Emerging issues in the measurement of rape victimization. *Violence Against Women, 17*(2), 2012-18.
doi:10.1177/1077801210397741

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349–354.

Davis, K. C., Gilmore, A. K., Stappenbeck, C. A., Balsan, M. J., George, W. H., & Norris, J. (2014). How to score the sexual experiences survey? A comparison of mine methods, *Psychology of Violence, 4*(4), 445–461. doi:10.1037/a0037494

DeGue, S., DiLillo, D., & Scalora, M. (2010). Are all perpetrators alike? Comparing risk factors for sexual coercion and aggression. *Sexual Abuse: A Journal of Research and Treatment, 22*(4), 402–426. doi:10.1177/1079063210372140

Deitz, S. R., Blackwell, K. T., Daley, P. C., & Bentley, B. J. (1982). Measurement of empathy toward rape victims and rapists. *Journal of Personality and Social Psychology, 43*(2), 372–384. doi:10.1037/0022-3514.43.2.372

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*(12), 1203–1218.
<http://doi.org/10.1016/j.jbusres.2008.01.009>

Foa, E. B., Cashman, L., Jaycox, L., & Perry, K. (1997). The validation of a self-report measure of posttraumatic stress disorder: The Posttraumatic Diagnostic Scale. *Psychological*

Assessment, 9(4), 445–451. doi:10.1037/1040-3590.9.4.445

Freeman, A. J., Schumacher, J. A., & Coffey, S. F. (2015). Social Desirability and Partner Agreement of Men's Reporting of Intimate Partner Violence in Substance Abuse Treatment Settings. *Journal of Interpersonal Violence*, 30(4), 565–579.
doi.org/10.1177/0886260514535263

Gidycz, C. A., Warkentin, J. B., & Orchowski, L. M. (2007). Predictors of perpetration of verbal, physical, and sexual violence: A prospective analysis of college men. *Psychology of Men & Masculinity*, 8(2), 79. Doi:10.1037/1524-9220.8.2.79

Hamby, S., & Grych, J. (2013). Tracing the Threads of the Web: The Epidemiology of Interconnections Among Forms of Violence and Victimization. In: *The Web of Violence* (pp. 8–27). New York, NY, US: Springer. http://doi.org/10.1007/978-94-007-5596-3_2

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(3), 243–255.
doi:10.1111/1471-6402.00104

Hilton, N. Z., Harris, G. T., & Rice, M. E. (1998). On the Validity of Self-Reported Rates of Interpersonal Violence. *Journal of Interpersonal Violence*, 13(1), 58–72.
<http://doi.org/10.1177/088626098013001004>

Kolivas, E. D., & Gross, A. M. (2007). Assessing sexual aggression: Addressing the gap between rape victimization and perpetration prevalence rates. *Aggression and Violent Behavior*, 12(3), 315–328. Doi:10.1016/j.avb.2006.10.002

Koss, M. P., Abbey, A., Campbell, R., Cook, S., Norris, J., Testa, M., ... White, J. (2007).

- Revising the SES: A Collaborative process to improve assessment of sexual aggression and victimization. *Psychology of Women Quarterly*, 31(4), 357–370. Doi:10.1111/j.1471-6402.2007.00385.x
- Koss, M. P., & Gidycz, C. A. (1985). Sexual Experiences Survey: Reliability and validity. *Journal of Consulting and Clinical Psychology*, 53(3), 422–423. doi:10.1037/0022-006X.53.3.422
- Krahé, B., Haas, S., Vanwesenbeeck, I., Bianchi, G., Chliaoutakis, J., Fuertes, A., Matos, M. G., et al. (2016). Interpreting survey questions about sexual perpetration in cross-cultural research: A qualitative study with young adults from nine European countries. *Sexuality & Culture*, 20(1), 1–23. doi: 10.1007/s12119-015-9321-2
- Lemmer, G., Gollwitzer, M., & Banse, R. (2015). On the psychometric properties of the aggressiveness-IAT for children and adolescents. *Aggressive Behavior*, 41(1), 84–95.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276–282.
- Osman, S. L. (2011). Predicting rape empathy based on victim, perpetrator, and participant gender, and history of sexual aggression. *Sex Roles*, 64(7–8), 506–515. JOUR. <http://doi.org/10.1007/s11199-010-9919-7>
- Ouimette, P. C., Shaw, J., Drozd, J. F., & Leader, J. (2000). Consistency of reports of rape behaviors among nonincarcerated men. *Psychology of Men & Masculinity*, 1(2), 133–139.
- Ross, R. R., & Allgeier, E. R. (1996). 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, 26, 1587–1616. <http://doi.org/10.1111/j.1559-1816.1996.tb00087.x>

Sim, J., & Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, 85(3), 257–268.

Simpson, L. E., and Christensen, A. (2005). Spousal agreement regarding relationship aggression on the Conflict Tactics Scale-2. *Psychological Assessment*, 14(4), 423–432.

