1-5-2021

An Initial Test of the Tactic-First and Item-Order Hypotheses: Accounting for Response Discrepancies in Sexual Victimization Questionnaires

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

Erica L. Goodman

Alyssa M. Ciampaglia

Recommended Citation

An Initial Test of the Tactic-First and Item-Order Hypotheses: Accounting for Response Discrepancies in Sexual Victimization Questionnaires

RaeAnn E. Anderson[1] , **Erica L. Goodman[1]**,
and Alyssa M. Ciampaglia[2]

Abstract

Reported prevalence rates of sexual violence range widely in the literature, even in the same sample when using two different questionnaires of the same construct. These discrepancies are concerning as they indicate we may be underestimating the rate of sexual violence and, therefore, resources and treatment for victims. Thus, the current study aimed to investigate two mechanisms that may contribute to differences in reported prevalence rates across the literature and discrepancies within studies: the tactic-first and item-order hypotheses. Participants were 265 MTurk workers whom all completed the Post-Refusal Sexual Persistence Scale (PRSPS), then were randomly assigned to one of two versions of a tactic-first Sexual Experiences Survey (T-SES). Experimental conditions varied in the item order of the T-SES, one condition received the traditional hierarchical item order ($n = 130$) while the other received a randomized item order ($n = 135$). Our results suggest strong support for the tactic-first hypothesis; victimization prevalence rates on the T-SES were double compared to the traditional SES (54.1 vs. 19.8%) in prior research. Further, in both conditions, victimization prevalence rates were statistically equivalent between the PRSPS (62.6%) and the tactic-first SESs (56.2 and 54.1%), $\chi^2(1) < 2.5$, $p \approx .1$ -.7, contrary to prior research. We did not find support for the item-order hypothesis; there were few differences between item-order conditions. Our findings indicate that unintentional underreporting remains a threat to validity in sexual violence assessment, and continued research into the mechanisms of measurement is warranted.

Keywords

rape, measurement, assessment, sexual assault, sexual violence, gender

1   Department of Psychology, University of North Dakota, Grand Forks, ND, USA
2   Children's Hospital of Philadelphia, Center for Fetal Diagnosis and Treatment
3   Northwestern University, Department of Psychiatry & Behavioral Sciences, Feinberg School of Medicine
4   Children's Hospital of Philadelphia, Center for Fetal Diagnosis and Treatment

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 any form of sexual contact without consent (Basile et al., 2014). This serious public health problem can take many forms, ranging from unwanted sexual contact (e.g., groping) to the most violent and criminal, rape (e.g., sexual intercourse by incapacitation or force; Basile et al., 2014). Approximately 50% of American women will be harmed by sexual violence (e.g., sexual victimization) in some form (Black et al., 2011). Sexual victimization is associated with a wide range of physical and mental health outcomes (Dworkin et al., 2017) in addition to threatening public safety. Yet, the true scope of sexual violence remains unknown, partially due to measurement issues. For example, estimates of sexual violence victimization in men range from 2 to 73%, depending on the definition and measurement strategy used (Peterson et al., 2010). Estimates are less extreme in range for women but are still variable; estimates of rape victimization in college women range from 1.8 to 34% (Fedina et al., 2018). The way in which individuals are asked about sexual victimization greatly affects their answers. Without precise and accurate information, research on risk factors and the development and delivery of prevention, justice and treatment programs are hindered. Thus, research that examines why prevalence rate estimates vary is necessary. The goal of this study was to test two different explanations for variation in sexual violence measurement and, therefore, prevalence rates: the tactic-first and item order hypotheses.

A Brief History of Sexual Violence Measurement

The most common measure of sexual victimization is The Sexual Experiences Survey (the SES: Koss & Oros, 1982) and variations thereof, such as the officially revised and updated, Sexual Experiences Survey – Short Form Victimization, which we will refer to as the 2007 SES (Koss et al., 2007). A dramatic and field-changing innovation of the SES was the use of behaviorally specific language to describe sexual acts. This vastly increases the rate of identified victims[1] (Fisher et al., 2003). Although the mechanism is unclear, we suggest that the use of behaviorally specific language de-stigmatizes experiences for participants, facilitating disclosure. Providing a behavioral description also provides participants a clear frame of reference to compare their own experiences to and provides researchers with precise information. Indeed, using a question like, "have you ever been raped?" results in at least half as many cases being identified for women (Wilson & Miller, 2016) and even more missed cases for men (Artime et al., 2014). The use of behaviorally specific measures such as the SES have dramatically improved the recognition and prevention of sexual violence (Fisher et al., 2005; Hamby, 2014; Krebs, 2014; National Research Council, 2014).

Discrepancies in Measurement

Yet, as the above ranges of prevalence estimates suggest, there is room for improvement in estimating the true scope and impact of sexual violence. Recent research has explored how seemingly minor differences in questionnaires can result in sometimes large discrepancies in reported prevalence rates, even when using two questionnaires that incorporate behaviorally specific terms. For example, the Revised Conflict Tactics Scales only considers intimate partners as perpetrators, and 40-50% of perpetrators of sexual violence are acquaintances (Black et al., 2011). Following, the Conflict Tactics Scales should identify fewer cases than the SES, which captures cases from any perpetrator. And yet, the Revised Conflict Tactics scales identify more cases of sexual victimization than the SES, in the range of 10 – 26 percentage points higher prevalence rates, in multiple studies (Anderson et al., 2018a; Cook, 2002; Moreau et al., 2014), even when controlling for item content (Anderson et al., 2018a).

