2021

Test-Retest Reliabilities of Four Tactic-first Sexual Violence History Questionnaires

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

Monica A. Garcia

Douglas L. Delahanty

Psychology of Violence
2152-0828

Test-Retest Reliabilities of Four Tactic-first Sexual Violence History Questionnaires

RaeAnn E. Anderson[1,2], Monica A. Garcia[1] and Douglas L. Delahanty[1]

Abstract

Objective: The present study documented, compared, and contrasted the test-retest reliabilities of the victimization and perpetration forms of a Tactic-first Sexual Experiences Survey (T-SESs) and the Post-Refusal Sexual Persistence Scale (PRSPSs).

Methods: 243 Mechanical Turk workers (116 women, 124 men) completed four questionnaires in a randomized order via an anonymous web survey at Time 1 and approximately one week later at Time 2.

Results: There were consistent gender differences in test-retest estimates. When assessing a history of victimization in women, both the T-SES and the PRSPS demonstrated evidence of minimal to good reliability (κ > .61, ICC = .86-92) while for men the PRSPS (κ = .64) was more consistent than the T-SES (κ = .59). When assessing a history of perpetration, there were fewer gender differences although post-hoc analyses suggest potential gender differences in assessing substance use facilitated perpetration (κ .48-.83) but were limited by few cases. Continuous scoring approaches were the most reliable, dichotomous scores were mostly reliable, and categorical scores generally did not meet minimal acceptable standards. For the rape victimization acknowledgment items, we found strong evidence of reliability for women (κ = .89, *n* = 31) and suggestive evidence of reliability for men (*n* = 7). There were few differences in reliability between standard and extended versions of the questionnaires.

Conclusions: All four questionnaires exhibited good evidence of one-week test-retest reliability when scored continuously. Evidence of reliability was strongest with the populations and constructs most well studied – victimization history among women and perpetration history among men.

Keywords

sexual violence, rape, assessment, measurement, coercion

1 Kent State University, Psychological Sciences, 144 Kent Hall, Kent, OH 44221
2 University of North Dakota, Psychology, 2000 Columbia Hall, Grand Forks, ND 58201

Corresponding Author:
RaeAnn E. Anderson, University of North Dakota, Department of Psychology, 2000 Columbia Hall, Grand Forks, ND 58202, USA. Email: raeann.anderson@UND.edu.

Sexual violence is an experience of sexual contact without consent (Koss et al., 2007). The consequences of experiencing sexual violence (i.e., sexual victimization) are often severe and include poor mental and physical health as well as challenges in relationships, parenting, and employment (Martin et al., 2011). Yet, obtaining accurate assessments of sexual violence can be extremely challenging - as exemplified by a review of men's sexual victimization, which estimated that 2-73% of American males experience victimization (Peterson et al., 2010). These broad ranges are at least partially related to differences in measurement strategy such as varying definitions and methods of assessment (Anderson et al., 2019; Bouffard & Goodson, 2017). The inability to estimate prevalence rates hinders understanding of the scope of violence and the efficacy of prevention efforts, especially when cultural stigma already suppresses reporting and recognition of sexual violence as a serious problem (Wolitzky-Taylor et al., 2010). The goal of this study was to examine the one-week test-retest reliabilities of experimental questionnaires that use a tactic-first structure that may improve the identification of sexual violence: a tactic-first Sexual Experiences Survey – Short Form (T-SES) and the Post-Refusal Sexual Persistence Scales (PRSPSs).

Tools for Assessing Sexual Violence

The most common tools used to assess sexual violence in North American have been the different versions of the Sexual Experiences Survey (SES: Koss & Oros, 1982; Koss et al., 1987). Although the original SES had good psychometric properties it has since been revised substantially (Koss et al., 2007). These revisions have resulted in separate measures of victimization and perpetration experiences and our understanding of the psychometric properties of these questionnaires is still developing. Most researchers using the SES-Short Form Perpetration (the 2007 revised SES) modify it in some manner suggesting that even after these revisions, the SES-Short Form Perpetration does not fully meet researchers' needs (Anderson et al., 2019). An important area for research is how the variations and modifications of the SES, such as those tested here, may impact the psychometric properties of the questionnaire in addition to the changes in prevalence rates.

One modification is to use tactic-first items. On the SES, each item begins with a description of a sexual behavior followed by a list of tactics that were used to coerce that specific sexual behavior. In tactic-first modifications, the order of these clauses is swapped; each item is oriented to the tactic followed by sub-items linking the tactic to a sexual behavior, see Figure 1. This simple change has dramatic outcomes; tactic-first versions of the SES record higher prevalence rates in both cross-sectional and experimental work (Abbey et al., 2005; Anderson et al., 2019; Schuster et al., 2020) and have initial evidence of convergent validity (Abbey, et al., 2019; Abbey et al., 2006). It is unclear why tactic-first versions increase prevalence rates, one very simple explanation is that the items are simply easier to read and understand. Another is that tactics are more effective memory retrieval cues because they have a stronger memory trace. In other words, coercive tactics are a more unique experience than sexual behavior; even if a person has experienced a large number of coercive sexual assaults, they have likely experienced consensual sex more often given that the average American adult has consensual sex approximately once per week starting in the late teen years (Herbenick et al., 2010; Reece et al., 2010). Crime statistics suggest that while repeated rape is too common, it rarely occurs as often as once per week (Daigle et al., 2008). The more common an experience, the less specific details are remembered. Thus, tactics may have a more distinctive memory trace.

In considering perpetration behavior, a focus on tactics is also more conceptually clear. Although a perpetrator may initiate sexual violence with a specific sexual outcome in mind, the tactics are the behavior the perpetrator has complete control of. The sexual outcome is influenced by both the perpetrator and the putative victim/survivor's behavior. Indeed, this is the entire logic of feminist self-defense interventions, women cannot control perpetrator's behavior, but they can fight back (and bystanders can intervene: Orchowski et al., 2018). In other words, focusing on tactics focuses on the means rather than the ends of sexual violence, and on the target of any perpetration prevention intervention: the behavior of perpetrators.

The Post-Refusal Sexual Persistence Scales (the PRSPS: Struckman-Johnson et al., 2019) have always adopted a tactic-focused measurement strategy and were developed to capture gender differences in rape (Struckman-Johnson et al., 2003). Partially because of the

tactic-focused design, the description of sexual behaviors is minimized in comparison to the SES and structure of the PRSPSs is simpler. Finally, one study suggests the PRSPS better fits the theoretical model of sexual perpetration (Testa et al., 2015).

Measurement Models for Sexual Violence and Reliability

In a latent measurement model, a questionnaire is developed wherein each item is presumed to represent a single aspect of the entire possible universe of items that sample the unobservable, underlying construct. In contrast, the measurement of sexual violence is the assessment of a history of behaviors or experiences. Thus, assessment of sexual violence history is more similar to checklists of events rather than each item representing some portion of an underlying, latent construct.

