**Marquette University**

**e-Publications@Marquette**

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

***This paper is NOT THE PUBLISHED VERSION*.**

Access the published version via the link in the citation below.

*Psychology of Violence*, Vol. 12, No. 2 (2022): 116-126. [DOI](https://doi.org/10.1037/vio0000403). This article is © American Psychological Association and permission has been granted for this version to appear in [e-Publications@Marquette](http://epublications.marquette.edu/). American Psychological Association does not grant permission for this article to be further copdpied/distributed or hosted elsewhere without the express permission from American Psychological Association.

“It still happened”: A Mixed-Methods Analysis of College Students’ Rationales for Endorsing Acts of Violent Victimization

Kristen Yule

Department of Psychology, Marquette University, Milwaukee, WI, USA

John Grych

Department of Psychology, Marquette University, Milwaukee, WI, USA

# Abstract

*Objective:* Measures of relationship violence have been criticized for failing to distinguish intentional acts of aggression from behaviors that occur in a playful context. However, efforts to reduce this concern by modifying the questionnaires’ instructions have not reliably reduced reporting rates or improved the criterion validity of the measures. This experimental study investigated how respondents who were randomly assigned to one of three instruction conditions perceived and responded to a measure of relationship violence. *Method:* Undergraduate students (*N* = 150) reported on partner violence victimization using the *Conﬂict in Adolescent Dating Relationships Inventory* (*CADRI*; Wolfe et al., 2001). They received either the standard instructions, instructions to exclude acts that occur while joking around or roughhousing, or instructions to include only intentional acts. Participants then took part in a think aloud procedure and a structured cognitive interview to examine their reasons for endorsing or not endorsing items. *Results:* Instructions designed to eliminate reports of behaviors that occur in playful contexts or are not intended to harm did not reduce overall rates of victimization. Respondents’ perceptions of their partners’ intent and the impact of the behavior on them were particularly salient when deciding whether to endorse items, but many participants reported behaviors simply because they occurred. *Conclusion:* These ﬁndings underscore the importance of victims’ subjective perceptions and experiences in determining whether to endorse a particular behavior and offer insight into why instructions not to report certain kinds of behaviors do not consistently decrease rates of violence or improve the measure’s validity.

# Keywords:

intimate partner violence, measurement, validity

Understanding the causes and consequences of intimate partner violence depends on accurately measuring violent behavior that occurs in a relationship. Violence typically is defined as intentional acts perpetrated to harm another (e.g., DeWall et al., 2011; Hamby & Grych, 2013; World Health Organization, 2002), and the most widely used measures of relationship violence ask respondents to report how often certain behaviors (e.g., threatening, pushing, hitting) occurred during a discrete period of time (e.g., 1 year). Assessing specific, concrete behaviors is designed to reduce the ambiguity that would arise from asking whether “abuse” or “violence” occurred; however, many of the behaviors included on these measures also can occur in nonviolent contexts. For example, individuals may grab or push their partner when joking around or roughhousing without intending to hurt them. Such behaviors can be considered “false positives” that inflate estimates of victimization and reduce the measure’s validity (Hamby, 2016a, 2016b; Lehrner & Allen, 2014), and questionnaires such as the *Conflict Tactics Scale* (Straus et al., 1996) and *Conflict in Adolescent Dating Relationships Inventory* (*CADRI*; Wolfe et al., 2001) have been criticized for failing to distinguish between playful behavior and true acts of violence (e.g., Hamby, 2016a; Sargent et al., 2020; Sillito, 2012).

Attempts have been made to address this problem by modifying the instructions of relationship violence questionnaires to direct respondents not to report behaviors that occurred while joking or playing around (Hamby, 2016b; Jouriles et al., 2009; Sargent et al., 2020). However, these modifications have not increased the criterion validity of the measures (Jouriles et al., 2009; Sargent et al., 2020). It is not clear why instructions that should produce more accurate reports of violence do not appear to do so. Therefore, the present study sought to provide insight into this question by assessing respondents’ thought processes while completing a widely used measure of intimate partner violence. Our goal was to better understand why respondents chose to endorse or not endorse particular items and to examine whether altering the instructions affected their responses.

# Assessment of Intimate Partner Violence

The goal of self-report measures of relationship violence is to accurately distinguish behaviors that meet the criteria for violence from those that may appear similar but are not intended to harm the partner. This is difficult to do, particularly in samples of adolescents and young adults, since participants often report engaging in ostensibly violent behaviors while joking around or roughhousing (Arriaga, 2002; Foshee et al., 2007; Frías, 2016; Jouriles et al., 2009). For example, a study of undergraduate women found that 20% of those who endorsed at least one item on the *Conflict Tactics Scale* later reported these aggressive behaviors had occurred while joking around or roughhousing, and another 38% reported instances of both playful and intentional aggression (Lehrner & Allen, 2014). In a sample of American and Australian undergraduates, Ackerman (2018) similarly found 22% of acts endorsed on the *Conflict Tactics Scale* were viewed as accidents or joking around when respondents were asked to describe the context in which a behavior occurred.

