

Marquette University

**e-Publications@Marquette**

---

Psychology Faculty Research and Publications

Psychology, Department of

---

2020

## **Do Instructions Intended to Reduce False Positives Improve the Measurement of Physical Partner Violence Victimization Among Adolescents and Young Adults?**

Kelli Sargent

Kristen Yule

Kate Bridges

Ernest N. Jouriles

John Grych

Follow this and additional works at: [https://epublications.marquette.edu/psych\\_fac](https://epublications.marquette.edu/psych_fac)



Part of the [Psychology Commons](#)

---

Marquette University

**e-Publications@Marquette**

***Psychology Faculty Research and Publications/College of Arts and Sciences***

***This paper is NOT THE PUBLISHED VERSION; but the author's final, peer-reviewed manuscript. The published version may be accessed by following the link in the citation below.***

*Psychology of Violence*, Vol. 10, No. 2 (2020): 152-161. [DOI](#). This article is © American Psychological Association and permission has been granted for this version to appear in [e-Publications@Marquette](#). American Psychological Association does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from American Psychological Association.

# Do Instructions Intended to Reduce False Positives Improve the Measurement of Physical Partner Violence Victimization Among Adolescents and Young Adults?

Kelli S. Sargent

Southern Methodist University

Kristen Yule

Marquette University

Kate Bridges and Ernest N. Jouriles

Southern Methodist University

John H. Grych

Marquette University

## Abstract

**Objective:** We examine how instructions to exclude behaviors occurring in playful or joking contexts influence the measurement of physical partner violence victimization. Specifically, we demonstrate how such instructions influence the prevalence and validity of self-reported victimization. **Method:** Study 1 used a Think Aloud procedure to evaluate thoughts of college students ( $n = 451$ ) reporting victimization experiences that occurred during high school. Participants were randomized to report on physical partner violence victimization with or without instructions to exclude playful acts. Study 2 experimentally evaluated whether the instructions affect the criterion validity of victimization scores with measures of depressive symptoms and emotion regulation among first-year college students ( $n = 615$ ). Study 3 sought to replicate findings from Study 2 in a community sample of 18–25-year-olds ( $n = 398$ ), using alternative violence items, response formats, and a different recall period. Study 4 utilized a short-term longitudinal design to replicate the pattern of findings from Studies 2 and 3, and examine how instructions influence self-reports of revictimization over a 2-month follow-up among first-year college students ( $n = 887$ ). Study 5 presents a single-paper meta-analysis that synthesizes prevalence rates across these four studies. **Results:** Overall, instructions designed to eliminate aggressive acts in joking contexts did not consistently influence prevalence rates of victimization or improve criterion validity over standard instructions. **Conclusions:** Instructions designed to exclude behaviors occurring in playful or joking contexts do not necessarily produce more valid self-reports of physical partner violence victimization, as compared with standard instructions.

The measurement of intimate partner violence (IPV) in adolescent and adult relationships has a long, controversial history. Violence is defined as intentional acts perpetrated to harm another (Hamby, 2016a, 2017; Hamby & Grych, 2013), and is typically measured using questionnaires that assess how often certain behaviors (e.g., pushing, hitting) occurred during a discrete period of time (e.g., Conflict in Adolescent Dating Relationships Inventory [CADRI], Wolfe et al., 2001; Conflict Tactic Scales 2 [CTS2], Straus, Hamby, Boney-McCoy, & Sugarman, 1996). These measures, sometimes referred to as “act” scales, ask about specific, concrete behaviors in an attempt to reduce the ambiguity and subjectivity of terms such as *abuse* and *violence*. Although widely used to investigate the prevalence and correlates of IPV, significant concerns have been raised about their validity (Hamby, 2016b). In particular, act scales have been criticized for disregarding the context in which behaviors occur (Frías, 2016; Hamby, 2017). Actions such as pushing or hitting can represent attempts to hurt or control a romantic partner, but also can occur when couples are roughhousing, joking, or being playful. Behavior intended to harm has a range of potential adverse psychological and physical effects, but the impact of the same behavior in a playful interaction may be benign. Endorsing a behavior on an act scale of violence that was not intended to harm may be considered a false positive that would inflate estimates of “true” violence victimization, reduce the validity of act scales, and potentially produce inaccurate conclusions about IPV (Hamby, 2016a, 2016b; Lehrner & Allen, 2014). This article presents a series of studies examining how instructions to exclude aggressive behaviors occurring in playful contexts influence the rates and validity of self-reported physical victimization in adolescents’ and young adults’ romantic relationships.

Research with adolescents and young adults shows that a substantial proportion of reported acts of violence occur in a playful context (Arriaga, 2002; Foshee, Bauman, Linder, Rice, & Wilcher, 2007; Frías, 2016; Jouriles, Garrido, Rosenfield, & McDonald, 2009). For example, approximately half of a sample of high school students in Spain (Fernández-González, O’Leary, & Muñoz-Rivas, 2013) and more than half of a sample of female U.S. undergraduate students (Lehrner & Allen, 2014) indicated that violent behaviors reported on act scales occurred while joking around or roughhousing. To address the “false positives” problem, some investigations have modified the instructions on self-report questionnaires, either by directing respondents not to report behaviors that occurred when joking or being playful or by asking respondents to identify the context in which reported acts occurred (Fernández-González et al., 2013; Jouriles et al., 2009). Others have designed interviews to assess respondents’ perceptions of the motives for IPV that occurred, and subsequently coded responses to identify which, if any, instances were characterized by joking or roughhousing (Foshee et al., 2007; Lehrner & Allen, 2014). These studies show that rates of IPV victimization among high-school adolescents are markedly lower after aggressive acts that occurred in playful contexts were excluded (Arriaga, 2002; Fernández-González et al., 2013; Foshee et al., 2007; Frías, 2016; Jouriles et al., 2009; Lehrner & Allen, 2014).

Hamby (2016b) recently argued that false positives differentially impact rates of male- versus female-perpetrated violence because females are more likely to push, hit, and grab their partners in a playful way than males. Act scales generally produce either similar rates of violence for males and females or show that females engage in more aggression toward romantic partners than do men (Archer, 2000). These findings conflict with other sources of data showing that males are more violent than females (Catalano, 2012; Hamby, Finkelhor, & Turner, 2012). Arguably, more false positives in self-reports of female-perpetrated IPV would overestimate female aggression and negatively affect validity (Hamby, 2016b).