Figure 1

Example Items from the Questionnaires

The Sexual Experiences Survey-Short Form Victimization (SES-SFV)	The Tactic-first Sexual Experiences Survey-Short Form Victimization (T-SES for victimization)
The following questions concern your sexual experiences since the age of 14. These are personal questions, but we hope that you will be willing to answer them honestly. All of your answers will be kept confidential.	The following questions concern your sexual experiences since the age of 14. These are personal questions, but we hope that you will be willing to answer them honestly. All of your answers will be kept confidential.
4. A man put his penis into my butt, or someone inserted fingers or objects without my consent by:	1. Has someone ever overwhelmed you with continual arguments and pressure in order to:
a Telling lies, threatening to end the relationship, threatening to spread rumors about me, making promises I knew were untrue or continually verbally pressuring me after I said I didn't want to.	a Fondle, kiss, or sexually touch you, without your consent?
b Showing displeasures, criticizing my sexuality or attractiveness, getting angry but not using physical force, after I said I didn't want to.	b Make you have oral sex with them, without your consent?
c Taking advantage of me when I was too drunk or out of it to stop what was happening	c Attempt to make you have oral sex with them without our consent, but for some reason it did not happen?
d Threatening to physically harm me or someone close to me.	d Make you have anal sex with them without your consent?
e Using force, for example holding me down with their body weight, pinning my arms, or having a weapon.	e Attempt to make you have anal sex with them without your consent, but for some reason it did not happen?
…4-6 more items like this depending on sex	f Make you have sexual intercourse with them without your consent?
	g Attempt to make you have sexual intercourse with them without your consent, but

	for some reason it did not happen? ….8 more items like this
The Post-Refusal Sexual Persistence Scale-Victimization	The Post-Refusal Sexual Persistence Scale-Perpetration
Since of 14, has someone ever used any of the tactics on the list below to have sexual contact (kissed, fondled, genital touching, oral sex, anal sex, or sexual intercourse) with you after you have indicated "no" to their sexual advance?	Since age 14, which of the following strategies have you used to convince or try to convince a person to have sex (kissed, fondled, genital touching, oral sex, anal sex, or sexual intercourse) after they initially said "no":
1 Told you a lie of some kind (e.g., how much they liked/loved you)	1 Questioned their commitment to the relationship (e.g., saying "if you loved me, you would")
2 Threatened to break up with you	2 Used your older age to convince them
3 Tried to talk you into it by repeatedly asking or arguing	3 Got them drunk/high in order to convince them to have sex
4 Said they would blackmail you	4 Tied them up.
….20 more items like this	…24 more items like this

Note. Shaded sections refer to items that correspond in content on the different versions of the Sexual Experiences Survey

Each experience or behavior of sexual violence is not necessarily related and may have occurred in different incidents, perpetrated by different people, at different ages, et cetera rather than being caused by a single underlying trait or characteristic. Although perpetration may stem from an underlying shared trait, this is an empirical question yet to be answered. However, experiences of sexual victimization are not presumed to be caused by some underlying latent factor (Koss et al., 2007). To do so would be to suggest that there is some underlying trait within victims/survivors that facilitates their victimization, rather than their victimization being caused by the behavior of another person.

An induced or formative measurement model is more appropriate for assessing sexual victimization (and potentially perpetration: Diamantopoulos et al., 2008; Hulme, 2007; Koss et al., 2007). In formative measurement models, test-retest reliability is considered the most important form of reliability whereas internal consistency reliability (e.g. Cronbach's alpha) is appropriate for latent models (Diamantopoulos et al., 2008). Test-retest reliability estimates the stability of scores over time; in other words, are participant's responses stable from one week to another? In the case of sexual violence, test-retest reliability provides confidence in labeling a person's behavior or experience as rape, a serious crime; thus our focus on test-retest reliability.

Prior Estimates of Reliability – Victimization

Research suggests poorer reliability for men than women on the SES-Short Form Victimization (the revised SES for victimization) in assessing a history of victimization; with nearly one-third of cases classified differently across time for men (men: Anderson et al., 2018; women: Johnson et al., 2017). Only one study has reported kappa, which corrects for chance agreement and most estimates were < .6, the recommended cut-off for reliability (Littleton et al., 2018). Further, the worst estimates were for categorical scores (e.g., severity scores). Thus, the test-retest reliability of the revised SES for victimization may be poor, particularly for categorical scores and for men.

Embedded at the end of the SES and the subsequent revisions is the SES acknowledgment item, "have you ever been raped?" (Koss et al., 2007; Koss & Oros, 1982) This item, in combination with responses to the behaviorally specific questions, is used to categorize those who experience rape as either acknowledged, i.e., using the label rape, or unacknowledged, using an alternative and minimizing term to label their experience, such as "miscommunication" (Littleton et al., 2007). Lack of acknowledgment is common and in fact, the norm (Wilson & Miller, 2016); thus, the importance of using behaviorally specific items to detect rape. Lack of acknowledgment likely stems from internalized stigma around rape and the likelihood of negative reactions when disclosing (Ullman et al., 2007). An entire literature on the relationship between rape acknowledgment, clinical outcomes, and victimization risk has developed underscoring the importance of this construct; yet we were unable to find any estimates of test-retest reliability of the acknowledgment item.

Prior Estimates of Reliability – Perpetration

Prior estimates of test-retest reliability for the SES-SFP (the revised SES for perpetration) have been more promising but are still quite limited. Two small studies ($N \approx 70$) have found agreement between perpetration categories ranging from 82-91% over a 2-week interval (Anderson et al., 2017; Johnson et al., 2017) and a kappa of .61 (Anderson et al., 2017). But Buday & Peterson (2015) found that when interviewing women who completed these items many women were endorsing items with *victimization* experiences in mind.

The Current Study

The goal of the current study was to document and compare the test-retest reliabilities for two tactic-first measures of sexual violence: a T-SES and the PRSPS. We particularly wanted to analyze the data in a way that the estimates could be compared to prior literature. We chose to focus on one-week test-retest reliability as this timeframe is long enough to reduce practice effects but short enough that new episodes of sexual violence are unlikely. Thus, we compared the test-retest scores for the PRSPSs to the T-SESs, a modified version of the SES and we followed the analytic plans of Anderson et al., (2017, 2018) and Johnson et al. (2017). We also report test-retest estimates for the acknowledgment item. Finally, we also report the reliabilities of the extended versions of these questionnaires, as the extended versions include new items designed to expand the literature on sexual violence (Koss et al., 2007); yet, questionnaire length can induce participant fatigue that can hinder reliability.

Hypotheses (H)/Research Questions (RQ):

1. We expected reliability estimates to be similar if not better than in previous research, given stronger memory trace for tactics.
2. We hypothesized that the PRSPSs would have better evidence of reliability than the T-SESs given the simpler structure of the PRSPSs.
3. Given the lack of any prior research, we made no hypothesis regarding the reliability of the rape acknowledgment item.
4. We expected to find consistent gender differences, with victimization estimates being more reliable for women compared to men, and perpetration estimates more reliable for men compared to women.
5. We made no hypotheses about the reliability of the extended versions of the questionnaires compared to the standard versions.