Thus, it appears that scores on measures of relationship violence combine reports of actual violence with “false positives” (Hamby, 2016a, 2016b; Lehrner & Allen, 2014), potentially reducing their validity, or ability to accurately assess the construct they are designed to assess (e.g., DeVon et al., 2007; Haynes et al., 1995). Several studies have attempted to reduce “false positives” by directing respondents not to endorse behaviors that occurred in a playful context. In a series of four experimental studies, Sargent et al. (2020) compared the prevalence and criterion validity of self-reported victimization when respondents received either standard instructions or instructions to exclude behaviors that occurred while joking around. The results showed that, across the four studies, these instructions did not consistently reduce prevalence rates of victimization or improve the criterion validity of the scale for men or women. These findings replicated the only other study to our knowledge to examine whether modifying the instructions improved the criterion validity of a self-report measure of relationship violence (Jouriles et al., 2009).

An alternative approach to improving the validity of victimization measures is to try to increase the rate of “true positives” by directing respondents to report only behaviors that they believe were intended to harm them. Hamby (2016a) examined how altering the instructions in two ways affected college students’ self-reports of victimization. In addition to a statement that directed respondents to exclude instances of aggression occurring during “horseplay or joking around,” some respondents received instructions to endorse only behaviors that occurred when their partner was angry. Men’s reports of victimization did not differ significantly across conditions, but women reported less victimization with the “anger” qualifier and *more* victimization with the “not during horseplay or joking around” qualifier. Hamby (2016a) did a similar test with a sample of women recruited from social service agencies that serve victims of domestic violence and found that the rates of violence reported did not differ for groups receiving the standard instructions, the “joking” qualifier, and the “angry” qualifier.

Directing participants not to endorse a subset of the potentially violent behaviors they have experienced would be expected to decrease reports of victimization on these questionnaires, and reducing “false positives” should increase their accuracy. The findings showing that these outcomes do not occur raise important questions about how respondents interpret the questions on these measures and decide whether or not to endorse particular items. To shed light on the thought processes that respondents engage in while completing measures of relationship violence, Sargent et al. (2020) used a “think aloud” procedure (Ericsson & Simon, 1980) in which participants spoke aloud the thoughts they were having while completing the survey. Nearly 20% of those who received instructions to exclude acts that occurred in a joking or playful context still endorsed items that happened in such situations. These participants’ comments indicated they endorsed the items because they were unsure about their partner’s intentions or because the acts were perceived as joking or playful by one partner but not the other. These findings suggest that deciding whether a behavior is playful is more complex than it may appear, and the perception of intentionality may differ for the perpetrator and victim. Consequently, instructions to not report acts of “playful aggression” may not effectively screen out behaviors that are not meant to harm. The decision to endorse an item also may be influenced more by respondents’ perception of the effect of the behavior than its intent. For example, being pushed, shoved, or slapped may still hurt (physically and/or emotionally) even if it occurred while joking or playing around. Another possibility is that respondents simply focus on whether particular behaviors occurred and do not attend to instructions to discount behaviors occurring in particular contexts. In other words, the nuances included in the instructions may not be salient when respondents decide whether to endorse an item. However, Sargent et al. (2020) coded only whether respondents reported aggression that occurred in playful contexts and thus does not shed light on why they chose to endorse or not to endorse particular behaviors.

# The Present Study

The goal of the present investigation was to better understand how respondents decide whether or not to endorse items on a widely used self-report measure of relationship violence (*CADRI*; Wolfe et al., 2001). Knowing how respondents perceive the instructions and questions on the measure and their rationale for reporting particular behaviors may provide insight into the meaning of scores on relationship violence scales. We were particularly interested in assessing the extent to which respondents focused on the intent and the impact of a behavior when deciding whether to endorse an item and on the effect of modifying the instructions on their reports and reasoning. In addition, because prior studies investigating this issue have focused on physical victimization (Hamby, 2016b; Sargent et al., 2020), we extended this work by including items that represent psychological violence. Behaviors such as insulting a partner, threatening to end the relationship, or ridiculing a partner in front of others are common forms of psychological aggression and may have a negative emotional and/or psychological impact even if such statements are delivered in a joking or playful manner. Indeed, Jouriles et al. (2009) found that participants rated their experiences of psychological aggression as more unpleasant and more intentionally hurtful than their experiences of physical aggression.

The present study used two methods to assess how respondents perceive and respond to the *CADRI*. First, we utilized an open-ended “think aloud” procedure (Ericsson & Simon, 1980) in which participants were asked to express whatever thoughts occurred to them while they filled out the questionnaire. The think aloud procedure does not constrain participants’ responses or provide any cues about what kinds of responses are desirable and so increases confidence that their comments are self-generated. However, because participants may not spontaneously voice all their thoughts, we also administered a structured cognitive interview to identify factors that influence respondents’ decision to report victimization. Participants were randomly assigned to one of three conditions prior to completing any measures. In the first, they read the standard instructions used on the *CADRI*, which ask respondents to report how often particular acts of psychological and physical aggression occurred with a partner. For the second condition (“Joking”), a phrase was added before each item directing participants to exclude behaviors that occurred while “roughhousing or joking around.” In the third condition (“Intentional”), a phrase was added after each item directing participants to endorse only behaviors that occurred “on purpose.”

The present study aimed to answer the following research questions:

1. Do instructions to ignore aggressive behaviors in playful contexts or report only behaviors viewed as intentional affect respondents’ endorsement of physical and psychological victimization on a self-report measure of relationship violence, and do results vary based on the sex of participants?
2. Do instructions to ignore aggressive behaviors in playful contexts or report only behaviors viewed as intentional affect participants’ reasons for endorsing or not endorsing item? In particular, to what extent do participants focus on the intent and effect of a behavior when deciding whether to endorse and item, and do responses vary based on participant sex?