In most studies attempting to reduce false positives, rates of victimization changed similarly for males and females, with males reporting equal or greater amounts of victimization than females (Fernández-González et al., 2013; Frías, 2016; Jouriles et al., 2009). An exception is a study conducted by Hamby (2016b), in which an experimental methodology was used. Participants were randomized to report physical violence victimization using four items with standard instructions or instructions to exclude acts perpetrated in a joking context. In both conditions, there was also a fifth item assessing sexual violence victimization that did not ask about a joking context. Hamby found that females reported more victimization than did men when they completed a questionnaire in which each physical violence item was prefaced by the qualifier, “Not including horseplay or joking around.” However, although the qualifier produced higher rates of female than male victimization, it was not because the qualifier *reduced* reports of male victimization, as predicted. Rather, females reported more victimization in the joking qualifier condition than they did when responding to the standard instructions, whereas male reports of victimization were similar in both conditions.

If measures that exclude playful behavior reduce false positives, scores yielded by these measures should be more valid. For example, these scores should correlate more strongly with other variables reliably related to victimization, compared with scores from measures that include false positives. Surprisingly, this question only has been examined directly in a single study. Jouriles and colleagues (2009) compared the criterion validity of an act scale that instructed respondents to exclude behavior

that occurred while playing around with a version using the standard instructions. The modified version led to fewer reports of victimization, but was not more strongly related to psychological distress and anxiety than the standard version without the qualifier.

## The Current Research

We investigated the effects of instructions used in self-report measures of physical IPV victimization on prevalence estimates and criterion validity in five studies using either college students or community emerging adults as participants. Study 1 examines the thought processes of participants completing a victimization scale with and without instructions to exclude behaviors that occurred during “horseplay or joking around” using a “Think Aloud” procedure. Studies 2, 3, and 4 examine whether adding this qualifier influences prevalence rates of victimization and improves criterion validity compared with a measure of the same behaviors without the qualifier. Depressive symptoms and emotion regulation difficulties are two well-documented and common correlates of physical violence victimization in intimate relationships (Devries et al., 2013; Ehring & Quack, 2010; Exner-Cortens, Eckenrode, & Rothman, 2013; Kaura & Lohman, 2007), and difficulty regulating emotional responses also has been implicated as a risk factor for revictimization (Noll & Grych, 2011). Studies 2, 3, and 4 evaluate whether the addition of a joking/horseplay qualifier to violence items strengthens the association between violence victimization and these criterion variables. Study 4 also evaluates how use of the qualifier may influence correlations between victimization experiences over time. Previous victimization experiences are theorized to alter how individuals approach and respond to later situations, increasing risk for future victimization (Noll & Grych, 2011), and victimized youth are at risk for future victimization (Exner-Cortens et al., 2013; Jouriles, Choi, Rancher, & Temple, 2017). Because studies on violence victimization often use different recall periods, violence items, and response formats, these were varied across the four studies. Study 5 quantitatively summarizes these patterns of victimization by sex and condition by conducting a single-paper meta-analysis (McShane & Böckenholt, 2017).

## Study 1

In Study 1, we hypothesized that (1) participants receiving instructions to exclude acts that occurred during “horseplay or joking around” would report lower prevalence of IPV than those receiving instructions that do not include this qualifier, and (2) among reports of victimization, there would be fewer false positives in the qualifier condition than using standard instructions.

## Method

### *Participants*

Participants ( $n = 451$ ) were recruited from two midsize U.S. universities: one in the Midwest (M;  $n = 189$ ) and one in the South (S;  $n = 262$ ). Participants were recruited from undergraduate psychology subject pools and offered extra course credit for participating. Participants did not differ by university on sex, race, or ethnicity,  $ps > .56$ ; however, participants at M were slightly younger ( $M_{age} = 18.16$ ) than at S ( $M_{age} = 19.97$ ). The sample was predominantly female (73%;  $n = 331$ ) and White (75%;  $n = 336$ ), and included participants who identified as Asian (7%;  $n = 33$ ), “other” (6%;  $n = 29$ ), Black (4%;  $n = 20$ ), American Indian/Alaska Native (1%;  $n = 5$ ), “more than one race” (6%,  $n = 25$ ), and Native Hawaiian or Other Pacific Islander (<1%;  $n = 3$ ). In a separate ethnicity question, 77 participants

identified as Hispanic (17%). Most participants ( $M_{age} = 19.22, SD = 1.76$ ) were first-years (57%;  $n = 259$ ), 95 were sophomores (21%), 54 were juniors (12%), and 43 were seniors (10%).

### *Procedures*

All procedures were approved by the institutional review board at both universities. In a private room, participants completed a questionnaire assessing physical victimization in romantic relationships in high school. Before completing the measure, participants were randomly assigned to one of two conditions. The "Standard" condition ( $n = 223$ ) included items without the phrase "Not including horseplay or joking around," and the "Qualifier" condition ( $n = 228$ ) included this phrase before each item. Participants in both conditions were instructed to "think aloud" as they filled out the questionnaire, speaking aloud "all the things that go through your mind as you're choosing your answer" (Ericsson & Simon, 1980). Participants were told that the study was "trying to learn about how people approach answering questions like these, and how they decide on their final answer." An interviewer walked through three practice questions with each participant, and left the room for the participant to respond to all study questions. All responses were audio recorded for coding.

### *Measures*

Participants completed four items from the five-item Partner Victimization Scale (PVS; Hamby, 2016b). The items without the qualifier are as follows: "My partner threatened to hurt me and I thought I might really get hurt," "My partner pushed, grabbed, or shook me," "My partner hit me," and "My partner beat me up." We did not include the fifth PVS item, which assesses sexual violence, as it does not include the qualifier in the original PVS. A sample item with the qualifier was "Not including horseplay or joking around, my partner pushed, grabbed, or shook me." Respondents select *yes* or *no* for each item (Hamby, 2016b). For each item, two sets of independent coders at each university rated if the aggressive act was noted to have occurred (*yes* = 1, *no* = 0), and whether it was described as occurring when joking around or roughhousing (*not qualified* = 0, *qualified as joking* = 1). Participants who reported at least one "yes" response for an aggressive act that occurred while joking were coded as reporting a false positive (interrater agreement  $K = .92$  and  $.91$  at university S and university M, respectively).