The unexpected discrepancies in sexual violence measurement are not limited to the Revised Conflict Tactics Scales and the SES. An in-depth comparison between the 2007 SES and the Post-Refusal Sexual Persistence Scale (PRSPS) also documented large discrepancies, suggesting a broader problem in the violence literature (see also: Cascardi & Muzycyn, 2016; DiLillo et al., 2006). To wit, the 1987 SES identified 45 cases while the PRSPS identified 73 (12.1 vs. 19.6%: Strang et al., 2013). And perhaps most alarmingly, of the total 93 affirmative cases, only 25 cases overlapped (e.g., were identified perpetrators by both questionnaires, Strang et al., 2013). Results were even starker using the 2007 SES; the PRSPS identified 2.6x as many cases (54.3 vs. 19.7%). Notably, the 2007 SES is more behaviorally specific than the 1987 SES, which theoretically should have decreased discrepancies but did the opposite. In replications of these discrepancies between the SES and the PRSPS, the degree of discrepancy is often largest for men – particularly when assessing men's victimization (Anderson & Delahanty, 2019). Further, while controlling for

---

[1] Although we use the phrase "victims" for clarity in this specific context, we do NOT wish to label any individual's experience for them. Individuals may want to use the term victim, survivor, neither, or both and we affirm their choice and autonomy to do so.

content does decrease discrepancies (Testa et al., 2015; Anderson & Delahanty, 2019), it does not eliminate them, suggesting the contribution of additional mechanisms (Anderson & Delahanty, 2019).

These studies also illustrate that some tactics appear harder to adjust or control for than others. Verbal coercion tactics are consistently more likely to be discrepant, perhaps due to their more context-driven nature (Anderson & Delahanty, 2019; Testa et al., 2015). Strang and colleagues (2013) detailed a veritable laundry list of possible explanations for these discrepancies that have not yet been investigated, including the operationalization of consent, tactic-first vs. behavior-first items, different response formats, and the order of items. Two possible explanations raised by Strang et al., (2013) and Anderson et al., (2018a) that remain unexplored and would impact many different violence questionnaires, not just the 2007 SES, are the use of tactic-first items and item-order issues.

Reasons for Discrepancies: Tactic-first items

The SESs use detailed, behaviorally specific, compound items; the sexual behavior is the "stem" or the first part of the item with each possible tactic that was used to coerce this sexual behavior listed underneath (Figure 1). In contrast, the PRSPS contains an initial instruction set describing what sexual behavior is. The PRSPS defines sexual behavior as "sexual contact (kissed, fondled, genital touching, oral sex, anal sex, or intercourse)," and then each item listed below is a possible tactic used to coerce sexual behavior. For example, a PRSPS item is "tried to talk you into it by repeatedly asking", see Figure 1. Thus, each PRSPS item is a tactic, whereas 2007 SES items are a combination of sexual behavior + tactic(s). Tactic-first items look something like this:

"Someone forced me [tactic = force] to have sex [behavior = sex] without my consent." Whereas a sexual behavior-first item might look like this:

"I had vaginal sex [behavior = vaginal sex] without my consent by force [tactic = force].

One experimental study demonstrated the power of tactic-first items (Abbey et al., 2005). Abbey and colleagues randomly assigned participants to either a tactic-first modified SES or a traditional, sexual behavior-first SES. The tactic-first version resulted in higher prevalence rates (74%) than the traditional, sexual behavior-first version (62%: Abbey et al. 2005); the effect was even starker for perpetration. This is one possible explanation for discrepancies between the 2007 SES and the Revised Conflict Tactics Scales and the 2007 SES and the PRSPS[2]. This finding is also consistent with a recent systematic review of nearly 80 studies that found tactic-first questionnaires reliably produced larger prevalence rates than sexual behavior-first questionnaires (Anderson et al., 2019).

The mechanism of tactic-first items producing higher prevalence rates is unclear but could be related to memory retrieval. Tactics may have a sharper memory trace – they may be easier to recall because the experience is more unique (Conway & Bekerian, 1987) and negatively valenced (Kensinger & Ford, 2020). Sexual behaviors, even when coerced by violent tactics, would share some overlapping features with consensual sexual experiences and more frequent experiences are harder to recall in detail (Roediger et al., 2010). Negatively valenced memories are easier to retrieve (Kensinger & Ford, 2020). Prior research on the use of behaviorally specific items to assess rape suggests that the specificity of items is essential for triggering memory and facilitating memory retrieval (Fisher, 2009). In sum, evidence is building that describing tactics first or emphasizing tactics over the sexual behavior is a mechanism of discrepancy; we refer to this as the tactic-first hypothesis.

Reasons for discrepancy: Item-order

Another possible explanation for measurement discrepancies is the order of the items in the questionnaire, particularly in regard to the placement of more severe items. The PRSPS and Revised Conflict Tactics Scale items appear in a randomized or semi-randomized order; the severity of the tactics (e.g., violence, criminality) is dispersed throughout the questionnaire. In contrast, SES items are administered in a hierarchical order, starting with the least severe items and building in violence severity and criminal implications. Less hierarchical and more randomized order of items results in higher prevalence rates, at least for Revised Conflict Tactics Scales items (Dietz & Jasinski, 2007; Ramirez & Straus, 2006). The mechanism behind this is unclear, but perhaps randomization de-stigmatizes the items, facilitating greater intentional disclosure by participants. It is also possible that randomized items promote greater engagement – participants are forced to concentrate more and thus, may disclose more (Swain et al., 2008). We refer to this possible mechanism of discrepancy as the item-order hypothesis.

2 It is also reasonable to call PRSPS items tactic-focused rather than tactic-first given the structure of the PRSPS. We use tactic-first for rhetorical emphasis.

Figure 1. Example Items from the Questionnaires
The Sexual Experiences Survey-Short Form Victimization (the 2007 SES)

---

1. A man put his penis into my vagina, or someone inserted fingers or objects without my consent by:

     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.

     b   Showing displeasures, criticizing my sexuality or attractiveness, getting angry but not using physical force, after I said I didn't want to.

     c   Taking advantage of me when I was too drunk or out of it to stop what was happening

     d   Threatening to physically harm me or someone close to me.

     e   Using force, for example holding me down with their body weight, pinning my arms, or having a weapon.