# Method

## Participants

Participants were recruited from an undergraduate psychology participant pool at a midsize, private university in the Midwestern United States and offered extra course credit. Participants were told the study investigated how people approach answering questionnaires and how they decide on their final answer. They completed a set of questionnaires that included the *CADRI*. The sample for Session 1 (*n* = 190) was 65% women, ranged in age from 18 to 30 years old (*M* = 19.84, *SD* = 1.23), and the majority identified their race as White (68%), with smaller numbers identifying as Asian (14%), Hispanic (8%), Black (4%), “more than one race” (5%), and American Indian/Alaska Native (1%). A majority of participants identified as heterosexual (*n* = 142, 75%). More than half (62%) indicated they were not currently in a relationship, 37% reported they were dating one person, and 1% were dating more than one person. One hundred fifty-three participants returned for Session 2 (81%). There were no significant differences on any demographic variables, including sex, ethnicity, race, or age (*p*s > .20), between participants who participated in the second session and those who did not, or across the three experimental conditions. Chi-square goodness-of-fit tests indicated that the sample included a somewhat larger percentage of women (65% vs. 54%; χ2 = 8.89, *p* = .01) than the undergraduate population of the university at the time the study was conducted, but did not differ in race or ethnicity. An a priori power analysis using G\*Power (Faul et al., 2007) showed the study had adequate power (.80) to detect a medium to large effect in a model that tests main and interaction effects for participant sex and instruction condition. The rationale for choosing this effect size was based on prior research that reported a medium-sized effect for the interaction between instruction wording and participant gender on reports of relationship violence in a sample of undergraduate students (Hamby, 2016a).

## Procedure

Participation involved two sessions. In the first, participants completed a series of baseline questionnaires online using the Qualtrics survey platform. The second took place in-person in a private room in the Psychology Department and involved completing questionnaires, participating in the “think aloud” procedure, and participating in the cognitive interview.

In the first session, after providing informed consent, Qualtrics randomly assigned participants to one of three groups that differed in the instructions for completing the *CADRI* (Wolfe et al., 2001). One group (*n* = 64) received the standard questionnaire instructions in which respondents are asked to indicate how often they had been the recipient of a series of aggressive behaviors in a romantic relationship on a 5-point scale (0 = *Never*, 4 = *Four times or more*) in the past year. For the second group (“Joking” condition; *n* = 63), each behavior was preceded by the phrase “Not including roughhousing and joking around …” (e.g., “Not including roughhousing and joking around, my partner kicked, hit, or punched me”). For the third group (“Intentional” condition; *n* = 63), each behavior was followed by the phrase “… on purpose” (e.g., “My partner kicked, hit, or punched me on purpose”). Once they completed the questionnaires online, participants were directed to a website to schedule a time to complete the second, in-person part of the study. The average length of time between the first and second session was approximately 3 weeks.

For the second session, participants completed a paper and pencil version of the victimization measure they had completed online in the first session. They remained in the same instruction condition that they were assigned to in the first session. Participants in all three conditions were instructed to “think aloud” as they filled out the questionnaire, saying out loud “all the things that go through your mind as you are choosing your answer” (Ericsson & Simon, 1980). Participants were told 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 each participant through three practice questions and left the room for the participant to respond in private once it was clear that they understood the task. After participants completed the “think aloud” procedure, a cognitive interview was conducted to further examine participants’ thought processes while completing the *CADRI*. All responses to the “think aloud” task and cognitive interview were audio recorded and transcribed verbatim for coding. Trained research assistants conducted all interviews.

## Measures

Participants reported on their experiences of physical, emotional, and verbal victimization in romantic relationships over the past 12 months using the 10-item short form of the *CADRI* (Wolfe et al., 2001). Sample items include “They threw something at me” and “They threatened to end the relationship.” Respondents reported how often these behaviors “occurred with a romantic partner in the past year, including exes,” on a 5-point scale (0 = *Never*, 4 = *Four times or more*). Participants completed the same version of the scale (i.e., standard instructions, “joking” qualifier, “on purpose” qualifier) at both time points. Previous research demonstrates that the *CADRI* has moderate test–retest correlations over a 2-week period (Wolfe et al., 2001) and good convergent validity with other measures of physical and psychological aggression (Cascardi et al., 2019). The *CADRI* showed good internal consistency for the standard (α = .94), joking (α = .85), and intentional (α = .96) instruction conditions.

### Think Aloud Task

Responses during the “think aloud” procedure were coded and analyzed using a conventional content analysis approach (Hsieh & Shannon, 2005), which involves deriving codes directly and inductively from the data. Following guidelines on qualitative content analysis (Hsieh & Shannon, 2005; Zhang & Wildemuth, 2009), coding categories were developed by identifying key ideas or themes from a subset of participant responses. The initial codes were refined by combining some and splitting others into subcategories. Using the revised coding scheme (see Table 1), responses were then coded by two research assistants who had not participated in the initial development of the coding manual and were not aware of the hypotheses and design of the present study.  
  