### *Results*

Participants assigned to the two conditions did not differ on sex, ethnicity, race, age, or year in school,  $ps > .16$ . All 451 participants provided complete data on victimization experiences. For Hypothesis 1, a post hoc power analysis calculated using G\*Power 3.1 indicated power = .88 to detect a small-to-medium effect ( $w = .15$ ). Forty-two (9%) participants reported experiencing at least one act of IPV victimization during high school: 25/223 (11%) in the Standard condition, and 17/228 (7%) in the Qualifier condition. More participants at S ( $n = 31$ ) reported victimization than did students at M ( $n = 11$ ),  $\chi^2(1) = 4.81, p = .028$ . Thus, a logistic regression was used to determine if victimization (any vs. none) differed by condition (Standard vs. Qualifier) when including university (M vs. S) as a covariate. Victimization did not differ by condition,  $b = -0.42, p = .21$ . In the Standard condition, 16 females and nine males reported victimization. In the Qualifier condition, 13 females and four males reported victimization.

Only 42 participants reported experiencing at least one act of IPV victimization, and a post hoc power analysis for Hypothesis 2 indicated low power of 0.26. Nine of the 25 (36%) participants reporting

victimization in the Standard condition indicated that at least one act occurred while joking around (i.e., were classified as false positives). Three of these students were female and the remaining six were male. Three of the 17 (18%) participants in the Qualifier condition who reported physical victimization described the behavior as occurring in a joking context, two of which were female and the other male. Fisher's exact test indicated that false positives did not differ by condition among participants reporting victimization,  $p = .17$ .

Examples of comments made by participants who reported experiencing victimization in a joking or playful context in the Standard condition included the following: "We wrestled sometimes. Nothing really aggressive . . . I was pushed a couple times, but more playful pushing and I did some pushing, too . . . So, I would say yes? I mean technically, yeah, she did push me, but more playful," "Yeah, but not in a real, serious manner. Definitely just joking around. I thought she was joking, she thought she was joking, everything went well," and

We would jokingly push in a non-hurtful way. And grabbed and like shaking in ways that . . . were more playful and we understood that it was meant to be a non-combative thing, so . . . I guess yes? But with the justification of it in an entirely nonthreatening and more cutesy kind of way.

## Discussion

Study 1 examined participants' thought processes as they responded to questions on victimization experiences in romantic relationships while they were in high school. Victimization did not differ by condition, and in fact, acts of violence that occurred in joking or playful contexts were reported in *both* conditions (36% and 18% of victimization in the Standard and Qualifier conditions, respectively); although prevalence of playful aggression was twice as high in the Standard condition, this difference was not statistically significant, perhaps due to the fact that the analysis was not sufficiently powered. Thus, results provide some indication that act scales capture "false positive" reports of physical IPV that occur in a playful context, but use of the qualifier "Not including horseplay or joking around," before each violent act does not completely eliminate reports of playful aggression.

It is interesting to consider why playful acts of aggression were still reported by three participants, even when explicitly instructed to exclude such acts. In the current sample, these instances were characterized by confusion regarding the perpetrator's motives. That is, participants reported feeling unsure about whether their partner was joking or described acts of aggression considered playful by one partner and not the other. This suggests that participants may struggle to contextualize relationship behavior, despite attempts to clarify aggressive acts using a qualifier.

## Study 2

Study 2 examines victimization rates using the standard instructions versus the qualifier, as well as associations between victimization and both depressive symptoms and emotion regulation difficulties. Study 1 suggests that scales that do not explicitly instruct respondents to exclude acts that occur in joking contexts capture more of these acts, compared with scales with standard instructions. If these acts are indeed false positives, the use of the qualifier "Not including horseplay or joking around" should improve the criterion validity of violence victimization scores yielded by IPV scales. Thus, we hypothesize as follows: (1) There will be fewer reports of victimization in the Qualifier than Standard

condition, and (2) associations of victimization with depressive symptoms and emotion regulation difficulties will be stronger in the Qualifier than Standard condition.

## Method

### *Participants*

Participants ( $n = 624$ ) were recruited from university M and university S. This sample did not overlap with the sample from Study 1. At M, 650 participants were randomly selected from the entire first-year class ( $N = 1,547$ ) and invited via e-mail to take part in the study; 282 (43%) participated and received a \$10 e-gift card that could be used at a wide range of retailers. Participants at S ( $n = 342$ ) were recruited from required first-year Wellness courses, and offered course credit for participating. Alternative assignments were offered for those who chose not to participate. M had a greater proportion of female participants than S,  $\chi^2(1) = 13.44, p < .001$ , but did not differ from S on any other demographic variables,  $ps > .12$ . The overall sample (43% male) was predominantly White (73%), and also included Asian (13%), Hispanic (5%), "more than one race" (5%), Black (3%), Native Hawaiian or Other Pacific Islander (<1%), and American Indian/Alaska Native (<1%) participants. Two participants did not report their race. Age ranged from 18 to 28 years ( $M = 18.31, SD = 0.69$ ).

### *Procedures*

Participants at both universities were informed that study participation involved completing questionnaires as part of a larger study on how high school experiences influence college behavior. Participants completed all questionnaires using the Qualtrics survey platform. Participants were randomized within university to report high school victimization using the instructions used by standard act scales ( $n = 310$ ) or with the "Not including horseplay or joking around" qualifier preceding each item ( $n = 314$ ).

### *Measures*

#### *Physical victimization*

Participants completed the same victimization measure used in Study 1, reporting *yes* or *no* to each of the four items. Coefficient  $\alpha$  in the current sample was .64 for the standard act and .73 for the scale with the qualifier. Test-retest correlations for a 9-week period for physical assault CTS items has been documented in previous research ( $r = .68$ ; Vega & O'Leary, 2007). Items were summed, with scores ranging from 0 to 4.

#### *Depressive symptoms*

Participants completed the Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977). This 20-item scale asks respondents to rate how often they experienced depressive symptoms in the past week on a scale from 0 (*rarely or none of the time*) to 3 (*most or all of the time*). Coefficient  $\alpha$  in the current sample was .90. CES-D scores relate positively to other self-report measures of depression (Radloff, 1977).

#### *Difficulties in emotion regulation*

Participants completed seven items from the Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004), rating how true each item is about them on a scale from 1 (*not true about me*) to 4 (*mostly true about me*). Sample items include "When I'm upset, it takes me a long time to feel better,"



and “When I’m upset, I feel out of control.” Coefficient  $\alpha$  was .71 in the current sample. DERS scores relate to social problems and internalizing symptoms in college samples (Kaufman et al., 2016).

## Results

Two participants in the Qualifier condition did not complete the IPV victimization scale, four students did not report their sex, and data from three participants were excluded due to reporting patterns indicating that they were not paying attention; thus, the final analytic sample was  $n = 615$ . Post hoc power analyses indicated sufficient power of 0.96 to detect a small-to-medium effect ( $f = .15$ ) for Hypothesis 1, and power of 0.96 to detect a medium effect ( $q = 0.3$ ) for Hypothesis 2. Participants at M reported greater victimization ( $M = 0.21$ ) than students at S ( $M = 0.12$ ),  $F(1, 618) = 3.94, p = .048$ . Participants did not differ by condition on any demographic variables,  $ps > .07$ .