Method

Participants

Participants were drawn from 466 Amazon Mechanical Turk Workers who consented and opened the Time 1 survey. Only North American, adult, Workers with a high quality and completion ratings from prior tasks (>90%) were eligible to participate. The survey was advertised as "Questionnaires about Sexual Behavior 1 – Kent State University." Time 1 participants were invited to participate in Time 2 one week later; 277 participants consented and opened the Time 2 survey. 274 of these participants were able to be matched with their Time 1 data using their Amazon MTurk IDs.

Analytic Sample

Of these 274 participants, 46.7% were women (n = 128), 51.5% were men (n = 141) and 1.8% identified as a gender minority of some type (n = 5). Participants were mostly heterosexual (n = 235 – 85.8%), some were bisexual (n = 28 - 10.2%), gay (n = 10 - 3.6%), and one did not identify as any of the aforementioned identities (n = 1 - 0.4%). Participants were mostly Caucasian (85.0%, n = 233); 24 identified as African American (8.8%), 20 as Asian American (7.3%), and 2 as Native American (0.7%). A small number (8.0% - n = 22) identified as Hispanic or Latino/a. The average age of participants was 32.76 (SD = 7.22), range = 20 – 68. In comparison to those who did not participate in Time 2, the analytic sample was more likely to identify racially as White (69.8% vs 85.0%, $\chi 2(1) = 15.68$, $p < .001$), more likely to report victimization (Time 1 PRSPS, $\chi 2(1) = 5.348$, $p = .021$), and more likely to report perpetration (Time 1 T-SES, $\chi 2(1) = 9.511$, $p = .002$).

Procedures

The following procedures were approved by the Institutional Review Board of Kent State University. Data were collected in April 2018 using Qualtrics to administer the online questionnaires. After providing informed consent, participants completed the six study questionnaires (four sexual violence questionnaires, a discriminant validity questionnaire, and a demographics form) in a randomized order. After six days, participants were emailed an invitation to complete Time 2. Participants were sent an additional two reminder emails over the course of the following week and then survey access was disabled. The mean number of days between Time 1 and Time 2 was 8.66 days (SD = 2.5), median 8, and the mode was 7 days.

Measures

The Tactic-First Sexual Experiences Survey

The T-SES used in this study is a modification of the tactic-first SES used by Abbey et al., (2005). All T-SES items are compound items beginning with a stem that describes a tactic followed by seven sub-items that represent the sexual behavior coerced: sexual touching, attempted oral sex, oral sex, attempted anal sex, anal sex, attempted vaginal sex, vaginal sex. This structure is an inversion designed to mirror the revised SES (Koss et al., 2007). The nine tactic stem items represent the following: verbal pressure (2), verbal criticism (1), alcohol/drug incapacitation (3-4), threats of physical force (1), use of physical force (1), and multiple perpetrator attacks (1). The acknowledgment item, "have you ever been raped?" was administered last. Crossing the nine tactic items by the seven sexual behavior sub-items resulted in a total of 63 items on the victimization version and 70 items on the perpetration version as the perpetration version contains an additional alcohol/drug facilitation item.

Changes Made to the T-SESs

We made a number of modifications to the questionnaire as published in 2005, reflecting advances in the science of sexual violence measurement since the idea of a tactic-first SES was first introduced (Abbey et al., 2005). Thus, the questionnaire used in this study is not

a standard questionnaire, but an experimental tool designed to help investigate questions related to measurement strategy and stimulate further research. We use the term T-SES not to legitimize our questionnaire as an official variant of the SES, but rather for simplicity.

We dropped mention of the gender of the perpetrator consistent with Anthony & Cook (2012)'s findings that gender-neutral instructions are more inclusive, less heteronormative, and do not appear to substantially alter psychometric properties. We separated the threat of force and use of physical force content into two items, consistent with Koss et al., (2007) and with Anderson et al., (2017)'s suggestion that a focus on specific tactics, rather than combining tactics, may improve validity and highlight potential intervention targets. We also added three items bringing the total number of items to 63 (victimization) and 70 (perpetration), respectively. Added items were drawn from the long form of the revised SES published by Koss et al., (2007). Two items were added to both forms - one describing encouraging or pressuring someone to consume substances and one regarding multiple perpetrator attacks, in order to stimulate further research on these tactics. The perpetration form included one item regarding giving someone substances without their knowledge; we only included this on the perpetration form given that this question is difficult to answer from the victimization perspective and may be better captured with other items. The T-SES items were given in a hierarchical order per Koss et al., (2007), although the sexual behavior stems were randomized within each tactic item given suggestions that randomization of items may improve disclosure rates (Dietz & Jasinski, 2007).

The Post-Refusal Sexual Persistence Scales

(PRSPS: Struckman-Johnson et al., 2003). The first iteration of the PRSPS contained 19 items; each item lists a single tactic that was used to coerce any sexual behavior such as, "threatened to break up with you." Sexual behavior is defined broadly in the instructions as "genital touching, oral sex, or intercourse"; in this study we expanded this instruction set by adding kissing, fondling, and anal sex consistent with Strang, et al., (2013). The types of tactics assessed by the PRSPSs include enticement, verbal coercion, misuse of authority, alcohol/drugs, and physical force; thus, the PRSPSs assess two categories of tactics not included on the revised SES (enticement and misuse of authority). PRSPS items are administered in a randomized order. Construct validity for both victimization and perpetration forms were demonstrated in the original study by eliciting written descriptions of the sexual violence incidents corresponding to endorsed items (Struckman-Johnson et al., 2003).

Changes Made to the PRSPSs

We added items to equalize content with the T-SESs to control for variance introduced by content differences as prior research suggests this can be a large source of discrepancy between the PRSPS and the SES (Strang et al., 2013). Thus, in this study, the victimization version contained 24 items while the perpetration version contained 28 items. The items added were for the following content: verbal coercion (3), threats of physical harm (1) and multiple perpetrator attacks (1) for a total of 24 items. Given Strang and Peterson's (2017) findings that perpetrators sometimes skip endorsing items that contain long lists of behaviors unless they can endorse every behavior, we separated some content into multiple perpetration items. For example, "took advantage of the fact that you were drunk or high" may adequately capture a victimization experience but could correspond to multiple perpetration behaviors including giving someone alcohol without their knowledge, verbally pressuring someone who is intoxicated, et cetera. Materials are available on osf.io. Finally, the version of the PRSPS we tested is very similar to revisions made to the PRSPS in Strang et al (2013) that they refer to as the Sexual Strategies Scale.

Traditional vs. Extended Questionnaires

Although participants were presented with one questionnaire, we scored two "versions" of each questionnaire. The first version was scored with only the traditional items that appear on each questionnaire in order to provide some comparison between this study and

prior estimates of reliability. These traditionally scored questionnaires are referred to simply by the name introduced in the Methods (e.g., the T-SES and PRSPS) in following analyses. The second scoring version included the additional items added for this study; these questionnaires are referred to extended versions (e.g., the T-SES extended).