Table 1. *Coding Scheme for the Think Aloud Task*

|  |  |  |
| --- | --- | --- |
| Code description | Sample responses from participants for *endorsing* items | Sample responses from participants for *not endorsing* items |
| Playing around | “I recall a couple times with my ex they did that as a joke so I would say 3 times.” | “Well the person that I’m seeing jokingly calls me a loser but I know he’s kidding because he usually says it when I do something good. So, it’s very obvious that he’s being sarcastic. So, I’m still go with never because kidding isn’t really real. |
| No intent to hurt | “I am going to say once. They kind of played with my hair and pulled it a little bit but I don’t think the intent was to hurt.” | “For this one, I still don’t think they did it on purpose to hurt me, so I am going to say never.” |
| Action normalized | “I think everyone’s boyfriend or significant other threatens to end their relationship.” | “Um no I mean of course like all arguments kind of come with like a oh you know could this kind of be the end of things um but I think it was all in really logical talking but I don’t think that was a real threat.” |
| Action hurt | “My last boyfriend definitely did that. I am going to say three times. I mean in my opinion I think he thought it was joking around, but it wasn’t okay to say in front of other people and it hurt my feelings a lot.” | “This has never happened. Sure, there’s times where I felt they’re being rude, but they don’t ever mean to do anything deliberately. If he hurts my feelings, he always tries to clarify his intentions and that he’s sorry.” |
| Context clarification | “No, well he shook me if I was like out of it, so I would say like once. But that was like a shook like hey you’re not this crazy person but yeah it happened.” | “I remember one of them saying something verbally about maybe emotionally hurting me. Not actually like physically hurting me. So, I mean if this one means verbally I’d have to say probably once, but if it means physically, I’d probably say never so I’m going to assume this is physical and say never. |
| Action endorsed without explanation | “That has happened a few times.” | N/A |
| Never happened | N/A | “Never. That has not happened in the past year.” |

Note: N/A = Not applicable for code.

Participants’ verbal responses to the 10 *CADRI* items were coded into one of seven categories: (a) *Playing around:* participant described their partner’s behavior as occurring with joking or playful intentions; (b) *No intent to hurt:* participants stated an action occurred accidentally and/or without an intention to cause harm; (c) *Action normalized*: participant described the behavior as being normal or typically occurring during a fight or argument; (d) *Action hurt*: participant indicated they were hurt physically and/or emotionally by their partner’s behavior; (e) *Context clarification*: participant described details about the behavior and the situation in which it occurred; (f) *Action endorsed without explanation*: participant indicated a behavior occurred but did not elaborate further; and (g) *Never happened*: participant stated that a behavior never occurred with a partner. Responses of participants who did not have a partner in the past 12 months also were coded as “Never Happened.” Interrater reliability for the “think aloud” task was good; κ values ranged from .73 to 1.00, with an average value of .86. All participant responses were coded into one of these seven categories.

### Cognitive Interview

After completing the “think aloud” procedure, participants took part in a structured interview that was developed for this study to assess how students interpreted the *CADRI*. This type of cognitive interview also is known as a retrospective, tailored interview, in which an interviewer asks scripted probes about questionnaire items after completing the measure (Willis, 2005). Participants were provided with a blank copy of the *CADRI* to reference while responding to a series of questions. The questions focused on participants’ perceptions of the intent and effect of the behaviors and asked if they would or would not endorse items under the following conditions: (a) the behavior had occurred but their partner was joking, playing around, or roughhousing; (b) the behavior had occurred but they thought that their partner did not mean to hurt them; and (c) their partner did not mean to hurt them but the behavior did hurt either physically or emotionally. The last question assessed how participants weigh the roles of intent and impact; is a behavior discounted if it was perceived as unintentional even though it caused harm, or do participants focus more on the effect than the intent of a behavior when it causes harm? After answering each question participants were asked to explain why they would or would not endorse the behavior. Because the questions are phrased hypothetically, all participants could respond regardless of whether they had been in a relationship in the prior 12 months. The cognitive interview questions and coding scheme with sample responses can be found in Table 2.  
  
Table 2. *Items and Coding Scheme for the Cognitive Interview*

|  |  |  |
| --- | --- | --- |
|  | 1. Would you answer “yes” to an item if a particular behavior had occurred but your partner was joking, playing around, or roughhousing? Why or why not? |  |
|  | 1. Would you answer “yes” to an item if a particular behavior had occurred but you thought that your partner did not mean to hurt you? Why or why not? |  |
|  | 1. Would you answer “yes” to an item if you thought that your partner did not mean to hurt you but the behavior did hurt either physically or emotionally? Why or why not? |  |
| Code description | Sample responses from participants endorsing items | Sample responses for participants not endorsing items |
| Playing around | “Yes, if they ridiculed me or made fun of me in front of others. Not like in a mean way, but just like joking around. I don’t know. Like in front of friends…yes, probably” | “If it’s a mutual understanding of joking not one-sided, then no.” |
| Severity dependent | “I think if they did it aggressively, it doesn’t matter if they meant it or not. So yes.” | “If he didn’t mean to I would have said no, but it depends on how badly he hurt you because if it was “you sent me to the hospital” kind of bad then anyone would be like yeah. But if he didn’t then it’s like  ‘you didn’t mean to then no’ so depends on severity.” |
| Still happened | “It still happened and that is not right. Even if they meant it or did not, it happened.” | “Um no. I would not because—well—I wouldn’t because like I said, I mean, they might not have meant it but it still happened.” |
| No intent to hurt | “Yes. Just because sometimes he could have been doing like kind of like backhanded compliments, but he didn’t do it on purpose.” | “I would say no just because if it’s an accident and you know they weren’t trying to hurt you.” |
| Reference scale | “Yep. It never asks if they meant to do it, there was never an intention question. It was just did this happen or did this not happen and how many times.” | “No, it doesn’t seem like that’s what the questions were asking.” |
| Don’t know | “Ahh, yes. Umm I don’t know cause maybe my judgement’s wrong. I don’t know.” | “No because it’s…I don’t know, that’s just like different I guess.” |