### Hypothesis 1

Fifty-seven of the 615 participants (9%) reported experiencing at least one act of physical violence victimization in high school: 27/307 (9%) in the Standard condition and 30/308 (10%) in the Qualifier condition. Victimization did not differ by condition,  $\chi^2(1) = 0.16, p = .67$ . In the Standard condition, 19 females and eight males reported victimization. In the Qualifier condition, 16 females and 14 males reported victimization.

### Hypothesis 2

As shown in Table 1, participants’ reports of victimization in both the Standard condition and Qualifier condition were associated with depressive symptoms, but difficulties in emotion regulation was only related to reports of victimization in the Standard condition. Fisher  $r$ -to- $z$  transformations were computed to compare the magnitude of the association between victimization and the criterion variables across conditions; no differences emerged. We also examined whether males and females differed in the strength of the associations between the two versions of the victimization measure and the criterion variables; no sex differences were found. This pattern of results did not change when controlling for university.

Table 1 Correlations Between Physical Victimization and Criterion Variables by Condition and Sex

Criterion variable	Study 2			Study 3		
	Standard	Qualifier	$z$	Standard	Qualifier	$z$
Depressive Symptoms						
Total sample	.21**	.13*	1.02	.31**	.25**	0.64
Female	.22**	.17*	0.48	.31***	.18*	1.13
Male	.18*	.07	.90	.34*	.48***	0.91
Difficulties in emotion regulation						
Total sample	.18**	.12	0.76	.25**	.23**	0.21
Female	.17*	.20*	-0.29	.21*	.13	0.67
Male	.15	-.01	1.30	.38**	.38**	0.00

*Note.* Intimate partner violence was scored continuously, with scores ranging from 0 to 4 for Study 2 and 0 to 16 for Study 3. The  $r$ -to- $z$  transformations were used to test the difference between correlation coefficients. “Total sample” correlations were analyzed with both males and females combined, reported by condition.

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

## Discussion

Study 2 examined the criterion validity of PVS physical violence items (Hamby, 2016b) with and without the “joking or horseplay” qualifier. Similar to Study 1 and past research (Vagi, O’Malley Olsen, Basile, & Vivolo-Kantor, 2015), 9% of participants reported physical victimization in a romantic relationship during high school. IPV victimization was similar for both forms of the measure. In addition, both forms were associated with depressive symptoms, as well as difficulties in emotion regulation among females. The magnitude of the associations did not differ across conditions. Thus, the findings indicate that including the qualifier “Not including horseplay or joking around” before violent acts does not significantly influence prevalence rates of physical IPV victimization or the magnitude of correlations between physical IPV victimization and either depressive symptoms or emotion regulation difficulties. However, findings from Study 2 are restricted to reports provided by college students of physical IPV victimization that occurred during high school, and to the four physical violence victimization items from Hamby’s (2016b) PVS.

## Study 3

Study 3 was designed to test if results from Study 2 replicated using a different sample (participants recruited from Amazon’s Mechanical Turk [MTurk] vs. participants recruited from universities), a different time frame for reporting on victimization (past 12 months vs. during high school), and a different set of physical violence items (items from the CADRI; Wolfe et al., 2001) with each item having a continuous, rather than a dichotomous (yes/no), response scale. Additionally, the qualifying statement preceding each behavior was modified slightly to “Not including roughhousing and joking around.” Similar to Study 2, we hypothesized as follows: (1) There will be fewer reports of victimization in the Qualifier than Standard condition, and (2) associations of victimization with depressive symptoms and emotion regulation difficulties will be stronger in the Qualifier than Standard condition.

## Method

### Participants

Participants ( $n = 424$ ) were recruited from MTurk, one of the largest crowdsourcing markets used to recruit people to complete brief online tasks for small payments. MTurk is widely used for research and tends to recruit a greater diversity of participants as compared with university student samples. A brief description of the study, including estimated duration and compensation, was posted on the MTurk website with a survey link, advertised to MTurk workers residing in the United States between the ages of 18 and 25. Participants were predominately female (68%) and White/Caucasian (64%), and included participants who identified as Black/African American (11%), Asian (10%), Hispanic (8%), “more than one race” (5%), American Indian/Alaska Native (1%), and “other” (1%). Participants age ranged from 18 to 25 years ( $M = 22.78$ ,  $SD = 1.86$ ).

## Procedure

Respondents completed questionnaires via Qualtrics, which randomized participants to report victimization using the “Not including roughhousing and joking around” qualifier ( $n = 196$ ) or standard format ( $n = 202$ ). MTurk payments rarely exceed \$1.00 an hour for completed tasks (Paolacci, Chandler, & Ipeirotis, 2010), and work quality tends to be more variable as payment increases (Wu, Corney, & Grant, 2014). Participants were compensated \$0.25 for completing the survey, which took an average of 7 min ( $SD = 4$ ).

## Measures

### *Physical victimization*

Participants reported physical victimization in romantic relationships over the past 12 months using four items from the CADRI Physical Abuse subscale (Wolfe et al., 2001). Specifically, participants reported how often these behaviors “occurred with a romantic partner in the past year, including exes,” using the CADRI response scale from 0 (*never*) to 4 (*four times or more*). Items assessed the extent to which a partner “threw something at me,” “kicked, hit, or punched me,” “slapped me or pulled my hair,” and “pushed, shoved, or shook me.” For participants randomized to the qualifier condition, the “Not including roughhousing and joking around” qualifier was placed before each item. Coefficient  $\alpha$  was .92 with the qualifier and .91 using the standard act form. Previous research demonstrates moderate test–retest correlations over a 2-week period of the CADRI Physical Abuse subscale ( $r = .64$ , Wolfe et al., 2001). Items were summed as an index of victimization frequency.

### *Depressive symptoms*

As in Study 2, participants completed the CES-D (Radloff, 1977). Coefficient  $\alpha$  was .94.

### *Difficulties in emotion regulation*

Participants completed 18 items from the Difficulties in Emotion Regulation Scale—Short Form (DERS-SF; Kaufman et al., 2016), rating how often each item applied (1 = *almost never* to 5 = *almost always*). The DERS-SF has demonstrated similar psychometric properties to the full measure (Kaufman et al., 2016). Coefficient  $\alpha$  in the current sample was .92.