Scoring

Each item was administered with the response scale, "how many times? 0, 1, 2-5, 6-9, 10?" We tested dichotomous, categorical, and continuous scores consistent with prior research (Davis et al., 2014; Anderson et al., 2017, 2018a). We coded responses of "1" or greater on any item as endorsement of that item. Categorical scores, sometimes called severity or ordinal scores, reflect the most severe tactic endorsed by a participant: enticement (PRSPS only), misuse of authority (PRSPS only), verbal coercion, substance intoxication, physical threats/harm, multiple perpetrators. For continuous scores responses were added to a total score.

Analytic Plan and Power

We based our data analytic strategy on the methods used in prior literature to maximize comparability between our results and prior research. To that end, we report percent agreement, kappa, and correlations. Kappa of .80 is considered the cut-off for clinical constructs (strong agreement) with .6 considered a minimum (McHugh, 2012). Intraclass correlations were selected given that the ICC assumes that the data come from the same group or construct whereas Pearson's r or Spearman's rho are ideal for measuring the degree of association between two different constructs (Koo & Li, 2016).

In prior research the number of women with positive cases of victimization have ranged from 174 – 189, with correlations between scores in the r = .52 range (Littleton et al., 2019; Johnson et al., 2017). Sample sizes for men have historically been much smaller; ranging from 11-21 positive cases of victimization with correlations in the range of r = .04 - .53. For assessing perpetration in men, the number of positive cases has ranged from 11-17 with correlations in the r = .57 - .69 range. Research on the critical point of stability for correlations in the range of .5 suggests that a sample size of 34-68 would be adequate (Schönbrodt and Perugini, 2013). Our sample of 69+ women and 47+ men with victimization histories suggest correlational analyses may be appropriately powered. For perpetration, our sample of 24+ women and 40+ men with histories of perpetration may be under-powered; yet are 2-3x larger than prior research. We report 95% confidence intervals with 1000 bootstrapped samples to further contextualize our effect sizes. However, we recommend future research investigate the use of more complex models that are specially adapted for over-dispersed and count-type datasets.

Data Quality and Cleaning

We required MTurk Workers at Time 1 and Time 2 to be highly rated and have completed >90% of their previous tasks. Of the Workers who opted to participate in Time 2, the vast majority provided matching IDs (98.9%, unlike prior college samples: Anderson et al., 2017; 2018). We excluded 31 participants who did not complete at least 80% of either set of questionnaires so that we would have adequate data between questionnaires to compare. Most who were excluded for missing data were the same 24/25 participants who did not pass our two attention check items "please select 2-5 for this question". Given mixed findings on attention checks questions as an indicator of data quality (Hauser and Schwarz, 2016), and our other measures of data quality we did not further restrict our sample. Skewness and kurtosis were within normal limits (+/- 3.0) for all variables except continuous variable which was kurtotic for victimization and both skewed and kurtotic for perpetration, thus we report Spearman's rho.

Results

Victimization

Descriptive Findings

On the T-SES, 59.5% of women (69/116) and 38.0% of men (46/121) reported a history of some type of sexual victimization on the T-SES, whereas on the PRSPS, 69.0% of women (80/116), and 50.4% of men (61/121) reported victimization. Table 1 summarizes the number of positive cases (e.g., prevalence rates) across the questionnaires.

Table 1

Prevalence Rates by Questionnaire and Gender in this Sample

Questionnaire	% Women	% Men
Victimization		
Tactic-SES	59.5	38.0
Tactic-SES extended	61.2	39.7
PRSPS-V	69.0	50.4
PRSPS-V extended	69.0	50.4
Perpetration		
Tactic-SES	20.7	32.3
Tactic-SES extended	23.1	39.5
PRSPS-P	29.3	45.2
PRSPS-P extended	31.6	46.8

Note. SES = Sexual Experiences Survey, PRSPS = Post-Refusal Sexual Persistence Scale, V = victimization, P = perpetration

Dichotomous Scoring

For the T-SES, agreement was substantial for women (κ = .70) but only moderate for men (κ = .59), not meeting the minimal cut-off of .6, see Table 2. Reliability for the PRSPS was in the substantial range (κ = .61 -.64) for women and men.

Categorical Scoring

Both questionnaires were borderline substantially reliable for women, (w κ = .71 - .65) with smaller effect sizes (.52 - .58) and broad confidence intervals. Neither met minimum reliability thresholds for men (w κ ≤.6 with effect sizes < .6).

Continuous Scores

Reliability was good for women and men on both the T-SES and the PRSPS. For women, the T-SES (ICC = .92) was significantly more reliable than the PRSPS (ICC = .86), z = 2.11, p = 0.03. And both questionnaires were more reliable for women (.92 - .86) than for men (.79), z ≥ 2.13, p < .03.

Rape Victimization Acknowledgment Item

Of the 31 women who responded "yes" at Time 1, 29 responded "yes" at Time 2 (93.5%), κ = .89 (strong). Of the seven men who responded "yes" at Time 1, five responded "yes" at Time 2 (71.4%), κ = .56 (weak).

Overall Summary of Reliability for Victimization Questionnaires

For women, both questionnaires met suggested thresholds for dichotomous and continuous scoring. For men only both questionnaires were reliable when scored continuously. Neither questionnaire was sufficiently reliable when score categorically for either gender. Estimates were in general stronger for women than men.

Perpetration

Descriptive Results

On the T-SES for perpetration, 20.7% of women (24/116) and 32.3% of men (40/124) reported a history of sexual perpetration. On the PRSPS, 29.3% of women (34/116) and 45.2% of men (56/124) reported sexual perpetration.

Dichotomous Scoring

Both questionnaires were reliable when scored dichotomously, for women, κ = .60-.68 and men, κ = .60-.65 see Table 2.

Categorical Scoring

Both questionnaires were sufficiently reliable for women when scored categorically (κ = .60-.65) while for men effect size estimate suggest it may not be (<.6).

Continuous Scores

Both questionnaires were reliable for men and women. For women, the T-SES (ICC = .92) was more reliable than the PRSPS (ICC = .83), z = 2.86, p = 0.004 while they were nearly equivalent for men (.86-.87).

Overall Summary of Reliability of Perpetration Questionnaires

Both questionnaires were acceptable when scored dichotomously or continuously for either gender. Neither questionnaire was acceptable for men when scored categorically.

Standard vs. Extended Versions

Victimization

Victimization prevalence rates on the extended versions were equivalent on the PRSPS and higher than the standard T-SES (women: 59.5 vs. 61.2/ men 38.0 vs. 39.7). Although there were some numerical differences (see Table 3), there were no interpretative differences in reliability estimates between standard vs. extended versions of the T-SES and the PRSPS in assessing victimization.