A coding scheme for the cognitive interview was developed using the same process as the one for the think aloud procedure described above. Responses to the three interview questions were coded into one of six categories: (a) *Playing around:* participant stated that their partner was joking or playing around when the behavior occurred; (b) *Severity dependent:* participant indicated that endorsement of an action depends on the severity or potential harmfulness of the behavior; (c) *Still happened*: participant stated that the action occurred and/or had an effect regardless of the partner’s intention; (d) *No intent to hurt:* participant indicated that their partner did not intend to harm them or the behavior was an accident; (e) *Reference scale:* participant referred to the questionnaire’s directions or items to explain why they would or would not endorse an action; and (f) *Don’t know*: participant indicated that they were unsure whether they would endorse the behavior or not. For the cognitive interview, κ values ranged from .80 to 1.00, with an average value of .95. All participant responses were coded into one of these six categories.

## Data Quality Procedure

All data from the paper and pencil version of the *CADRI* used in the second session were double entered into an electronic database by two trained research assistants and then checked by flagging and evaluating conflicting double-entered data. Data inspection for all variables involved the evaluation of outliers, normality, and missing data and revealed missing data (>60%) for three participants (2%). These cases were dropped from study analyses.

# Results

## Research Question 1: Reported Rates of Victimization

First, we tested whether respondents’ reports of psychological and physical victimization differed according to the instructions they received (i.e., standard instructions, “joking” qualifier, “on purpose” qualifier), participant sex, or the interaction of instruction condition and sex. At the baseline assessment (*n* = 190), 124 participants (65%) reported victimization in the past 12 months: 35 (14 men, 21 women) had been randomized to the Standard condition (*M* = 2.58, *SD* = 5.35), 42 (16 men, 31 women) had been randomized to the Joking condition (*M* = 1.82, *SD* = 3.12), and 47 (6 men, 22 women) had been randomized to the Intentional condition (*M* = 3.75, *SD* = 8.37). At follow-up (*n* = 150), 85 participants (57%) reported at least one act of victimization in the past 12 months: 25 (13 men, 12 women) were in the Standard condition (*M* = 2.22, *SD* = 3.73), 31 (7 men, 24 women) were in the Joking condition (*M* = 2.76, *SD* = 3.48), and 29 (9 men, 20 women) were in the Intentional condition (*M* = 3.80, *SD* = 5.12). Reports of victimization at baseline and follow-up were significantly correlated (*r* = .54, *p* = .001). The first hypothesis was tested with a two-way analysis of variance (ANOVA; Condition × Sex) of baseline victimization scores. It revealed that reports of victimization at baseline (*n* = 190) did not differ by instruction condition (*p* = .37) or sex (*p* = .46). The instruction condition by sex interaction also was nonsignificant (*p* = .67). Similarly, victimization scores assessed at follow-up (*n* = 150) showed no condition, sex, or condition by sex effects (*p*s > .20).

## Research Question 2: Reasons Reported for Endorsing and Not Endorsing Victimization

Next, we examined participants’ responses on the think aloud task and cognitive interview. We present the cognitive interview first because it inquired directly about the reasons for why respondents would or would not endorse particular items. There was considerable variation in participants’ responses to the three questions on the cognitive interview. As Table 3 shows, about a third of participants indicated that they would endorse a behavior that occurred while their partner was joking, playing around, or roughhousing; just over half said they would endorse a behavior if they thought their partner did not mean to hurt them; and a large majority stated they would endorse a behavior that hurt them either physically or emotionally even if they thought their partner did not mean to hurt them. A two-way multivariate ANOVA (Condition × Sex) was conducted to assess whether there were significant differences between instruction condition and/or sex of participants on their responses to the three cognitive interview questions at follow-up (*n* = 150). Instruction condition and participant sex were entered as fixed factors and participants’ responses to the cognitive interview items (0 = *No,* 1 = *Yes*) were entered as dependent variables. Results showed that the multivariate effect for instruction condition, Wilks’ Lambda, *F*(6, 284) = 2.18, *p* = .04,  = .04, and participant sex, Wilks’ Lambda, *F*(3, 142) = 4.21, *p* = .01,   
 = .08, were significant, but not their interaction (*p* = .30). Univariate analyses then were conducted to test for differences on each of the three items. As Table 4 shows, significant differences were found for the first and third questions on the cognitive interview at follow-up (*n* = 150).

Table 3. Percentage of Participants Who Endorsed Items on the Cognitive Interview

|  |  |  |  |
| --- | --- | --- | --- |
|  | Cognitive Interview Items |  |  |
| Group | 1. Report a behavior that occurred when partner was joking, playing around, or roughhousing  *n (%)* | 2. Report a behavior that occurred but thought partner did not mean to cause harm  *n (%)* | 3. Report a behavior that hurt either physically or emotionally even though partner did not mean to hurt them  *n (%)* |
| Condition |  |  |  |
| Standard (*n = 49*) | 19 (39) | 32 (65) | 45 (92) |
| Joking *(n = 54)* | 12 (22) | 27 (50) | 43 (80) |
| Intentional *(n = 47)* | 21 (45) | 27 (57) | 40 (85) |
| Sex |  |  |  |
| Male participant *(n = 54)* | 13 (24) | 29 (54) | 41 (76) |
| Female participant *(n = 96)* | 39 (41) | 57 (59) | 87 (91) |
| Total (N = 15) | 52 (35) | 86 (57) | 128 (85) |

*Note.* Percentages are based on the total number that responded “yes” for each respective cognitive interview question.