….4-6 more items like this

The Tactic-first Sexual Experiences Survey-Short Form Victimization (the T-SES)

---

1. Have you ever given someone of the opposite sex alcohol or drugs without their knowledge or consent in order to…:

     a   Fondle, kiss, or sexually touch them without their consent?

     b   attempt to make them have oral, anal, or vaginal sex with you, but for some reason sex did not happen?

     c   make them have oral sex with you?

     d   make them have vaginal sex with you?

     e   make them have anal sex with you?

….8 more items like this

The Post-Refusal Sexual Persistence Scale-Victimization (The PRSPS-V)

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?

     1.   Said they would blackmail you

     2.   Threatened to break up with you

     3.   Took advantage of the fact that you were drunk or high

     4.   Threatened you with a weapon

     5.   Acted together with another person (or people) to overwhelm you

…..19 more items.

The Current Study

The goal of the current study is to test the tactic-first and item-order hypotheses as mechanisms for the documented discrepancies between questionnaires in the violence literature broadly and the 2007 SES and PRSPS specifically. We seek to investigate whether changing the structure of the 2007 SES to incorporate features of the PRSPS (e.g., tactic-first items and randomized item order) will minimize or eliminate differences in prevalence rates. The strongest support for our hypotheses would be a *lack* of

statistical differences in the prevalence rates between the two measures. We use Anderson and Delahanty (2019) as a benchmark to which we compare our results. Anderson and Delahanty replicated and extended Strang et al., (2013). Anderson and Delahanty (2019) incorporated content differences, response format differences, and experimental procedures to successfully decrease the degree of documented discrepancies between the 2007 SES and the PRSPS, thus providing a baseline or control condition for this study.

We test the tactic-first hypothesis by utilizing a tactic-first version of the 2007 SES (the T-SES). We test the item-order hypothesis by varying the order of items on the T-SES experimentally. One experimental condition received randomized items (randomized T-SES), and one received hierarchical items starting with the least violent items (hierarchical T-SES).

Hypotheses:
1. Tactic-first hypothesis (Hypothesis 1a): We hypothesize that victimization prevalence rates will not be statistically different between the PRSPS and the hierarchical T-SES. We also hypothesize that there will be fewer discrepancies in victim identification between the hierarchical T-SES and the PRSPS compared to the traditional SES and the PRSPS in Anderson and Delahanty (2019). We repeated these analyses by tactic type to assess whether the tactic-first hypothesis is sensitive to item content: verbal coercion (Hypothesis 1b), alcohol (Hypothesis 1c), physical force (Hypothesis 1d), and multiple perpetrators (Hypothesis 1e). We expected discrepancies to be greatest for verbal coercion, consistent with prior research (H2b: Anderson & Delahanty, 2019; Strang et al., 2013; Testa et al., 2015).
2. Item-order hypothesis (Hypothesis 2a): We hypothesize that there will be no statistical difference in victimization prevalence rates between the PRSPS and the randomized T-SES. We further hypothesize that the randomized T-SES will identify more cases than the hierarchical T-SES. Finally, we hypothesize that there will be even fewer discrepancies between the randomized T-SES and the PRSPS than between the hierarchical T-SES and PRSPS. As with Hypothesis 1, we also investigated the item order hypothesis by type of tactic in Hypotheses 2b-e.
3. Consistent with prior research, we expected that even though prevalence rates would be higher for women, the degree of discrepancy would be greater for men
4. Finally, we tested for order of administration effects given that Anderson and Delahanty found that giving the 2007 SES first increased responding on the 2007 SES while suppressing responding on the PRSPS.

Method

Study 1

Participants were 265 (142 men, 122 women; one participant did not indicate a response) recruited from Amazon's Mechanical Turk (Mturk). MTurk is a platform that pairs workers with paid tasks (Litman, Robinson, & Abberbock, 2017). Participants were between the ages of 19 and 64 (M = 32.11, SD = 7.41). Participants were primarily heterosexual (87.5%), some identified as bisexual (9.1%), gay (1.5%), queer (.8%), asexual (.4%) and lesbian (.4%). Participants were largely White (78.5%), followed by Black (11.7%), Asian American (9.4%), and 2.6% as "other"; 7.9% identified as Hispanic or Latinx.

Measures

The Tactic-first Sexual Experiences Survey (Abbey et al., 2005). Because of the impact of item content, for this study, we added items compared to Abbey et al., (2005). Specifically, we added made to penetrate items and presented all participants with all items making the instrument more gender neutral. Prior work suggests using gender-neutral items does not change the psychometric properties of the SES (Anthony & Cook, 2012). Thus, the T-SES used in this study contained nine tactic items crossed by seven sexual behavior sub-items for a total of 63 items. Each item was answered on a frequency scaled response format "How many times? 0, 1, 2-5, 6-9, 10+" rather than the traditional 0, 1, 2, 3+ times. We chose to use this larger response format consistent with research suggesting that the average number of incidents is higher than the original response format (0, 1, 2, 3+) suggests, especially for those in abusive relationships (M = 12.6, Straus et al., 1996) and that response format impacts reported prevalence rates (Cuccolo & Anderson, 2020). The T-SES has demonstrated comparable validity to the 2007 SES (Abbey et al., 2019).

The Post-Refusal Sexual Persistence Scale (PRSPS: Struckman-Johnson et al., 2003). The original PRSPS contains 19 items that each list a single tactic that was used to coerce sexual behavior, see Figure 1 for example items. To match content included on the 2007 SES we added five items to the PRSPS addressing verbal coercion (3), threats of physical harm (1), and multiple perpetrator attacks (1) for a total of 24 items. We also used the same response format as used for the T-SES: 0, 1, 2-4, 6-9. 10+ and the same timeframe "since age 14". The PRSPS has demonstrated good validity in prior research; PRSPS scores were consistent with written accounts of the same victimization experiences (Struckman-Johnson et al., 2003).

Scoring. We utilized dichotomous scores consistent with prior research comparing these measures (Anderson & Delahanty, 2019; Strang et al., 2013) and suggestions that these scores are most reliable for men (Anderson et al., 2018b). We coded responses of "1" or greater on any item as endorsement of victimization. Finally, we did not compute Cronbach's alpha as a measure of reliability as this is a measure of internal consistency reliability for latent constructs; there is no latent construct within participants that is theorized to cause sexual victimization.