Perpetration

Perpetration prevalence rates on the extended versions were higher than the standard versions although the differences were small, up to 2.4 percentage points for women and up to 7.2 for men. Although there were some numerical differences (see Table 3), the extended versions of the T-SES and PRSPS generally performed very similarly. Exceptions were all for women; dichotomous scores on the T-SES did not meet standards nor did PRPS categorical scores.

Post-Hoc Analysis

Given the poor kappas for categorical scores, we also computed follow-up analyses to evaluate reliability for each tactic individually. Categorical scores assign each person a category corresponding to the most severe tactic they experienced; thus, error variance is compounded and variably distributed across participants who range in the number of tactics they experienced. Given that the

Table 2

Summary of Test-Retest Reliability Estimates by Primary Questionnaire, Gender, and Scoring Type with Effect Sizes and 95% Confidence Intervals (n = 116 women, 107 men)

	% agreement T-SES/ PRSPS	K [CI] or WK [CI] T-SES/ PRSPS	Effect size T-SES/ PRSPS	Spearman's rho T-SES/ PRSPS	ICC T-SES/ PRSPS	Minimum met?
Victimization Findings						
Women						
Dichotomous	85.3%/ 83.6%	.70[.56, .83]/ .61[.44, .75]	.70/ .62	—	—	Yes
Categorical	71.6%/ 63.8%	.71[.61, .81]/ .65[.55, .76]	.58/ .52	—	—	No
Continuous	—	—	—	.84 [.76, .91]/ .77 [.65, .87]	.92 [.88, .94]/ .86 [.80, .90]	Yes
Men						
Dichotomous	81.0%/ 81.8%	.59[.43, .74]/ .64 [.50, .77]	.59/ .64	—	—	Yes - PRSPS
Categorical	71.9%/ 63.6%	.60[.46, .73]/ .56 [.43, .68]	.49/ .45	—	—	No
Continuous	—	—	—	.65 [.48, .79]/ .66 [.52, .79]	.79 [.70, .85]/ 79 [.70, .85]	Yes
Perpetration Findings						
Women						

Dichotomous	87.1%/ 87.1%	.60[.39, .78]/ .68[.53, .82]	.60/ .68	—	—	Yes
Categorical	87.1%/ 87.1%	.60[.48, .85]/ .65[.51, .80]	.60/ .68	—	—	Yes
Continuous	—	—	—	.67 [.48, .83]/ .73 [.59, .85]	.92 [.89, .95]/ .83 [.76, .89]	Yes
Men						
Dichotomous	82.3%/ 83.1%	.60[.45, .75]/ .65[.52, .79]	.61/ .66	—	—	Yes
Categorical	75.8%/ 71.8%	.60[.47, 74]/ .62 [.51, 74]	.53/ .54	—	—	No
Continuous	—	—	—	.68 [.53, .81]/ .69 [.55, .81]	.87 [.81, .91]/ .86 [.81, .90]	Yes

Note. T-SES = Tactic-First sexual Experiences Survey. PRSPS = Post-Refusal Sexual Persistence Scale. Phi effect sizes were used for dichotomous scores and Cramer's V for categorical scores. Chi-square (dichotomous scores) and likelihood ratios (categorical scores) were used to assess statistical significance of percent agreement, all analyses were significant, $p < .05$. Weighted kappa is reported for categorical scores. ICC's of .7 and above was deemed acceptable based on prior conventions (Post, 2016)

Table 3

Reliability of the Extended Questionnaires

	% agreement T-SES ext/ PRSPS ext	κ or $_{WK}$ T-SES ext/ PRSPS ext	Effect Size T-SES ext/ PRSPS ext	Spearman's rho T-SES/ PRSPS	ICC T-SES ext/ PRSPS ext
	Victimization Findings				
Women					
Dichotomous	85.3%/ 83.6%	.70[.57, .83]/ .61[.44, .77]	.71, .62	—	—
Categorical	69.8%/ 59.5%	.71 [.62, .81]/ .65 [.55, .75]	.60, .51	—	—
Continuous	—	—	—	.85[.76, .91]/ .77 [.65, .86]	.91 [.88, .94]/ .84 [.77, .89]
Men					
Dichotomous	80.2%/ 81.8%	.58[.41, .73]/ .64[.50, .77]	.58, .64	—	—
Categorical	66.9%/ 61.2%	.59[.46, .72]/ .56[.43, .68]	.45, .46	—	—
Continuous	—	—	—	.67[.52, .79]/ .65[.49, .78]	.82 [.74, .87], .79 [.70, .86]
	Perpetration Findings				
Women					
Dichotomous	83.8%/ 85.5%	.54[.35, .70], /.66[.49, .80]	.54/ .66	—	—
Categorical	82.9%/ 78.6%	.61[.43, .81], /.66[.53, .79]	.66/ .56	—	—

Continuous	—	—	—	.61[.39, .77]/ .74[.59, .84]	.94 [.91, .96]/ .84 [.76, .89]
Men					
Dichotomous	83.7%/ 84.7%	.70[.57, .82]/ .69[.56, .81]	.70/ .880	—	—
Categorical	78.2%/ 69.4%	.70[.58, .82]/ .62[.51, .74]	.56/ .50	—	—
Continuous	—	—	—	.73[,60, .84]/ .72[.60, .84]	.86 [.79, .90]/ .86 [.80, .90]

Note. T-SES = Tactic-First sexual Experiences Survey. PRSPS = Post-Refusal Sexual Persistence Scale. Phi effect sizes were used for dichotomous scores and Cramer's V for categorical scores. Weighted kappa is reported for categorical scores. ICC's of .7 and above was deemed acceptable based on prior conventions (Post, 2016).

number of cases per tactic was sometimes very small by gender, ranging from 7 – 33, we report findings for the entire sample using the standard questionnaires. This data is available in a Table at the following website: https://osf.io/74m6c/ .

Supplemental Table 1

Reliability analysis for each type of tactic using entire sample

Tactic	T-SES *n*, % agreement, κ , κ 95% CI	T- SES gender differences in κ	PRPS *n*, % agreement, κ, κ 95% CI	PRPS gender differences in κ
Victimization, N = 241				
Enticement	—		120, 81.3%, .63, .52-.73	—
Verbal coercion	107, 83.4%, .66, .57-.75		121, 78.4%, .57, .46-.67	—
Misuse of authority	—		54, 85.1%, .58, .44-.69	—
Substance use	60, 87.1%, .66±, .54-.76	men κ = .53	74, 85.1%, .64±, .53-.74	men κ = .48
Physical force	57, 87.9%, .66±, .54-.77	men κ = .46	57, 85.9%, .61±, .48-.72	men κ = .46
Multiple perpetrator	25, 92.5%, .65, .49-.79		20, 92.5%, .55, .35-.72	—
Perpetration, N = 243				
Enticement	—		68, 83.1%, .56, .43-.68	
Verbal coercion	62, 84.8%, .60, .48-.71		63, 86.4%, .65, .53-.75	
Misuse of authority	—		18, 92.6%, .49, .26-.67	

Substance use	22, 92.6%,	.60±,	.41-.76	women κ = .83, men κ = .48	27, 90.1%,	.57±, .40-.72	women κ = .67
Physical force	19, 96.3%,	.75±,	.58-.90		22, 93.4%,	.60±, .40-.77	women κ = .71*
Multiple perpetrator	17, 96.3%,	.72±,	.51-.87	women κ = .83*	15, 94.2%,	.43, .15-.66	

Note. CI = confidence interval. Only gender differences outside of +/-.10 compared to the total sample are reported. Column *n* is for Time 1 cases.*there were some errors in computing CIs due to the small sample such that the upper CI included 1.00

Victimization

Kappas on the T-SES tactic scores (verbal, substance use, physical force, multiple perpetrator) ranged from .65-.66. For the PRSPS, kappas ranged from .55 - .64. For the PRSPS, kappas for verbal coercion (.57) misuse of authority (.58) and multiple perpetrator (.55) did not meet standards. We repeated these analyses by gender; kappas more than .10 worse for men were for substance use and physical force on both questionnaires.