## Results

Data from 26 participants were excluded due to reporting patterns indicating that they were not paying attention, resulting in an analytic sample of  $n = 398$ . Post hoc power analyses indicated power of  $>0.99$  to detect a medium effect for Hypothesis 1, and power of 0.90 to detect a medium effect for Hypothesis 2. Participants in the two experimental conditions did not differ on sex, ethnicity, race, age, or years of education,  $ps > .30$ .

### *Hypothesis 1*

Across the 398 participants, 114 (29%) reported at least one act of physical IPV victimization within the past year: 55 (27%) in the Standard condition and 59 (30%) in the Qualifier condition. Mean victimization scores ( $M = 1.47$ ,  $SD = 3.22$ ) did not differ by condition,  $F(1, 396) = .10$ ,  $p = .76$ . In the Standard condition, 41 females (28%) and 18 males (33%) reported victimization; in the Qualifier condition, 26 females (20%) and 29 males (42%) reported victimization. A two-way analysis of variance of victimization scores revealed an overall sex difference, with males ( $M = 2.38$ ) reporting higher victimization scores than females ( $M = 1.07$ ),  $F(1, 394) = 13.88$ ,  $p = .001$ , but scores did not differ by

condition,  $p = .54$ . The sex by condition interaction also was nonsignificant,  $p = .10$ . Within the Qualifier condition, victimization differed by sex,  $F(1, 194) = 15.19$ ,  $p = .001$ , with males ( $M = 2.72$ ) reporting more victimization than females ( $M = .87$ ). There was no sex difference within the Standard condition,  $p = .15$ .

### *Hypothesis 2*

Correlations of physical victimization with the criterion variables are presented in Table 1. Both forms of the measure were associated with depressive symptoms and emotion regulation difficulties. Fisher  $r$ -to- $z$  transformations resulted in no differences between conditions on either criterion variable, even when examining males and females separately.

### Discussion

Study 3 examined the same research questions as Study 2 in a different sample and using a different measure of physical IPV victimization. Rates of reported victimization were considerably higher in Study 3 than in Study 2, but as in Study 2, use of the qualifier did not significantly influence rates of reported victimization. However, it should be acknowledged that significantly more females reported victimization than males when responding to the standard instructions. This sex difference stands in contrast to much of the existing scientific literature that relies heavily on act scales (Archer, 2000). The results from Study 3 also indicated that both forms of the measure had similar associations with criterion variables. These results are similar to those of Study 2. Together the results from Studies 2 and 3 suggest that criterion validity, with depressive symptoms and difficulties in emotion regulation, does not differ whether or not respondents are instructed to exclude behaviors that occur in a playful context.

### Study 4

Studies 2 and 3 evaluated criterion validity using measures of depressive symptoms and emotion regulation difficulties. Another well-documented consequence of victimization experiences is increased risk for future victimization (Exner-Cortens et al., 2013; Jouriles et al., 2017). A short-term longitudinal design was employed to evaluate how the “Not including horseplay or joking around” qualifier may influence correlations between victimization experiences over time and examine victimization rates over a 2-month period. Previous research has documented acts of physical IPV within time frames as brief as 2 months (Jouriles, McDonald, Garrido, Rosenfield, & Brown, 2005); thus, we estimated this interval would effectively provide revictimization estimates within an academic semester. Previous criterion associations with depressive symptoms and emotion regulation difficulties were also tested. We hypothesized as follows: (1) There will be fewer reports of victimization in the Qualifier than Standard condition, (2) associations of victimization with depressive symptoms and emotion regulation difficulties will be stronger in the Qualifier than Standard condition, and (3) associations of victimization at Time 1 with victimization at Time 2 will be stronger in the Qualifier than Standard condition.

### Method

#### Participants

Participants ( $n = 994$ ) were recruited from first-year required Wellness courses at university S and offered course credit for participating. Alternative assignments were offered for those who chose not

to participate. The overall sample (47% male) was predominantly White (81%) and Non-Hispanic (90%; coded separately from Race). Participants also identified as Asian (9%), “more than one race” (6%), Black (3%), Native Hawaiian or Other Pacific Islander (<1%), and American Indian/Alaska Native (<1%) participants. Eight participants did not report their race. Participant age ranged from 18 to 25 years ( $M = 18.22$ ,  $SD = 0.55$ ).

## Procedures

Participants were informed that study participation involved completing questionnaires at two time points. Participants completed all questionnaires online using Qualtrics survey platform. Before baseline questionnaires, participants were randomized to report victimization with a “Not including horseplay or joking around” qualifier ( $n = 500$ ), or using the standard format ( $n = 473$ ). Approximately 2 months later, participants were sent electronic links to complete a follow-up survey.

## Measures

### *Victimization*

Participants responded to same four CADRI items described in Study 3 (Wolfe et al., 2001) at baseline and 2-month follow-up assessments. At baseline, participants were instructed to report how often the behaviors occurred in the past year. At follow-up, participants reported victimization within the past 2 months. Coefficient  $\alpha$  was .80 at baseline and .66 at 2-month follow-up in the standard act condition, and .86 at baseline and .80 at follow-up for the scale in the qualifier condition. IPV victimization was summed.

### *Depressive symptoms*

At baseline, participants completed the CES-D (Radloff, 1977), rating how often they experienced depressive symptoms in the past week. Coefficient  $\alpha$  was .90.

### *Difficulties in emotion regulation*

At baseline, participants completed 18 items from the Difficulties in Emotion Regulation Scale—Short Form (DERS-SF; Kaufman et al., 2016), rating how often each item applied (1 = *almost never* to 5 = *almost always*). Coefficient  $\alpha$  was .75.

## Results

The data from 107 students were excluded due to reporting patterns indicating that they were not paying attention, and one student did not report their sex, resulting in a baseline sample of  $n = 886$ , of which 856 provided follow-up IPV data (retention = 97%). Post hoc power analyses indicated power of 0.84 to detect a small effect size ( $f = .10$ ) for Hypothesis 1, and power of >0.99 to detect a medium ( $q = 0.3$ ) effect for Hypotheses 2 and 3. Participants in the two experimental conditions did not differ on any demographic variables, including sex, ethnicity, race, or age,  $ps > .15$ .