First, in response to the question of whether respondents would endorse a behavior that occurred when their partner was joking, playing around, or roughhousing, univariate ANOVAs revealed a significant main effect of participant sex. Men (*M* = .24) were less likely than women (*M* = .43) to state they would endorse an act that occurred when their partner was playing around, joking, or roughhousing. Instruction condition narrowly missed traditional levels of significance (*p* = .06): Participants in the Joking condition (*M* = .22) were marginally less likely than those in the Standard (*M* = .39) and Intentional (*M* = .45) conditions to indicate they would endorse a behavior that occurred in a playful context.

Second, in response to the question of whether respondents would report a behavior that hurt either physically or emotionally even though their partner did not mean to hurt them (Question 3), results showed significant main effects of instruction condition and participant sex. Participants in the Joking condition (*M* = .72) were less likely to say that they would endorse a behavior they perceived to be harmful but unintentional than participants in the Standard condition (*M* = .92); however, 80% of those in the Joking condition still stated that they would endorse such an item. Men (*M* = .74) were less likely than women (*M* = .91) to state that they would endorse a behavior that hurt them when their partner did not mean to harm them, though 76% indicated they would endorse the item.

Table 4. Multivariate Analysis of Variance of Instruction Condition by Sex for Participant Endorsement of Acts During the Cognitive Interview

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1. Report a behavior that occurred when partner was joking, playing around, or roughousinga |  |  |  | 2. Report a behavior that occurred but thought partner did not mean to cause harma |  |  |  | 3. Report a behavior that hurt either physically or emotionally even though partner did not mean to hurt thema |  |  |  |
| Source | *df* | *F* | *P* |  | *df* | *F* | *P* |  | *df* | *F* | *P* |  |
| Conditionb | 2 | 2.92 | .06 | .04 | 2 | 1.17 | .31 | .02 | 2 | 3.99 | .02 | .05 |
| Sexc | 1 | 5.62 | .02 | .04 | 1 | .77 | .38 | .01 | 1 | 8.29 | .01 | .05 |
| Condition x Sex | 2 | .26 | .78 | .01 | 2 | .07 | .93 | .01 | 2 | 2.97 | .0 6 | .04 |
| Error | 144 |  |  |  | 144 |  |  |  | 144 |  |  |  |
| Total | 150 |  |  |  | 150 |  |  |  | 150 |  |  |  |

a Act endorsement on cognitive interview questions: 0 = no, 1 = yes

b Condition: 1 = Standard, 2 = Joking, 3 = Intentional

c Participant sex: 0 = male participant, 1 = female participant

## Reasons for Endorsing Victimization Items

Next, we examined the comments participants made while explaining why they chose to endorse particular items during the think aloud task and cognitive interview. We conducted two-way multivariate ANOVA (Condition × Sex) to assess whether there were differences between instruction conditions and/or sex of participants on the frequency of coded responses during the think aloud and cognitive interview at follow-up (*n* = 150). These analyses indicated that participants’ responses did not differ by condition, sex, or their interaction for either the think aloud task (*p*s > .30) or cognitive interview *(p*s > .10). Therefore, we combined data across instruction conditions and sex when describing participants’ reasoning for endorsing acts of victimization during the think aloud task (see Table 5) and cognitive interview (see Table 6).  
  
Table 5. Comments Made When Endorsing Acts of Victimization on the CADRI During the Think Aloud Task (N = 150)

|  |  |
| --- | --- |
| Code description | Frequency *n (%)* |
| Playing around | 17 (20) |
| No intent to hurt | 47 (55) |
| Action normalized | 36 (42) |
| Action hurt | 21 (25) |
| Context clarified | 47 (55) |
| No explanation | 26 (31) |

*Note.* Percentages are based on the total number of participants that provided each reason as justification for endorsing an act of victimization for at least one item on the *CADRI.* They do not sum to 100 because participants could provide multiple reasons for each act endorsed. A total of 85 participants (57%) endorsed at least one act of victimization during the think aloud task.

Table 6. Reasons for Endorsing Acts of Victimization During the Cognitive Interview (N = 150)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cognitive interview question | Playing around  *n* (%) | Depends on severity  *n* (%) | Still happened  *n* (%) | No intent to hurt  *n* (%) | Directions and items  *n* (%) | Don’t know  *n* (%) |
| 1. Would report a behavior that occurred when partner was joking, playing around, or roughousing (*n* = 52, 35%) | 5 (9) | 9 (17) | 26 (50) | 6 (12) | 6 (12) | 0 (0) |
| 2. Would report a behavior that occurred but thought partner did not mean to cause harm (*n* = 86, 57%) | 1 (1) | 4 (5) | 66 (77) | 9 (10) | 5 (6) | 1 (1) |
| 3. Would report a behavior that hurt either physically or emotionally even though partner did not mean to hurt them (*n* = 128, 85%) | 1 (1) | 24 (19) | 87 (68) | 8 (6) | 7 (5) | 1 (1) |

*Note*. Percentages are based on the total number that responded “yes” for each respective cognitive interview question.

### Think Aloud Task

As shown in Table 5, many participants made more than one codable statement when endorsing items during the think aloud task. Across conditions, the most common responses were coded as “No intent to hurt” and “Context clarification.” That is, more than half of the respondents endorsed an aggressive act even though they indicated that they did not believe that their partner intended to hurt them or described details about the situation in which the behavior had occurred. For example, one participant stated, “Maybe once, but it probably wasn’t something serious, it was probably that they threw something in my direction so that I could catch it and then I missed.” Just under half normalized a behavior that they had endorsed (e.g., “People do that all the time”), and 20% endorsed a behavior while noting that it occurred while joking around (e.g., “Uh, yeah just messing around but yeah so I’ll say 3 times”).