Perpetration

Kappas on the T-SES ranged from .60 to .7. Estimates for the PRSPS ranged from .43 - .65 with enticement (.56), misuse of authority (.49), substance use facilitation (.57), and multiple perpetrator attacks (.43) failing to meet minimum thresholds. When repeating these analyses by gender, T-SES substance use facilitation improved for women (.83 from .60) and worsened for men (.48). Similarly, on the PRPS, substance use was more reliable for women (.67 from .57) when analyzed separately.
Discrepancy Rates and the Tactic-First Hypothesis (Hierarchical T-SES Compared to PRSPS-V Among n-130)

Kappa ranged from .66 - .72 and percent agreement ranged from 83.85 – 95.35%, indicating good agreement. These values are reported in Table 1. Analyses below focus on only chi-square findings for clarity.

Discussion

One barrier to sexual violence research has been the lack of precision in the measurement of violence; and specific to the assessment of sexual violence, the lack of test-retest reliability data. Our findings document initial evidence of reliability for a set of tactic-first questionnaires. We found that reliability estimates reflect the biases of the current literature – reliability estimates for victimization for men were poorer as were perpetration estimates for women.

Victimization Findings

Our findings highlight the complexity of assessing sexual victimization. Generally, there were few differences between the T-SES and the PRSPS (H2) and gender and scoring approach appears to be more salient factors in explaining differences in reliability than questionnaire. The only approach to assessing victimization that was reliably consistent across genders were continuous scores (H4). Indeed, although we hypothesized that the reliability estimates reported in this study would be stronger than those reported in previous literature (and they generally were: H1), we were surprised that our estimates still failed to meet minimum standards in many cases. This suggests that percent agreement as reported in prior research (Johnson et al., 2017; Anderson et al., 2017; Anderson et al., 2018) is a misleading assessment of reliability for sexual violence questionnaires. The poor reliability of categorical scores is consistent with Littleton et al., (2018)'s findings, the one prior study that reported kappa. We suggest that reliability for categorical scores is worse than dichotomous or continuous scores because categorical scores compound measurement error by coding only the most severe experience of a variable range of experiences. We recommend researchers avoid this scoring method.

The reliability of the T-SES rape acknowledgment item was good for women (H3). In spite of the growing literature on rape acknowledgment and clinical outcomes (Wilson & Miller, 2016), little has been reported on the psychometric properties of this item. Reliability of this item was also good for men but given the sample size (n = 7), we consider this finding very preliminary. In addition, using only the acknowledgment item would under-estimate cases by orders of magnitude, in this study, reducing the identified cases of sexual victimization by at least half for women (26.5 % vs. 59 – 69%) and even more for men (5.8% vs. 33 – 50%).

Perpetration and Post-hoc Analysis Findings

Perpetration findings mirrored victimization findings – although gender differences were less pronounced. We were surprised there was not stronger evidence of reliability for the PRSPS given

Testa's findings on structure and that the PRSPS was designed with gender differences in mind (H2: Struckman-Johnson al., 2003). In general, there were few differences between the questionnaires especially for perpetration. Post-hoc analyses suggested which tactics are least reliable and highlighted potential gender differences. This suggests that future research concentrate on better understanding the victimization experience via verbal coercion, misuse of authority, and multiple perpetrators. Notably, multiple perpetrator attacks are less common, and our reliability estimates may reflect poor power due to the small number of cases overall and especially when analyzed by gender. We were surprised by how poor reliability estimates were for perpetration via substance use facilitation on the T-SES, especially in comparison to the same content on the PRSPS. However, Littleton et al., (2018) also found low kappas estimates for substance use victimization. This finding may reflect the uncertainty introduced by alcohol in sexual situations – while initial consumption may be consensual or planned, continued consumption introduces uncertainties that may be difficult to assess afterwards given the initial context and pharmacological effects of alcohol. Secondly, these gender differences may reflect false positive rates for women on perpetration measures (Buday and Peterson, 2015). Poor reliability, specific to women reporting on perpetration via substance use may reflect that some of these women were actually thinking of victimization experiences (H4). We have strong reservations about using these questionnaires to assess perpetration in women without further research. There were also few differences between the standard and extended versions of the questionnaires which is assuring for users who wish to adding items from the long form as we did (H5).

Limitations

While this study represents one of the larger samples to date our sample size overall was still small and subsequently, contained only a small number of positive cases to analyze. We did not have an adequate sample size to test for effects related to sexual or gender minority status. This is an important point as sexual and gender minority people experience sexual violence at a much higher rate compared to heterosexual people (Rothman, Exner, & Baughman, 2011; James et al., 2016), and currently available measures may not fully capture their experiences (Dyar at el., 2019). It is imperative that this type of work continue with a range of populations – our findings are specific to American adults on MTurk. MTurk participants tend to be more educated, less religious, less likely to be employed, less extraverted and have lower self-esteem (Goodman et al., 2013). The MTurk population also tends to learn and change rapidly; thus, methods for assessing data quality are uncertain and evolving (Hauser et al., 2018; Chandler and Shapiro, 2016). We did not randomly assign participants to the standard versus extended versions of the questionnaires. Finally, some research suggests more complex data analytic approaches are needed for this type of data (Swartout et al., 2015).

Research Implications

Our findings suggest that tactic-first questionnaires increased reported prevalence rates and are at least somewhat reliable measures of sexual violence. As suggested by Buday and Peterson (2015), the gender differences found here reflect the literature – existing questionnaires contain embedded gender biases that reflect a gender stereotypic range of experiences. For example, recent research suggests that being made to penetrate may be a somewhat common form of victimization against men by women (Anderson et al., 2020). It is possible that qualitative research to develop new items is necessary.