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(3), 283–316. doi:10.1177/019251396017003001

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA, USA: Allyn & Bacon/Pearson Education. <http://doi.org/10.1037/022267>

Testa, M., Hoffman, J. H., Lucke, J. F., & Pagnan, C. E. (2015). Measuring sexual perpetration in college men: A comparison of two measures. *Psychology of Violence*, 5(3), 285–293. doi:10.1037/a0037584

Tables

Table 1

Spearman's Rho Correlations between Tactic Frequency Scores as a Potential Measure of Internal Consistency, n = 402

Tactic	1	2	3	4	5
1. verbal coercion, a	—	.64	.45	.48	.39
2. verbal coercion, b		—	.39	.49	.48
3. intoxication, c			—	.43	.34
4. physical threats, d				—	.67
5. physical force, e					—

Note. All correlations were significant at $p < .001$.

Table 2

Cross-tabulation of SES-SFP Lifetime Sexually Aggressive Behavior Categories at Time 1 and Time 2, (n = 66)

Time 1 Category	Time 2 Category				Total
	None <i>n</i> (% of 54)	Sexual Contact <i>n</i> (% of 4)	Sexual Coercion <i>n</i> (% of 2)	Rape/ Attempted Rape <i>n</i> (% of 6)	
None	49 (90.7)	2	2	1	54
Sexual Contact	2	2 (50.0%)	0	0	4
Sexual Coercion	0	1	0 (0.0%)	1	2
Rape/Attempted Rape	2	0	1	3 (50.0%)	6
Total	53	5	3	5	66

Note. Entries appearing in bold indicate the percentage of participants within each Time 1 category who reported the same highest level of sexual perpetration at Time 2; there was 81.8% agreement across categories.

Table 3

Patterns of SES-SFP Item Endorsement in Two Time Periods, n = 402

Item	<u>Prior Years'</u> <i>n</i> , % endorsed in sample	<u>Past Year</u> <i>n</i> , % endorsed in sample
1. Fondled, kissed, or rubbed private areas	63, 15.7%	38, 9.5%
2. Had oral sex or had someone perform oral sex	29, 7.2%	21, 5.2%
3. Penis, fingers, or objects into a woman's vagina	24, 6.0%	17, 4.2%
4. Put penis, fingers, or objects into someone's butt	10, 2.5%	12, 3.0%
5. Tried to have oral sex or have someone perform oral sex	30, 7.5%	28, 7.0%
6. Tried to put penis, fingers, or objects into someone's vagina	28, 7.0%	15, 3.7%
7. Tried to put penis, fingers, or objects into someone's butt	17, 4.2%	12, 3.0%

Note. Items numbered following their order on the SES-SFP. Prior years' operationalized as between 14 years of age and the past year. Bolded values are statistically significant at $p < .05$.

Table 4

Convergent Validity: SES-SFP Partial Correlations Controlling for Social Desirability, ($n = 402$)

	CTS2- Sexual Coercion $r_{part} (df)$	Aggression Questionnaire (AQ) $r_{part} (df)$	Rape Empathy Scale (RES) $r_{part} (df)$
SES-SFP: Prior Years'			
Highest Category	.27***(401)	.20**(401)	-.16**(391)
Total	.19***(401)	.14**(401)	-.25***(391)
Dichotomous	.27***(401)	.22***(401)	-.15**(391)
SES-SFP: Past Year			
Highest Category	.31***(401)	.18***(401)	-.29***(391)
Total	.18***(401)	.07(401)	-.29***(391)
Dichotomous	.30***(401)	.17**(401)	-.26***(391)

$p < .05$; ** $p < .01$; *** $p < .001$

Note. Cases were eliminated listwise. Prior years' operationalized as between 14 years of age and the past year.

Table 5

Relationship between SES-SFP Category Scores and Convergent Measures

Rape Empathy Findings		
SES-SFP Score, <i>n</i>	<i>M</i>	<i>SD</i>
<u>Category scores by sexual outcome obtained</u>		
No perpetration, 298	103.45 ^a	18.99
Sexual contact, 27	99.44	14.50
Sexual Coercion, 17	102.47	17.88
Attempted rape/Rape, 52	91.48 ^a	20.63
<u>Category Scores by tactic used</u>		
No perpetration, 298	103.45 ^a	18.99
telling lies, verbal pressure, 11	105.09 ^b	19.78
getting angry, criticism, 14	97.36	15.89
intoxication, 48	99.15 ^c	15.15
physical threats, 2	84.50	20.51
physical force, 21	82.71 ^{abc}	23.02
Trait Aggression Findings		
SES-SFP Score, <i>n</i>	<i>M</i>	<i>SD</i>
<u>Category scores by sexual outcome obtained</u>		
No perpetration, 298	61.70 ^a	17.11
Sexual contact, 27	69.79 ^a	11.52
Sexual Coercion, 17	69.56	13.34
Attempted rape/Rape, 52	71.27 ^a	18.49
<u>Category Scores by tactic used</u>		
No perpetration, 298	61.70 ^b	17.11
telling lies, verbal pressure, 11	68.75	14.08
getting angry, criticism, 14	74.47 ^b	11.11
intoxication, 48	71.73 ^b	17.66
physical threats, 2	65.00	5.66
physical force, 21	66.52	15.43

Note. Paired superscripts indicate statistically significant differences between groups that share the same letter at $p < .05$