### Cognitive Interview

Reasons that participants gave for endorsing acts of victimization on the cognitive interview provide more insight into their thinking (see Table 6). For each of the three questions, the most common reason given was that the behavior happened. For example, one participant stated, “Yes. Even if they were kidding I would still put it had happened.” The other reasons were given by a small minority of participants; the most common of these concerned the impact of the behavior. For example, in response to Question 1, one participant said, “I would answer yes because like if it was obviously hard enough to hurt that’s more than joking around.”

## Reasons for Not Endorsing Victimization Items

Turning to the comments that participants made when they chose *not* to endorse particular items, we evaluated whether there were differences across instruction conditions and/or sex on the think aloud task and cognitive interview. Two-way multivariate ANOVA (Condition × Sex) indicated that codes did not differ by condition (*p*s > .11), sex (*p*s > .20), or their interaction (*p*s > .15) for the think aloud task at follow-up (*n* = 150). Therefore, we combined data across instruction conditions and sex to describe the comments participants made for not endorsing acts of victimization during the think aloud task (see Table 7). However, some differences did emerge on the cognitive interview (described below).  
  
Table 7. Comments Made When Not Endorsing Acts of Victimization on the CADRI During the Think Aloud Task (N = 150)

|  |  |
| --- | --- |
| Code description | Frequency  *n* (%) |
| Playing around | 35 (23) |
| No intent to hurt | 19 (13) |
| Action normalized | 4 (3) |
| Action hurt | 2 (1) |
| Context clarified | 18 (13) |
| No explanation | 149 (100) |

*Note*. Percentages are based on the total number that provided each respective rationale as justification for not endorsing an act of victimization for at least one item on the *CADRI*. They do not sum to 100 because participants could provide different reasons for each act not endorsed. A total of 149 participants (99%) did not endorse at least one act of victimization during the think aloud task.

### Think Aloud Task

The most common reason given for not endorsing a behavior was because it did not happen, which all participants reported for at least one item. All other reasons were provided relatively infrequently, with the most common being that the behavior occurred in a playful or joking context. A smaller percentage of participants indicated that they did not report a behavior because their partner did not mean to harm them or explained the circumstances in which the behavior occurred.

### Cognitive Interview

Turning to the cognitive interview, two-way multivariate ANOVA indicated that participants’ reasons for not endorsing an act that occurred when their partner was playing around, joking, or roughhousing (Q1) differed by instruction condition, *F*(2, 92) = 6.89, *p* = .002,  = .09, but not sex (*p* = .95) or their interaction (*p* = .34). To identify which codes were associated with instruction condition, we conducted Fisher’s exact tests, which can be used for categorical data even when cell counts are five or less. Only one significant finding emerged (see Table 8): Participants in the Joking condition were more likely than those in the Standard or Intentional conditions to refer to the directions and items on the *CADRI* as their reason for not endorsing an act that occurred in a playful context. For example, one participant said, “No, because it specifically says not including horseplay or joking around.” For the rest of the codes, the most common reason offered for not endorsing a particular behavior was that it occurred while joking or playing around, followed by viewing the behavior as not severe enough to warrant reporting.  
  
Table 8. Reasons for Not Endorsing Acts of Victimization During the Cognitive Interview (N = 150)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cognitive interview question | Playing around  *n* (%) | Depends on severity  *n* (%) | Still happened  *n* (%) | No intent to hurt  *n* (%) | Directions and items  *n* (%) | Don’t know  *n* (%) |
| 1. Would NOT report a behavior that occurred when partner was joking, playing around, or roughousing (*n* = 98, 65%) | 33 (34) | 26 (27) | 2 (2) | 8 (8) | 26 (26) | 3 (3) |
| Standard condition (*n* = 30, 61%) | 14 (47) | 9 (30) | 1 (3) | 3 (10) | 2 (7) | 1 (3) |
| Joking condition (*n* = 42, 78%) | 10 (24) | 8 (19) | 1 (2) | 0 (0) | 21 (50) | 2 (5) |
| Intentional condition (*n* = 26, 55%) | 9 (35) | 9 (35) | 0 (0) | 5(19) | 3 (11) | 0 (0) |
| 2. Would NOT report a behavior that occurred but thought partner did not mean to cause harm (*n* = 64, 43%) | 5 (8) | 12 (19) | 5 (8) | 27 (42) | 8 (12) | 7 (11) |
| 3. Would NOT report a behavior that hurt either physically or emotionally even though partner did not mean to hurt them (*n* = 22, 15%) | 2 (9) | 5 (23) | 0 (0) | 8 (36) | 2 (9) | 5 (23) |

*Note*. Percentages are based on the total number that responded “no” for each respective cognitive interview question.

The reasons given for not endorsing behaviors that were perceived as unintentional (Q2) or were hurtful even if they weren’t intended to harm (Q3) did not differ by instruction condition or sex (*p*s > .23), and thus responses to these two items are combined across groups (see Table 8). In response to both questions, the most common reason given by participants for not endorsing an item was that their partner did not intend to hurt them (e.g., “I would probably say no if it was an accident”), followed by references to the severity or impact of the act (e.g., “because it is not as serious”).