Procedures

Data were collected anonymously in April 2018 using Qualtrics to administer the online questionnaires. The study was advertised as "Questionnaires about Sexual Behavior 2 – Kent State University" and participants were compensated 75 cents. After completing informed consent, participants completed the demographics questionnaire and then the PRSPS and a T-SES. The order of the PRSPS and T-SES was randomly assigned. Note, for participants assigned to the randomized T-SES, even the order of the sexual behavior sub-items was randomized within each tactic. The Institutional Review Board of Kent State University approved all procedures.

Results

Data Cleaning

Any participant who completed at least one item on the T-SESs and the PRSPS were included; 41 participants were excluded who did not meet these criteria (n = 24 for PRSPS, n = 17 for T-SESs). Participants who were excluded for missing data were more likely to be women, $\chi^2(1)$ = 5.90, p = .015. In addition, we embedded one validity item within each the T-SES ("please select 0") and the PRSPS ("please select 1") to check participants attention. Participants who failed either of these items were excluded: 49 participants were excluded due to failing the validity check (44 from the PRSPS-check, 3 from the T-SESs checks). Participants who failed validity checks were not demographically different from included participants on age, race, sexual identity, gender, or race/ethnicity. The final analytic sample was 265.

Missing data for the remaining participants was minimal. Because missing data was so minimal, we used an available item analysis approach, consistent with Parent (2013)'s suggestions for this type of data. To wit, we replaced missing data with the modal value for that item. In this case, the modal value was zero, as the likelihood of endorsing any one victimization item was 6-40%. On the PRSPS, a small number of participants skipped one or more items item (0.4%). On the hierarchical T-SES, 18 participants (13.85%) skipped one or more items; however, randomized T-SES, 38 participants (60.32%) skipped one or more items. Significantly more participants skipped items on the randomized T-SES than the hierarchical T-SES ($\chi^2(1)$ = 60.77, p < .001).

Power Analysis

We computed a priori power analyses using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) and the results of Anderson and Delahanty (2019) to estimate likely effect sizes. Power analyses suggested a total sample of N = 198 to achieve Power = .80 for effects of w = .20 (small to medium) size. We also computed post-hoc power analyses to estimate achieved power. Achieved power estimates ranged from .85 to .99 suggesting our study is well powered.

Analytic Plan

The analyses below compared responses on the questionnaires within the conditions they were assigned. We computed cross-tabulations and chi-squares to test the null hypothesis that there is no relationship between prevalence rates and measure used (PRSPS and T-SES). When this overall chi-square was significant, we then tested whether the prevalence rates significantly differed. This was the primary analysis for our hypotheses. This extends Anderson and Delahanty's (2019) analysis by more clearly identifying the source of discrepancy as the initial chi-square tests for discrepancies between the observed and expected frequencies but does not specify the source of the discrepancy.

We also calculated percent agreement/discrepancy rates (percent of disagreement) and kappa to quantify the size of the discrepancy. In other words, the number of cases where the questionnaires agreed (PRSPS yes/T-SES yes or PRSPS no/T-SES no) divided by the total number of cases (N). In contrast to percent agreement, kappa estimates agreement while controlling for chance agreement. We interpreted kappa using McHugh (2012)'s guidelines wherein .6 is the minimal level of agreement.

Overall Victimization Prevalence Rates

Of the 265 participants who completed the PRSPS, 62.6% (n = 166) reported some form of sexual violence victimization. Of the 130 participants assigned to the hierarchal T-SES, 56.2% (n = 73) reported sexual victimization. Of the 135 participants who were randomly assigned to the randomized T-SES, 54.1% (n = 73) reported sexual victimization.

There was no difference in the prevalence rate of sexual victimization between the hierarchical T-SES and the PRSPS, $\chi 2(1) = 1.491$, p = .222 (Hypotheses 1) or between the randomized T-SES and the PRSPS, $\chi 2(1) = 2.680$, p = .102 (Hypothesis 2). There was also no difference between the T-SESs, $\chi 2(1) = 0.065$, p = .799. The next section details discrepancy rates and rates by tactic.

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.

H1a: Overall victimization. The hierarchical T-SES identified 73 victims while the PRSPS identified 78 victims; there were 21 discrepant cases. The overall chi-square was statistically significant, $\chi^2(1) = 58.51$, p < .001, indicating that the questionnaires significantly diverged overall even though victimization prevalence rates were not significantly different, $\chi^2(1) = 1.491$, p = .222. We next compared the discrepancy rate to Anderson and Delahanty (2019). The discrepancy rate in this study decreased (16.15%) compared to Anderson and Delahanty - 19.31%; but these rates were not significantly different ($\chi^2(1) = .71$, p = .40).

H1b. Verbal coercion. The hierarchical T-SES identified 66 cases and the PRSPS identified 70 cases; there was a total of 18 discordant cases. The overall chi-square was statistically significant ($\chi^2(1) = 67.20$, p < .001) but the proportion of cases identified by the hierarchical T-SES (66/130) compared to the PRSPS (70/130) was not ($\chi^2(1) = .25$, p = .620). The discrepancy rate did decrease in this study to 16.22%, compared to 20.06% in Anderson & Delahanty; but these rates were not significantly different ($\chi^2(1) = 1.02$, p = .313).

H1c: Substance-facilitated. The hierarchical T-SES identified 37 cases and the PRSPS identified 30 cases; there was a total of 17 discordant cases. The overall chi-square was significant ($\chi^2(1) = 57.67$, p < .001) but the proportion of cases identified by the hieratical T-SES (37/130) compared to the PRSPS (30/130) was not statistically different ($\chi^2(1) = .98$, p = .322). The discrepancy rate in this study unexpectedly increased - 13.08% - while the Anderson & Delahanty discrepancy rate was 11.59%; but these rates were not significantly different ($\chi^2(1) = .23$, p = .630). Note that Anderson and Delahanty tested a narrower range of tactics.