Estimates of reliability in this study were generally in the minimally acceptable range (H1). This is notable given the larger number of positive cases (24-73) in our study compared to prior work (Anderson et al., 2017; Anderson et al., 2018; Johnson et al., 2017). Yet, larger samples may be required. It is also possible that strong reliability for this type of behavioral measure is just more difficult to achieve for a number of reasons. Sexual victimization is not a latent construct and the theory of measurement for induced models is still being developed. Another complicating factor is that the time-period, e.g., "since you were 14" often spans years if not decades. Finally, some of the behaviors assessed may be highly context dependent. Using verbal means to 'repeatedly ask for sex" may be interpreted quite differently across participants and what may be coercive in one setting may be consensual but unwanted in another, belying simple behavioral descriptions.

Clinical and Policy Implications

This study provides valuable data for individuals measuring sexual violence. Our findings underscore the value of kappa instead of percent agreement as assessments of reliability. The present findings also underscore the need for further research on the use of tactic-first questionnaires and research designs that cut against the stereotypical, yet also very common paradigm of men as perpetrators/women as victims. Though research on this stereotypical dyad is necessary and important, our results underscore the frequency of men's victimization and women's perpetration and current science's limitation in understanding these phenomena. The frequency of men's victimization and women's perpetration suggests these phenomena also deserve much greater clinical attention – including screening and direct therapeutic intervention. Finally, although we tested experimental questionnaires modified to reflect recent advances in science, these tools still do not capture all salient factors. For example, the SES method for assessing the relationship to the perpetrator of violence is inconsistent with some health surveillance systems, such as those used by the World Health Organization (Garcia-Moreno et al., 2005). Another area for future research is examining the taxonomy of violence and how to create questionnaires that adequately represent these taxonomies (Bagwell-Gray, 2019; Cook & Parrott, 2009).

Conclusions

This study demonstrated adequate test-retest reliability for four tactic-first sexual violence questionnaires. The reliability of these tactic-first questionnaires was comparable to the traditional sexual behavior-first versions and better in some instances. Reliability was strongest for the questionnaires and populations that have historically been studied the most – women who experience victimization and men who perpetrate. Reliability was worse for the experiences least studied – men's victimization and women's perpetration. Our findings underscore the potential usefulness of tactic-first questionnaires while highlighting the need for more gender-expansive research in sexual violence.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The data were collected with the support of a small institutional grant awarded to Dr. Anderson via the Kent State University Applied Psychology Center. Dr. Anderson's efforts were also supported by a grant from the National Institute on Alcohol Abuse and Alcoholism 5K01AA026643-02/03. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agency.

ORCID iD

RaeAnn E. Anderson https://orcid.org/0000-0001-9938-0717
Douglas L. Delahanty https://orcid.org/0000-0002-9021-7064

References

Abbey, A., Parkhill, M. R., & Koss, M. P. (2005). The effects of frame of reference on responses to questions about sexual assault victimization and perpetration. *Psychology of Women Quarterly*, *29*(4), 364–373. https://doi.org/10.1111/j.1471-6402.2005.00236.x

Abbey, A., Helmers, B. R., Jilani, Z., McDaniel, C., & Benbouriche, M. (2019, August) Measuring Men's Sexual Aggression Against Women: Evaluating Robustness of Findings. In S. Cook (Chair), *Evaluating the Sexual Experiences Survey-Validity and Uses*. Symposium conducted at the meeting of the American Psychological Association, Chicago, IL, USA.

Abbey, A., Parkhill, M. R., BeShears, R., Clinton-Sherrod, A. M., & Zawacki, T. (2006). Cross-sectional predictors of sexual assault perpetration in a community sample of single African American and Caucasian men. *Aggressive Behavior*, *32*(1), 54–67. https://doi.org/10.1002/ab.20107

Anderson, R. E., Cahill, S. P., & Delahanty, D. L. (2017). Initial Evidence for the Reliability and Validity of the Sexual Experience Survey-Short Form Perpetration (SES-SFP) in College Men. *Journal of Aggression, Maltreatment & Trauma*.

Anderson, R. E., Cahill, S. P., & Delahanty, D. L. (2018). The Psychometric Properties of the Sexual Experiences Survey-Short Form Victimization (SES-SFV) and Characteristics of Sexual Victimization Experiences in College Men. *Psychology of Men and Masculinity*, *19*(1), 25–34. https://doi.org/10.1037/men0000073

Anderson, R. E., Goodman, E. L., & Thimm, S. S. (2020). The assessment of forced penetration: A necessary and further step toward understanding men's sexual victimization and women's perpetration. *Journal of Contemporary Criminal Justice*, 104398622093610. https://doi.org/10.1177/1043986220936108

Anderson, R. E., Silver, K. E., Ciampaglia, A. M., Vitale, A. M., & Delahanty, D. L. (2019). The Frequency of Sexual Perpetration in College Men: A Systematic Review of Reported Prevalence Rates From 2000 to 2017. *Trauma, Violence, & Abuse.* https://doi.org/10.1177/1524838019860619

Anthony, E. R., & Cook, S. L. (2012). Assessing the impact of gender-neutral language on disclosure of sexual violence. *Psychology of Violence*, *2*(3), 297–307. https://doi.org/10.1037/a0028562

Bagwell-Gray, M. E. (2019). Women's experiences of sexual violence in intimate relationships: Applying a new taxonomy. *Journal of Interpersonal Violence*, 088626051982766. https://doi.org/10.1177/0886260519827667

Bouffard, L., & Goodson, A. (2017). Sexual coercion, sexual aggression, or sexual assault: how measurement impacts our understanding of sexual violence. *Journal of Aggression, Conflict and Peace Research*, *9*(4), 269–278.

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. https://doi.org/10.1080/00224499.2014.967373

Chandler, J., & Shapiro, D. (2016). Conducting Clinical Research Using Crowdsourced Convenience Samples. Annual Review of Clinical Psychology, 12, 53–81. https://doi.org/10.1146/annurev-clinpsy-021815-093623

Cook, S., & Parrott, D. (2009). Exploring a taxonomy for aggression against women: can it aid conceptual clarity? *Aggressive Behavior*, *35*(6), 462–476. doi.org/10.1002/ab.20321

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 Nine Methods. *Psychology of Violence*, *4*(4), 445–461. https://doi.org/10.1037/a0037494

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

Dietz, T. L., & Jasinski, J. L. (2007). The effect of item order on partner violence reporting: An examination of four versions of the revised Conflict Tactics Scales. *Social Science Research*.