### *Hypothesis 1*

At baseline assessment, 67 students (8%) reported experiencing victimization in the past 12 months: 34 (21 males, 13 females) had been randomized to the Standard condition, and 33 (17 males, 16 females) had been randomized to the Qualifier condition. At follow-up, 42 students reported at least one act of victimization in the past 2 months: 23 (10 males, 13 females) were in Standard condition and 19 (10 males, nine females) were in the Qualifier condition. A two-way analysis of variance (Condition  $\times$

Sex) of baseline victimization scores revealed an overall sex difference, with males ( $M = 0.39$ ) reporting higher victimization scores than females ( $M = 0.17$ ),  $F(1, 885) = 6.95$ ,  $p = .009$ , but scores did not differ by condition,  $p = .23$ . The sex by condition interaction was also nonsignificant,  $p = .77$ . No sex or condition differences emerged at 2-month follow-up,  $ps > .08$ . Within the Standard condition, victimization differed by sex at baseline,  $F(1, 425) = 6.45$ ,  $p = .009$ , with males ( $M = 0.35$ ) reporting more victimization than females ( $M = 0.10$ ). There was no sex difference at 2-month follow-up,  $p = .41$ , and no sex differences within the Qualifier condition at either time point,  $ps > .12$ .

### Hypothesis 2

Table 2 summarizes criterion correlations with victimization scores at baseline. Victimization was not associated with depressive symptoms in the Standard or Qualifier conditions, and associations between victimization and depressive symptoms did not differ by condition, either for the full sample or when examining males and females separately. Victimization was not associated with difficulties in emotion regulation in either condition. However, a significant correlation was found for females in the Standard condition, and this was stronger, compared with the correlation for females in the Qualifier condition,  $z = 2.28$ ,  $p < .05$ .

Table 2 Study 4 Correlations Between Physical Victimization and Criterion Variables by Condition and Sex

Criterion variable	Standard	Qualifier	z
Baseline depressive symptoms			
Total sample	.02	.08	-0.89
Female	.23**	.16**	0.80
Male	-.03	.04	-0.69
Baseline difficulties in emotion regulation			
Total sample	.06	.06	0.00
Female	.28**	.08	2.28*
Male	-.00	.07	-0.69
2-month revictimization			
Total sample	.16**	.24**	1.24
Female	.25**	.19**	0.69
Male	.17**	.26**	-0.93

Note. Physical victimization was scored continuously, with possible scores ranging from 0 to 16. The  $r$ -to- $z$  transformations were used to test the difference between correlation coefficients.

\*  $p < .05$ . \*\*  $p < .01$ .

### Hypothesis 3

Baseline victimization was associated with victimization over the 2-month follow-up in both the Standard and Qualifier conditions,  $ps < .01$  (see Table 2). The strength of this association did not differ by condition when examining the full sample, or separately among males and females.

## Discussion

Study 4 examined the same research questions as Studies 2 and 3 and included an additional criterion variable: revictimization over a 2-month follow-up. Approximately 9% of participants reported victimization in the past year at baseline assessment. Rates of victimization did not differ by condition at baseline or follow-up. However, reports of victimization differed by sex in the Standard condition, with more males reporting victimization than females. Sex differences were not found using “Not including horseplay or joking around” qualifier.

Associations with criterion variables across both forms of the measure were mixed. Victimization was associated with depressive symptoms for females in both conditions, but these associations did not differ by condition. Emotion regulation was related to victimization only among females in the Standard condition. Finally, both forms of the measure were associated with revictimization, and again, the strength of these associations did not differ by condition. This pattern of associations may not generalize to other follow-up intervals. Overall, findings suggest that including the qualifier does not consistently provide a more valid assessment of IPV victimization when considering theoretically important correlates of such experiences.

## Study 5

We conducted a single-paper meta-analysis (SPM; McShane & Böckenholt, 2017) to synthesize our data regarding victimization prevalence rates. Meta-analysis generally pools results of multiple studies using a weighted average, thus providing more precise estimates than any individual study and increasing statistical power to reconcile potentially discrepant results. SPM accounts for patterns of variation and covariation among observations of a dependent measure (i.e., IPV victimization) from a set of studies measuring the same underlying phenomenon. Specifically, it accommodates variation resulting from differences in method factors among observations (i.e., participant sex and measure condition) and covariation of observations nested within studies and conditions. Notably, the SPM model demonstrates statistical sufficiency to the “gold standard” multilevel model fit of individual-level observations (McShane & Böckenholt, 2017).

## Method

Participant sex (male vs. female), measurement condition (qualifier vs. standard act condition), proportion reporting any victimization, and sample sizes were extracted from the four previous studies ( $n = 2,350$  participants). Data were analyzed using <https://blakemcshane.shinyapps.io/spmeta/> online software.

## Results

Pooled estimates of reported victimization are presented in Table 3. There was no simple effect of condition, estimate = 0.01,  $SE = 0.03$ , or sex, estimate =  $-0.01$ ,  $SE = 0.03$ , on victimization rates. There was also no interaction of condition by sex, estimate =  $-0.02$ ,  $SE = 0.03$ . The heterogeneity estimate was high,  $I^2 = 87.04\%$ , 95% confidence interval [79.59, 91.78], such that considerable variability across studies was not accounted for by measurement condition or sex.

Table 3 Meta-Analysis Pooled Estimates of Physical Victimization by Condition and Sex

Condition	Sex	Proportion	SE
Standard	Male ( <i>n</i> = 444)	0.145	0.05
	Female ( <i>n</i> = 714)	0.133	0.05
Qualifier	Male ( <i>n</i> = 462)	0.150	0.05
	Female ( <i>n</i> = 730)	0.115	0.05

*Note.* Sample sizes reflect total number of males or females in each condition, respectively. Proportion indicates the reported rate of victimization within each subgroup.

## General Discussion

Concerns about widely used measures of IPV victimization claim these measures are insensitive to context and thus conflate acts that occur while joking with intentional acts of aggression. These “false positives” may undermine the validity of such measures and call into question what we know about the prevalence, causes, and effects of IPV. Although attempts to eliminate false positives have been shown to influence reported rates of victimization in some studies, the extent to which this improves the criterion validity of the scale rarely has been tested empirically. We conducted four experimental studies comparing measures using standard measurement instructions with those that direct respondents to omit instances of joking around or roughhousing. We examined victimization rates and criterion validity of each format in different samples, utilizing different sets of items (from the PVS and CADRI), different response options (dichotomous and continuous), and different recall periods (during high school and past year).

In the present research, self-reports of victimization using standard instructions did capture instances of playful aggression (i.e., “false positives”). Specifically, in Study 1, when given standard instructions, 36% of participants who reported violence victimization endorsed at least one false positive. In contrast, only 18% of those given qualified instructions endorsed a false positive. This difference was not statistically significant, but it is consistent with the idea that individuals commonly report being the recipient of playful acts of aggression on self-reports and that instructions to omit such behaviors may reduce reports of playful victimization. Yet, instructions explicitly asking participants to not include acts occurring in a playful context do not completely eliminate reports of playful aggression.