H1d. Physical force. The hierarchical T-SES identified 34 cases and the PRSPS identified 33 cases; there were 13 discordant cases. The overall chi-square was significant ($\chi^2(1) = 70.95$, p < .001), but the proportion of cases identified by the hierarchical T-SES (34/130) compared to the PRSPS (33/130) was not, ($\chi^2(1) = .02$, p = .887). In comparison to the Anderson and Delahanty (2019), the discrepancy rate decreased by 1.44% (10.0% in this study compared to 11.44%); however, the rates did not significantly differ ($\chi^2(1) = .23$, p = .634).

H1e. Multiple perpetrators. The hierarchical T-SES identified 14 cases and the PRSPS identified 12 cases; there were 6 discrepant cases. The overall chi-square was significant ($\chi^2(1) = 71.84$, $p < .001$); but the proportion of cases identified by the hierarchical T-SES (14/130) compared to the PRSPS (12/130) was not statistically different ($\chi^2(1) = .17$, $p = .680$). Anderson and Delahanty (2019) did not assess multiple perpetrators, so the discrepancy rates could not be compared.

Item-order Hypothesis (randomized T-SES compared to PRSPS-V among n = 135)

H2a: Overall victimization. As noted above, there was not a significant difference in prevalence rates between the hierarchical and the randomized T-SES. Kappa ranged from .59 to .73 and percent agreement ranged from 79.85 – 94.81% indicating substantial to good agreement. These values are reported individually in Table 1.

The randomized T-SES identified 73 cases while the PRSPS identified 88 cases; there were 27 discrepant cases. The overall chi-square was significant ($\chi^2(1) = 49.54$, $p < .001$; but the proportion of cases identified by the randomized T-SES compared to the PRSPS was not ($\chi^2(1) = 2.68$, $p = .102$). This indicates that the discrepancy rate is significant, even though the victimization prevalence rates were not. The rate of discrepancy between the randomized T-SES and the PRSPS was 20.15%; the rate of discrepancy was not significantly different compared to the discrepancy rate in Hypothesis 1a ($\chi^2(1) = .71$, $p = .400$) or in Anderson and Delahanty (2019; $\chi^2(1) = .05$, $p = .823$).

H2b. Verbal coercion. The randomized T-SES identified 64 cases, the PRSPS identified 81 cases; there were 27 discrepant cases. The overall chi-square was significant ($\chi^2(1) = 50.93$, $p < .001$); but the proportion of cases identified by the randomized T-SES (64/135) compared to the PRSPS (77/135) was not ($\chi^2(1) = 2.50$, $p = .114$). The rate of discrepancy was 20.0%; the rate of discrepancy was not significantly different compared to the same rate in Hypothesis 1b ($\chi^2(1) = .63$, $p = .427$) or in Anderson and Delahanty (2019; $\chi^2(1) = .00$, $p = .987$).

H2c: Alcohol-facilitated. The randomized T-SES identified 46 cases, the PRSPS identified 40 cases; there were 16 discrepant cases. The overall chi-square was statistically different ($\chi^2(1) = 72.23$, $p < .001$) while the proportion of cases identified by the T-SES (64/135) compared to the PRSPS (77/135) was not, ($\chi^2(1) = 2.50$, $p = .114$. The rate of discrepancy was 11.85%; the same rate of discrepancy was not statistically different compared to the same rate in Hypothesis 1c ($\chi^2(1) = .092$, $p = .762$) or Anderson and Delahanty (2019; $\chi^2(1) = .01$, $p = .932$).

H2d. Physical force. The randomized T-SES identified 28 cases, the PRSPS identified 36 cases; there were 18 discrepant cases. The overall chi-square was statistically significant ($\chi^2(1) = 55.05$, $p < .001$); but the proportion of cases identified by the randomized T-SES (28/135) compared to the PRSPS (36/135) was not ($\chi^2(1) = 1.31$, $p = .253$). The discrepancy rate was 13.43%; the rate of discrepancy was not significantly different compared to the same rate in Hypothesis 1d ($\chi^2(1) = .71$, $p = .399$) or in Anderson and Delahanty (2019; $\chi^2(1) = .39$, $p = .533$).

H2e. Multiple perpetrators. The randomized T-SES identified 15 cases, the PRSPS identified 12 cases; there were 7 discrepant cases. The overall chi-square was significant ($\chi^2(1) = 69.56$, $p < .001$); but the proportion of cases identified by the randomized T-SES (15/135) compared to the PRSPS (12/135) was not ($\chi^2(1) = .37$, $p = .544$). The discrepancy rate was 5.19%; this rate was not different compared to the same rate in Hypothesis 1e ($\chi^2(1) = .04$, $p = .840$).

Additional analyses. We also re-analyzed verbal coercion, substance use, and physical force data strictly controlling for item content, consistent with past research suggesting this influences discrepancy (Anderson et al., 2018a; Anderson & Delahanty, 2019). The pattern of these results was identical to those reported above.

Gender Differences

H3. We hypothesized that the victimization prevalence rates for women would be higher for each questionnaire (PRSPS, hierarchical T-SES, randomized T-SES). The prevalence rate for women was higher than men's on the PRSPS, (71.31% vs. 54.93%, $\chi^2(1) = 7.48$, $p = .006$), and the randomized T-SES, 65% vs. 44.60%, $\chi^2(1) = 5.51$, $p = .020$. But there were no gender differences for the hierarchical T-SES (64.52% vs. 48.53%, $\chi^2(1) = 2.17$, $p = .068$). There were also no gender differences in degree of discrepancies between the questionnaires in the H1a-H2e.

Order of Administration

H4. We computed chi-squares to assess whether order of administration affected reported prevalence rates on the questionnaires. We computed chi-squares for each possible pattern of administration (6) as well as coding which questionnaire was first of the three possible. None of these analyses were significant, suggesting order of questionnaire administration had no effect on reported prevalence rates.