Dyar, C., Messinger, A. M., Newcomb, M. E., Byck, G. R., Dunlap, P., & Whitton, S. W. (2019). Development and Initial Validation of Three Culturally Sensitive Measures of Intimate Partner Violence for Sexual and Gender Minority Populations. *Journal of Interpersonal Violence*, 088626051984685. https://doi.org/10.1177/0886260519846856

García-Moreno, C., Jansen, H. A. F. M., Ellsberg, M., Heise, L., & Watts, C. (2005). WHO multi-country study on women's health and domestic violence against women. *Geneva: World Health Organization*, *204*, 1-18.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. Journal of Behavioral Decision Making, 26(3), 213–224. https://doi.org/10.1002/bdm.1753

Herbenick, D., Reece, M., Schick, V., Sanders, S. A., Dodge, B., & Fortenberry, J. D. (2010). Sexual behaviors, relationships, and perceived health status among adult women in the United States: results from a national probability sample. *The journal of sexual medicine*, *7*, 277-290.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. Behavior Research Methods, 48(1), 400–407. https://doi.org/10.3758/s13428-015-0578-z

Hauser, D., Paolacci, G., & Chandler, J. (2018). Common Concerns with MTurk as a Participant Pool: Evidence and Solutions. https://doi.org/10.31234/osf.io/uq45c

Hulme, P. A. (2007). Psychometric evaluation and comparison of three retrospective, multi-item measures of childhood sexual abuse. *Child Abuse and Neglect*, *31*(8), 853–869. https://doi.org/10.1016/j.chiabu.2007.03.016

James, S. E., Herman, J. L., Rankin, S., Keisling, M., Mottet, L., & Anafi, M. (2016). *The Report of the 2015 US Transgender Survey*. Washington, DC. Retrieved from https://transequality.org/sites/default/files/docs/usts/USTS-Full-Report-Dec17.pdf

Johnson, S. M., Murphy, M. J., & Gidycz, C. A. (2017). Reliability and validity of the Sexual Experiences Survey - Short Forms victimization and perpetration. *Violence and Victims*, *32*(1), 78–92. https://doi.org/10.1891/0886-6708.VV-D-15-00110

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163.

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. https://doi.org/10.1111/j.1471-6402.2007.00385.x

Koss, M. P., Gidycz, C. A., & Wisniewski, N. (1987). The scope of rape: Incidence and prevalence of sexual aggression and victimization in a national sample of higher education students. *Journal of Consulting and Clinical Psychology*, *55*(2), 162–170.

Koss, M. P., & Oros, C. J. (1982). Sexual Experiences Survey: A research instrument investigating sexual aggression and victimization. *Journal of Consulting and Clinical Psychology*, *50*(3), 455–457.

Littleton, H., Layh, M., Rudolph, K., & Haney, L. (2018). Evaluation of the Sexual Experiences Survey—Revised as a screening measure for sexual assault victimization among college students. *Psychology of Violence*. https://doi.org/10.1037/vio0000191

Littleton, H. L., Rhatigan, D. L., & Axsom, D. (2007). Unacknowledged rape: How much do we know about the hidden rape victim? *Journal of Aggression, Maltreatment & Trauma*, *14*(4), 57–74. 4

Martin, S. L., Macy, R. J., & Young, S. K. (2011). Health and economic consequences of sexual violence. In J. W. White, M. P. Koss, & A. E. Kazdin (Eds.), *Violence against women and children, Vol 1: Mapping the terrain* (pp. 173–195). American Psychological Association. https://doi.org/10.1037/12307-008

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Orchowski, L. M., Edwards, K. M., Hollander, J. A., Banyard, V. L., Senn, C. Y., & Gidycz, C. A. (2018). Integrating sexual assault resistance, bystander, and men's social norms strategies to prevent sexual violence on college campuses: a call to action. *Trauma, Violence, & Abuse*, 152483801878915. https://doi.org/10.1177/1524838018789153

Peterson, Z. D., Voller, E. K., Polusny, M. A., & Murdoch, M. (2010). Prevalence and consequences of adult sexual assault of men: review of empirical findings and state of the literature. *Clinical Psychology Review*, *31*(1), 1–24. doi.org/10.1016/j.cpr.2010.08.006

Post, M. W. (2016). What to do with "moderate" reliability and validity coefficients?. *Archives of physical medicine and rehabilitation*, *97*(7), 1051-1052.

Reece, M., Herbenick, D., Schick, V., Sanders, S. A., Dodge, B., & Fortenberry, J. D. (2010). Sexual behaviors, relationships, and perceived health among adult men in the United States: results from a national probability sample. *The journal of sexual medicine*, *7*, 291-304.

Rothman, E. F., Exner, D., & Baughman, A. L. (2011). The Prevalence of Sexual Assault Against People Who Identify as Gay, Lesbian, or Bisexual in the United States: A Systematic Review. *Trauma, Violence, & Abuse*, *12*(2), 55–66. https://doi.org/10.1177/1524838010390707

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize?. Journal of Research in Personality, 47(5), 609-612.

Schuster, I., Tomaszewska, P., Marchewka, J., & Krahé, B. (2020). Does Question Format Matter in Assessing the Prevalence of Sexual Aggression? A Methodological Study. Journal of Sex Research, 1–10. https://doi.org/10.1080/00224499.2020.1777927

Strang, E., & Peterson, Z. D. (2017). Unintentional Misreporting on Self-Report Measures of Sexually Aggressive Behavior: An Interview Study. *The Journal of Sex Research*, *54*(8), 1–13. https://doi.org/10.1080/00224499.2017.1304519

Strang, E., Peterson, Z. D., Hill, Y. N., & Heiman, J. R. (2013). Discrepant responding across self-report measures of men's coercive and aggressive sexual strategies. *Journal of Sex Research*, *50*(5), 458–469. https://doi.org/10.1080/00224499.2011.646393

Struckman-Johnson, C., Anderson, P. B., Struckman-Johnson, D., & Smeaton, G. (2019). The Post-Refusal Sexual Persistence Scale. In R. Milhausen, J. K. Sakuluk, T. D. Fisher, C. M. Davis, & W. Yarber (Eds.), *Handbook of Sexuality-Related Measures* (4th ed.). New York, NY, USA: Routledge.

Struckman-Johnson, C., Struckman-Johnson, D., Anderson, P. B., Struckman-Johnson, C., Struckman-Johnson, D., & Anderson, P. B. (2003). Tactics of sexual coercion: when men and women won't take no for an answer. *Journal of Sex Research*, *40*(1), 76–86. https://doi.org/10.1080/00224490309552168

Swartout, K. M., Thompson, M. P., Koss, M. P., & Su, N. (2015). What is the best way to analyze less frequent forms of violence? The case of sexual aggression. Psychology of Violence, 5(3), 305–313. https://doi.org/10.1037/a0038316

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

Ullman, S. E., Townsend, S. M., Filipas, H. H., & Starzynski, L. L. (2007). Structural models of the relations of assault severity, social support, avoidance coping, self-blame, and PTSD among sexual assault survivors. *Psychology of Women Quarterly*, *31*(1), 2337. doi.org/10.1111/j.1471-6402.2007.00328.x

Wilson, L. C., & Miller, K. E. (2016). Meta-Analysis of the Prevalence of Unacknowledged Rape. *Trauma, Violence & Abuse*, *17*(2), 149–159. doi.org/10.1177/1524838015576391

Wolitzky-Taylor, K. B., Resnick, H. S., McCauley, J. L., Amstadter, A. B., Kilpatrick, D. G., & Ruggiero, K. J. (2011). Is Reporting of Rape on the Rise? A Comparison of Women With Reported Versus Unreported Rape Experiences in the National Women's Study-Replication. *Journal of Interpersonal Violence*, 26(4), 807–832.