Overall, instructions to exclude joking or roughhousing behaviors did *not* consistently lead to lower rates of victimization. Indeed, the absence of a between-condition difference was confirmed by SPM results using the 2,350 participants in the four studies. Use of the qualifier also did not differentially affect male and female reports of victimization. This is consistent with previous studies showing that the relative rates of male and female aggression did not change when playful acts are excluded (Fernández-González et al., 2013; Frías, 2016; Jouriles et al., 2009). However, it differs from Hamby’s (2016b) study showing that excluding horseplay or joking resulted in more female reports of victimization, while not affecting male reports of victimization. A recent study that asked participants to describe the context in which aggressive acts occurred after they completed the CTS found that males were more likely than females to describe reported victimization as playful or accidental, and raises the possibility that asking respondents to report on the context of the acts after they are endorsed may have a different impact on prevalence rates than instructing participants to exclude instances of joking or playing around (Ackerman, 2018).



This set of studies also demonstrated that instructions to exclude aggression occurring in playful contexts did not meaningfully influence the criterion validity of victimization measures. Associations between victimization and theoretically related constructs (depressive symptoms, emotion regulation difficulties, revictimization) did not differ consistently across condition. This result emerged in sufficiently powered studies using different samples, measures, and recall periods. The absence of differences may be a result of several factors. One possibility is that violent acts that occur in playful contexts are not really “false positives” at all; being slapped, grabbed, shoved, or hit may be damaging (physically and/or emotionally) from the recipient’s perspective whether the behaviors occur in joking contexts or not. It also is possible that “playful” aggression is correlated with intentional aggression, so eliminating reports of “playful” aggression does not alter associations with criterion variables. Another possibility is that some individuals may reattribute a partner’s aggressive behavior as joking or roughhousing after the fact as a way of denying that they were a victim of violence and/or to justify the decision to remain in the relationship (Arriaga, 2002; Littleton & Henderson, 2009). Similarly, playful behaviors that occur in the context of existing IPV may be perceived as more violent than in relationships without such aggression. Thus, the presumed reason for the behavior may be a less valid indicator of violence than whether or not the behavior occurred.

### Limitations

A few limitations should be acknowledged. First, use of the qualifier was examined only in the context of physical victimization during adolescence and early adulthood, using predominately White convenience samples. It is possible that experiencing IPV has different meanings during different developmental periods, among different populations, and different types of relationships (e.g., casual dating, committed partnerships, marriage). In short, caution should be exercised in generalizing the present findings. This procedure was also retrospective, and prevalence may differ when measured with repeated or contemporaneous assessments (Jouriles et al., 2005). Additionally, the current studies assessed the use of the qualifier in reports of physical IPV victimization, using a relatively small number of physical violence items and aggregate scores. It is not clear if these results generalize to individual violent acts, acts of psychological IPV, or measures of perpetration. However, this type of qualifier should not necessarily be used to assess every type of aggression (e.g., sexual assault, which is always inappropriate regardless of context; Hamby, 2016b). Finally, we assessed specific criterion variables related to IPV victimization, but there are many other criterion variables that could be considered in future research (e.g., trauma symptoms, anxiety, interpersonal competence).

### Research Implications

To improve the assessment of violence, it is essential to continue to study how measurement methods impact our understanding of physical IPV. The present study underscores the need to empirically test attempts to improve the validity of IPV measures. Although several studies have shown that excluding behaviors that occur in a playful context reduce the prevalence of IPV victimization (Arriaga, 2002; Fernández-González et al., 2013; Foshee et al., 2007; Frías, 2016), these studies did not attempt to evaluate whether this practice improves criterion validity (see Jouriles et al., 2009, for an exception). Showing that changes to a measure alter reported rates of violence, on its own, is not evidence of improved validity; it is important to also document that changes incrementally strengthen associations

with meaningful constructs, including putative outcomes, risk factors, and protective factors (Jouriles & Kamata, 2016).

There are other ways that self-report measures of violence and victimization can be modified that might increase their validity. For example, rather than trying to reduce reports of “false positives” by instructing respondents to exclude playful behavior, respondents instead could be directed toward recalling “true positives.” That is, to report behaviors attributed as intentional or meant to harm, by including instructions or item-stems that assess perceived perpetrator motives. Use of clinical interviews to assess victimization may help identify these “true positives,” and function as an indicator of convergent validity. Another approach is to evaluate effects of different recall periods or data collection methods, such as cumulative assessments over time (Jouriles et al., 2005). Empirically investigating the incremental benefits of new methods, with special attention to enhancing racial and ethnic diversity of research samples, would provide valuable guidance for improving IPV measurement.

### Clinical and Policy Implications

From a health care services perspective, eliminating “false positives” when evaluating violence in adolescents’ and young adults’ romantic relationships could lead to important advances in risk assessment and treatment planning. However, our results indicate that asking individuals to exclude aggressive behaviors that occurred while playing around or roughhousing does not necessarily produce more accurate reports of victimization. Rather, this research suggests that inviting individuals to report all aggressive interactions regardless of the context provides as valid an estimate, and does not require respondents to decide whether particular acts should be reported. In a clinical or health care setting, following up a brief screening questionnaire with an interview provides the opportunity to elicit more details about particular acts and incidents and better determine whether they meet specified criteria for violence.

### Conclusion

In the present research, directing respondents to exclude behaviors that occur in a playful context demonstrated statistically comparable criterion validity to commonly used standard act instructions when assessing physical IPV victimization. Use of such modified instructions also did not differentially affect male and female reports. These data underscore the need to empirically investigate the accuracy and criterion validity of violence measures, and to consider how reference to factors such as the emotional qualities of the act, perceived intentionality, self-defense motivations, and extent of injury may affect both rates of aggression and the validity of measures (Hamby, 2016a, 2017).