Discussion

The way individuals are asked about their history of sexual victimization affects their responses. The goal of the present study was to examine two possible mechanisms of changes in prevalence rates in sexual violence measurement research: the tactic-first and item-order hypotheses. In this study, using the T-SES eliminated statistical differences in victimization prevalence rates between the T-SES and the PRSPS. However, there were few differences between the tactic-first and the item-order conditions (the hierarchical vs. randomized T-SES), inconsistent with predictions regarding the item-order hypothesis. Understanding the measurement of sexual violence and thereby increasing precision improves identification of victims/perpetrators, increases the rigor of research, and ultimately improves the treatment, adjudication, and prevention of sexual violence.

Tactic-first and Item-order Hypotheses

There were no statistical differences in victimization prevalence rates between the T-SES and the PRSPS in either condition. This is in contrast to prior research which has consistently found higher prevalence rates for the PRSPS (Anderson et al., 2019; Strang et al., 2013; Anderson & Delahanty, 2019) in support of our hypotheses (H1a, H2a). In other words, utilizing tactic-first items eliminated the statistical differences in victimization prevalence rates between the questionnaires. That we consistently found no difference in prevalence rates between the PRSPS and the T-SES suggests strong support for the tactic-first hypothesis. Surprisingly, we did not find support for our item-order hypotheses, in that there were no decreases in discrepancy rates between the hierarchical and randomized T-SES (e.g., tactic-first and item-order conditions, Hypothesis 2). There was also no difference in the victimization prevalence rate between the hierarchical and randomized T-SES (Hypothesis 2). We expected to find larger differences between conditions for verbal coercion, suggesting mixed support for this hypothesis (H1b, H2b). Our tactic type analyses do suggest verbal coercion is a more difficult tactic to assess as percent agreement was consistently lower for this tactic compared to others.

We purposefully used a variety of ways to assess the impact of differences in methodology for assessing victimization history. The research literature does not suggest there is any one measure of discrepancy and each measure we chose, victimization prevalence rates, percent agreement/discrepancy rates, and kappa all have slightly different practical meanings and/or statistical interpretations. For example, although there were no differences in victimization prevalence rates there were still differences in discrepancy rates between the two studies (thus the significant overall chi-square). Percent agreement/discrepancy rates were not different in comparison to Anderson and Delahanty (2019), which used the behavior-first 2007 SES, nor were there differences between conditions (item-order hypothesis). Yet, kappa estimates, a measure of agreement which corrects for chance, were improved by the use of the T-SES in both conditions in this study. At the largest value, kappa was .74 and at its lowest was .59 in this study. In Anderson and Delahanty (2019), the kappas did not exceed .63 and were particularly low for men (.44 at the lowest). This is significant in that kappa of .6 is the threshold of reliability; thus, the T-SES is a more similar construct to the PRSPS than the 2007 SES is (McHugh, 2012). Kappa estimates tended to be lower for the randomized T-SES than the hierarchical T-SES, since the item content was identical, this suggests that the ultra-randomization used in this study may cause participant confusion.

Gender and Order of Administration Issues

We were surprised to find no gender differences in discrepancy rates in this study. This suggests that the use of tactic-first items may be particularly important for capturing men's experiences. However, there is no construct validity data on the T-SES to verify that these experiences are being captured accurately. It is surprising and contrary to hypotheses that we did not find gender differences in victimization prevalence rates on the hierarchical T-SES (H3). However, given that the hierarchical T-SES rates (56.2%) were so similar to the randomized T-SES (54.1%), we suggest this finding may be a statistical artifact.

Finally, we did not find any order effects in spite of prior research documenting that these occur (Anderson & Delahanty, 2019; Anderson et al., 2018); it is possible that the use of tactic-first items also minimized order effects.

Practical Implications

Based on the evidence, it is clear that the way in which respondents are asked to recall sexual victimization greatly affects their answers. Thus, it is extremely important to clearly operationalize terms and use behaviorally specific language. When conducting in-person interviews, including in criminal justice contexts, questioning victims and perpetrators needs to be carefully considered. Using broad terms like "consent" or "rape" or even "assault" to describe experiences that participants may not label with those terms may discourage memory recall. The cognitive mechanisms behind comprehending the questions and activating memory for recall are still being discovered are personal context and emotional valence dependent (Conway & Bekerian, 1987; Kensinger & Ford, 2020). Within settings where interventions are provided, sexual violence risk reduction and prevention programs must rely on valid and reliable measures to determine if the program itself is working and to not misestimate program efficacy.

Our results harken Strang and colleagues (2013) notion that "unintentional underreporting" remains the biggest threat to validity in sexual violence assessment. To wit, the overall prevalence rates in this study were double the average prevalence rates for the 2007 SES reported in Anderson et al.'s (2019) systematic review (19.8 vs. 54.1-62.6%). This may be partially related to sample differences but is also likely related to measurement differences. Unintentional underreporting occurs when participants unknowingly leave out a sexual experience that falls within the operationalization of the questionnaire because of other factors, such as their interpretation of the terms, or the efficacy of items in cueing memory. Based on our results, this same unintentional underreporting clearly happens in victimization as well as perpetration research.

Research Implications

A This study highlights two of possibly many structural mechanisms that affect sexual violence reporting. Indeed, the tactic-first hypothesis could account for a variety of differences in prevalence rates documented in the literature (Anderson et al., 2018; DiLillo et al., 2006; Moreau et al., 2014). Further research that identifies and tests the bounds of these structural and design level influences on sexual violence questionnaires is needed. Another important element for further research is the bounds of behavioral specificity, what level of specificity is needed to cue memory retrieval, and what level is so specific as to confuse retrieval or discourage intentional disclosure (DiLillo et al., 2006).

Research that examines the impact of these changes on the psychometric properties of the questionnaires is important as well; does the use of tactic-first items or randomized items change the validity or reliability of a questionnaire? Research is also needed to understand the cognitive mechanisms behind these effects. It may be that descriptions of coercive tactics more effectively trigger the relevant autobiographical memories because the occurrence of coercive tactics is more unique or more negatively valenced than sexual behavior. Interview based research may be essential for carefully assessing the differences in how participants perceive various ways of constructing items and what memories are triggered. However, this type of research can be logistically challenging, as anonymity facilitates greater disclosure, especially for justice-involved women and other at-risk populations (Kubiak et al., 2012).

Finally, it is rather surprising that the present study did not find any significant gender differences in discrepancy rates given the large differences in sexual victimization prevalence rates and experiences between men and women (Peterson et al., 2010; Anderson et al., 2018b). This suggests that the structural changes to the questionnaires may have ameliorated heterosexist bias that may be intrinsic to the standard 2007 SES. This is another important issue for future research. Future work should place an emphasis on being particularly attentive to both understudied and at-risk groups because the dynamics of sexual violence do not always conform to the cisheteronormative model of a cisgender heterosexual woman being assaulted by a cisgender heterosexual man.

Limitations

This study illustrates the sensitivity of small measurement changes on psychometric outcomes in violence research. For example, the missing data rate for the randomized T-SES was significantly higher than

the other questionnaires. In the interest of true randomization, we randomized the order of the items, and the sub-stems underneath each item; however, given this missing data finding, it is possible that this ultra-randomization made the questionnaire more confusing for participants. Ultra-randomization may have also affected the psychometric properties, some kappa estimates and percent agreement rates were numerically lower in the item-order condition. We suggest further tests of the item-order hypothesis. We also suggest further research into the use of tactic-first or tactic-focused items; this study may not have been a true test of the tactic-first hypothesis given the structure of the PRSPS. As shown in Figure 1, sexual behavior is described in the initial instruction set even if not repeated in each item.

Our findings may also have been limited by age-related memory difficulties. The average age of our participants was in the 30s; asking participants to remember events from a decade or more prior may be quite different from asking college students to remember events spanning the prior several years.

Conclusions

This study examined two possible mechanisms of discrepancy between violence questionnaires using an experimental design. We found support for the tactic-first effect but not the item-order effect. This highlights how the structural elements of questionnaires may affect memory retrieval and thus, change rates of reported sexual violence in research and practice.

ORCID iD

RaeAnn E. Anderson https://orcid.org/0000-0001-9938-0717
Erica L. Goodman https://orcid.org/0000-0001-5819-8855
Alyssa M. Ciampaglia² https://orcid.org/0000-0001-8931-2470

References

Abbey, A., Helmers, B. R., Jilani, Z., McDaniel, M. C., & Benbouriche, M. (2019, August 8-11). Measuring men's sexual aggression against women: Evaluating robustness of findings. In Sarah L. Cook (Chair), *Evaluating the Sexual Experiences Survey-Validity and Uses [Symposia]*. Chicago, IL, Annual Meeting of the American Psychological Association. https://irp-cdn.multiscreensite.com/a5ea5d51/files/uploaded/APA2019_Program_190708.pdf

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

Anderson, R. E., Cahill, S. P., & Delahanty, D. L. (2017). Initial evidence for the reliability and validity of the Sexual Experiences Survey-Short Form Perpetration (SES-SFP) in college men. *Journal of Aggression, Maltreatment and Trauma*, *26*(6), 626–643. https://doi.org/10.1080/10926771.2017.1330296

Anderson, R. E., Cahill, S. P., & Delahanty, D. L. (2018a). Discordance between the Sexual Experiences Surveys-Short Forms and the Revised Conflict Tactics Scales in college men. *Psychology of Violence*. https://doi.org/10.1037/vio0000199

Anderson, R. E., Cahill, S. P., & Delahanty, D. L. (2018b). 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

Artime, T. M., McCallum, E. B., & Peterson, Z. D. (2014). Men's acknowledgment of their sexual victimization experiences. *Psychology of Men & Masculinity*, *15*(3), 313–323.

Basile, K. C., Smith, S., Breiding, M. J., Black, M. C., & Mahendra, R. (2014). *Sexual Violence Surveillance: Uniform Definitions and Recommended Data Elements, Version 2.0*. Retrieved from https://www.cdc.gov/violenceprevention/pdf/sv_surveillance_definitionsl-2009-a.pdf

Black, M.C., Basile, K.C., Breiding, M.J., Smith, S.G., Walters, M.L., Merrick, M.T., Chen, J., & Stevens, M.R. (2011). *The National Intimate Partner and Sexual Violence Survey (NISVS): 2010 Summary Report.* Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention.

Cascardi, M., & Muzyczyn, B. (2016). Concordant responding on the physical assault/abuse subscales of the Revised Conflict Tactics Scales 2 and Conflict in Adolescent Dating Relationships Inventory. *Psychology of Violence*, *6*(2), 303–312. https://doi.org/10.1037/a0039128

Conway, M. A., & Bekerian, D. A. (1987). Organization in autobiographical memory. *Memory & Cognition*, *15*(2), 119–132. https://doi.org/10.3758/BF03197023

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), 2012–2018. https://doi.org/10.1177/1077801210397741

Cuccolo, K., & Anderson, R. E. (2020, March 5). Impact of varying item response format on prevalence rates for sexual violence victimization and perpetration [Poster]. University of North Dakota Graduate Research Achievement Day, Grand Forks, ND, UND. https://und.edu/academics/graduate-school/_files/docs/grad-abstract-2020.pdf

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*, Vol. 36, pp. 353–373.