### References

- Ackerman, J. (2018). Assessing Conflict Tactics Scale validity by examining intimate partner violence over reporting. *Psychology of Violence, 8*, 207–217. 10.1037/vio0000112
- Archer, J. (2000). Sex differences in aggression between heterosexual partners: A meta-analytic review. *Psychological Bulletin, 126*, 651–680. 10.1037/0033-2909.126.5.651
- Arriaga, X. B. (2002). Joking violence among highly committed individuals. *Journal of Interpersonal Violence, 17*, 591–610. 10.1177/0886260502017006001

- Catalano, S. (2012). *Intimate partner violence, 1993–2010* (NCJ 239203). Washington, DC: Bureau of Justice Statistics.
- Devries, K. M., Mak, J. Y., Bacchus, L. J., Child, J. C., Falder, G., Petzold, M., . . . Watts, C. H. (2013). Intimate partner violence and incident depressive symptoms and suicide attempts: A systematic review of longitudinal studies. *PLoS Medicine*, *10*, e1001439. 10.1371/journal.pmed.1001439
- Ehring, T., & Quack, D. (2010). Emotion regulation difficulties in trauma survivors: The role of trauma type and PTSD symptom severity. *Behavior Therapy*, *41*, 587–598. 10.1016/j.beth.2010.04.004
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251. 10.1037/0033-295X.87.3.215
- Exner-Cortens, D., Eckenrode, J., & Rothman, E. (2013). Longitudinal associations between teen dating violence victimization and adverse health outcomes. *Pediatrics*, *131*, 71–78. 10.1542/peds.2012-1029
- Fernández-González, L., O’Leary, K. D., & Muñoz-Rivas, M. J. (2013). We are not joking: Need for controls in reports of dating violence. *Journal of Interpersonal Violence*, *28*, 602–620. 10.1177/0886260512455518
- Foshee, V. A., Bauman, K. E., Linder, F., Rice, J., & Wilcher, R. (2007). Typologies of adolescent dating violence: Identifying typologies of adolescent dating violence perpetration. *Journal of Interpersonal Violence*, *22*, 498–519. 10.1177/0886260506298829
- Frías, S. M. (2016). Casual attributions of dating violence perpetration and victimization in a national sample of high school Mexican students. *Journal of Family Violence*, *31*, 1019–1023. 10.1007/s10896-016-9878-z
- Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment*, *26*, 41–54. 10.1023/B:JOBA.0000007455.08539.94
- Hamby, S. (2016a). Advancing survey science for intimate partner violence: The Partner Victimization Scale and other innovations. *Psychology of Violence*, *6*, 352–359. 10.1037/vio0000053
- Hamby, S. (2016b). Self-report measures that do not produce gender parity in intimate partner violence: A multi-study investigation. *Psychology of Violence*, *6*, 323–335. 10.1037/a0038207
- Hamby, S. (2017). A scientific answer to a scientific question: The gender debate on intimate partner violence. *Trauma, Violence and Abuse*, *18*, 145–154. 10.1177/1524838015596963
- Hamby, S., Finkelhor, D., & Turner, H. (2012). Teen dating violence: Co-occurrence with other victimizations in the National Survey of Children’s Exposure to Violence (NatSCEV). *Psychology of Violence*, *2*, 111–124. 10.1037/a0027191
- Hamby, S., & Grych, J. (2013). *The web of violence: Exploring connections among different forms of interpersonal violence and abuse* (R. J. Johnson, Ed.). Dordrecht, the Netherlands: Springer. 10.1007/978-94-007-5596-3
- Jouriles, E. N., Choi, H. J., Rancher, C., & Temple, J. R. (2017). Teen dating violence victimization, trauma symptoms, and revictimization in early adulthood. *Journal of Adolescent Health*, *61*, 115–119. 10.1016/j.jadohealth.2017.01.020

- Jouriles, E. N., Garrido, E., Rosenfield, D., & McDonald, R. (2009). Experiences of psychological and physical aggression in adolescent romantic relationships: Links to psychological distress. *Child Abuse and Neglect*, *33*, 451–460. 10.1016/j.chiabu.2008.11.005
- Jouriles, E. N., & Kamata, A. (2016). Advancing measurement of intimate partner violence. *Psychology of Violence*, *6*, 347–351. 10.1037/vio0000014
- Jouriles, E. N., McDonald, R., Garrido, E., Rosenfield, D., & Brown, A. S. (2005). Assessing aggression in adolescent romantic relationships: Can we do it better? *Psychological Assessment*, *17*, 469–475. 10.1037/1040-3590.17.4.469
- Kaufman, E. A., Xia, M., Fosco, G., Yaptangco, M., Skidmore, C. R., & Crowell, S. E. (2016). The Difficulties in Emotion Regulation Scale Short Form (DERS-SF): Validation and replication in adolescent and adult samples. *Journal of Psychopathology and Behavioral Assessment*, *38*, 443–455. 10.1007/s10862-015-9529-3
- Kaura, S. A., & Lohman, B. J. (2007). Dating violence victimization, relationship satisfaction, mental health problems, and acceptability of violence: A comparison of men and women. *Journal of Family Violence*, *22*, 367–381. 10.1007/s10896-007-9092-0
- Lehrner, A., & Allen, N. E. (2014). Construct validity of the Conflict Tactics Scales: A mixed-method investigation of women's intimate partner violence. *Psychology of Violence*, *4*, 477–490. 10.1037/a0037404
- Littleton, H., & Henderson, C. E. (2009). If she is not a victim, does that mean she was not traumatized? Evaluation of predictors of PTSD symptomatology among college rape victims. *Violence Against Women*, *15*, 148–167. 10.1177/1077801208329386
- McShane, B. M., & Böckenholt, U. (2017). Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *Journal of Consumer Research*, *43*, 1048–1063. 10.1093/jcr/ucw085
- Noll, J. G., & Grych, J. H. (2011). Read-react-respond: An integrative model for understanding sexual revictimization. *Psychology of Violence*, *1*, 202–215. 10.1037/a0023962
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*, 411–419.
- Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401. 10.1177/014662167700100306
- Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised Conflict Tactics Scales (CTS2) development and preliminary psychometric data. *Journal of Family Issues*, *17*, 283–316. 10.1177/019251396017003001
- Vagi, K. J., O'Malley Olsen, E., Basile, K. C., & Vivolo-Kantor, A. M. (2015). Teen dating violence (physical and sexual) among U.S. high school students: Findings from the 2013 National Youth Risk Behavior Survey. *Journal of the American Medical Association Pediatrics*, *169*, 474–482. 10.1001/jamapediatrics.2014.3577
- Vega, E. M., & O'Leary, D. (2007). Test-retest reliability of the Revised Conflict Tactic Scales (CTS2). *Journal of Family Violence*, *22*, 703–708. 10.1007/s10896-007-9118-7
- Wolfe, D. A., Scott, K., Reitzel-Jaffe, D., Wekerle, C., Grasley, C., & Straatman, A. L. (2001). Development and validation of the Conflict in Adolescent Dating Relationships Inventory. *Psychological Assessment*, *13*, 277–293. 10.1037/1040-3590.13.2.277

Wu, H., Corney, J., & Grant, M. (2014, May). Relationship between quality and payment in crowdsourced design. *Computer Supported Cooperative Work in Design (CSCWD), Proceedings of the 2014 IEEE 18th International Conference* (pp. 499–504). New York, NY: IEEE.