DiLillo, D., Fortier, M. A., Hayes, S. A., Trask, E., Perry, A. R., Messman-Moore, T., Fauchier, A., & Nash, C. (2006). Retrospective assessment of childhood sexual and physical abuse: A Comparison of scaled and behaviorally specific approaches. *Assessment*, *13*(3), 297–312. https://doi.org/10.1177/1073191106288391

Dworkin, E. R., Menon, S. V., Bystrynski, J., & Allen, N. E. (2017). Sexual assault victimization and psychopathology: A review and meta-analysis. *Clinical Psychology Review*, *56*, 65–81.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fedina, L., Holmes, J. L., & Backes, B. L. (2018). Campus sexual assault: A systematic review of prevalence research from 2000 to 2015. *Trauma, Violence, & Abuse*, *19*(1), 76–93. https://doi.org/10.1177/1524838016631129

Fisher, B. S., Daigle, L. E., Cullen, F. T., & Turner, M. G. (2003). Acknowledging sexual victimization as rape: Results from a national-level study. *Justice Quarterly*, *20*(3), 535–574. https://doi.org/10.1080/07418820300095611

Hamby, S. (2014). Intimate partner and sexual violence research: Scientific progress, scientific challenges, and gender. *Trauma, Violence, & Abuse*, *15*(3), 149–158. https://doi.org/10.1177/1524838014520723

Kensinger, E. A., & Ford, J. H. (2020). Retrieval of Emotional Events from Memory. *Annual Review of Psychology*, *71*(1), 251–272. https://doi.org/10.1146/annurev-psych-010419-051123

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., & 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.

Krebs, C. (2014). Measuring sexual victimization: On what fronts is the jury still out and do we need it to come in? *Trauma, Violence, & Abuse*, *15*(3), 170–180. https://doi.org/10.1177/1524838014521028

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Moreau, C., Boucher, S., Hébert, M., & Lemelin, J. (2014). Capturing sexual violence experiences among battered women using the Revised Sexual Experiences Survey and the Revised Conflict Tactics Scales. *Archives of Sexual Behavior*, *44*(1), 223–231. https://doi.org/10.1007/s10508-014-0345-8

National Research Council. (2014). *Estimating the Incidence of Rape and Sexual Assault: Panel on Measuring Rape and sexual Assault in Bureau of Justice Statistics Household Surveys* (C. Kruttschnitt, W. D. Kalsbeek, C. C. House, & Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Eds.). https://doi.org/10.17226/18605

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

Ramirez, I. L., & Straus, M. A. (2006). The effect of question order on disclosure of intimate partner violence: An experimental test using the Conflict Tactics Scales. *Journal of Family Violence*, *21*(1), 1–9. https://doi.org/10.1007/s10896-005-9000-4

Roediger, H., Weinstein, Y., & Agarwal, P. (2010). Forgetting: Preliminary Considerations. In S. Della Sala (Ed.), *Forgetting* (pp. 15–36). Psychology Press.

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

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. https://doi.org/10.1177/019251396017003001

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

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing Three Sources of Misresponse to Reversed Likert Items. *Journal of Marketing Research*, *45*(1), 116–131.

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

Author Biographies

RaeAnn E. Anderson (Ph.D., University of Wisconsin-Milwaukee, 2015) is currently an Assistant Professor in Clinical Psychology at the University of North Dakota. She completed her post-doctoral training at Kent State University. Her research interests are understanding basic behavioral processes in sexual violence in order to inform sexual assault risk reduction and prevention programs, respectively.

Erica L. Goodman (Ph.D., University of North Dakota, 2020) has a doctoral degree in clinical psychology and is currently a post-doctoral fellow at the Depression and Anxiety Specialty Clinic (DASC) of Chicago. Her thesis used structural equation modeling to assess comorbid indicators of non-suicidal self-injury and eating disorder psychopathology. Her dissertation is investigating facets of loss-of-control eating in a non-clinical sample. Erica continues to do clinical work and research surrounding issues related to eating pathology and, recently, sexual violence perpetration and victimization.

Alyssa M. Ciampaglia (M.S., Saint Joseph's University, 2017) has a master's degree in experimental psychology. After defending her thesis and completing her degree, Alyssa began working as a clinical research coordinator at the Sleep Center at The Children's Hospital of Philadelphia. She presently coordinates for multiple NIH funded research projects. Her research interests are in promoting the health and wellbeing of gender and sexual minorities and advocating for intersectionality as well as diversifying methodology in clinical psychology.

Table 1. Percent Agreement, Prevalence Rates, and Kappa Between the Two Measures by Condition and Type of Victimization

	Overall victimization			Verbal coercion			Substance use facilitated		
	% agreement	Prevalence rates ± (%)	κ	% agreement	Prevalence rates ± (%)	κ	% agreement	Prevalence rates ± (%)	κ
	Overall victimization			*Verbal coercion*			*Substance use facilitated*		
A & D (2019) (2007 SES vs. PRSPS)	80.69*	38.6 vs. 49.0*	κ = .61	79.90*	28.2 vs. 42.6*	κ = .57	88.41*	13.4 vs. 17.8*	κ = .63
Hypothesis 1 (hierarchical T-SES vs. PRSPS)	83.85*	56.2 vs. 60.0	κ = .67	86.05*	50.8 vs. 53.8	κ = .72	86.92*	28.5 vs. 23.1	κ = .66
Hypothesis 2 (randomized T-SES vs. PRSPS)	79.85*	54.1 vs. 65.2	κ = .59°	80.00*	47.4 vs. 57.0	κ = .60	88.15*	34.0 vs. 29.6	κ = .73
				Physical force			*Multiple perpetrator*		
A & D (2019) (2007 SES vs. PRSPS)				88.56*	8.9 vs. 15.2*	κ = .54°	—	—	—
Hypothesis 1 (hierarchical T-SES vs. PRSPS)				90.00*	26.2 vs. 25.4	κ = .72	95.35*	10.8 vs. 9.2	κ = .74
Hypothesis 2 (randomized T-SES vs. PRSPS)				86.57*	20.7 vs. 26.7	κ = .63	94.81*	11.1 vs. 8.9	κ = .71

Note: ± SES prevalence rates always listed first. ° indicates kappa values fall below the "moderate" interpretive range. * indicates a statistical difference in discordance rates or prevalence rates. A & D (2019) = Anderson & Delahanty, 2019. SES = Sexual Experiences Survey, PRSPS = Post-Refusal Sexual Persistence Scale