# Discussion

Widely used measures of relationship violence have been criticized for failing to adequately distinguish reports of violent behaviors that are intended to hurt the partner from similar behaviors that occur in playful contexts and are not intended to harm. The latter have been considered “false positives” that overestimate the frequency of violence and reduce the accuracy of the scales. However, adding instructions to not report behaviors that occurred while joking around or roughhousing has not improved the validity of these measures nor consistently reduced the number of behaviors reported (e.g., Jouriles et al., 2009; Sargent et al., 2020). The present study investigated this apparent paradox by examining college students’ reasons for endorsing or not endorsing items on a measure of relationship violence, the *CADRI*. Consistent with prior research (Jouriles et al., 2009; Sargent et al., 2020), we found that *CADRI* scores did not differ among participants who were randomly assigned to one of three instruction conditions: standard instructions, instructions to not report behaviors that occurred in a playful context, and instructions to report only behaviors that were perceived to be intentional. These data provide further evidence that instructions designed to screen out behaviors that occur in a playful context do not reduce overall rates of reporting nor do instructions directing participants to report only intentional behaviors. Further, the findings provide insight into the reasons that college students endorse or fail to endorse items and offer potential explanations for why modifying the instructions may not change their responding.

We explored participants’ reasoning for reporting and not reporting items using a think aloud procedure that elicited their spontaneous thoughts while filling out the measure and a structured cognitive interview that systematically inquired about their reasons for reporting particular kinds of behaviors. With a few exceptions (see below), participants receiving different instructions did not differ in their explanations for endorsing or not endorsing items. The cognitive interview included questions that highlighted the intent and/or impact of particular behaviors by asking respondents whether they would endorse a behavior if it occurred while roughhousing or joking around, if it were perceived to be unintentional, and if it were perceived to be unintentional but caused physical or psychological harm. Just over one third of the participants (35%) stated that they would report a behavior that occurred in a playful context. The rate was slightly lower for participants receiving the “joking” instructions but the difference was not statistically significant. Across instruction conditions, the most common reason (50%) given by participants who said they would report acts that occurred in a playful context was that the behavior happened; another 17% stated that it would depend on how severe the impact was. Thus, for some participants, the fact that the behavior occurred was sufficient reason to report it. In contrast, the majority of participants who stated that they would *not* report a behavior that happened while roughhousing or joking around attributed their decision either to their partners’ lack of intent (34%) or the benign impact of the behavior (27%).

The other two cognitive interview questions asked explicitly about reporting behaviors that were unintentional. Just over half (57%) of participants stated that they would endorse such an item when the impact of the behavior was not specified, while a large majority (85%) indicated that they would report an unintentional behavior that had caused physical or psychological harm. These findings highlight the salience of the impact of behavior to respondents: when a behavior causes harm, its effect outweighed its perceived intent for most participants. Participants in the Joking condition were less likely than those in the Standard condition to say they would endorse a behavior they perceived to be hurtful but unintentional, but 80% still indicated they would report a behavior under these conditions and so it is not the case that these instructions lead most respondents to discount harmful behaviors that occurred while roughhousing or joking around.

The primary reason given for reporting behavior described as unintentional was the same as for behavior that occurred while joking around: because it happened. Across instruction conditions, over three quarters of participants gave this reason for endorsing a behavior even if they thought their partner did not mean to hurt them, and over two thirds for endorsing a behavior that hurt them even if it were unintentional. Those participants who stated that they would *not* endorse these items most often referred to the behavior being unintentional or not being harmful as the reason for not reporting it.

Similar themes emerged in participants’ comments during the think aloud procedure. Respondents often received more than one code when completing the *CADRI*, suggesting that they take multiple factors into consideration when deciding whether particular acts should be reported. The number of participants who described details of the situation (Context clarification = 55%) indicates that many thought carefully about the specific nature of the interaction when deciding whether to endorse an item. Participants often commented on the intentionality and impact of a behavior whether or not they endorsed it. A substantial percentage described an act as unintentional (56%), normative (42%), or playful (20%), but endorsed the item anyway. For example, a participant who endorsed the item “They threw something at me” stated, “Maybe once, but it probably wasn’t something serious, it was probably they threw something in my direction so I could catch it and then I missed.” Consistent with the responses on the cognitive interview, it appears that some participants endorsed behaviors they perceived as unintentional and not harmful simply because they happened. The most common reason participants spontaneously offered for *not* reporting a particular behavior (other than stating that the behavior did not happen) was because their partner was playing around (just under one quarter of the sample), with smaller numbers commenting that their partner did not mean to hurt them or described the context in which the behavior occurred.

Integrating the responses to the think aloud task and cognitive interview leads to several insights about how respondents respond to self-report measures of relationship violence and why instructions to discount behaviors that occur in a playful context may not significantly reduce reporting. First, given that two thirds of the participants indicated that they would not endorse a behavior that occurred while joking around or roughhousing, it appears that many respondents choose not to report playful behaviors when they complete a violence measure and so instructions to exclude acts that occur while joking around or roughhousing would have little impact in those cases. However, research shows that respondents often do endorse behaviors that occur while joking or roughhousing (e.g., Ackerman, 2018; Fernández-González et al., 2013), and some of the participants in this study did as well. The most common reason participants gave for endorsing a behavior that occurred in a playful context was because it happened, and the percentage giving this reason did not differ significantly when participants were instructed not to report behaviors that occurred while joking around or to report only intentional behaviors.

One possible reason that modifying the instructions did not significantly change reporting is that participants did not pay attention to the instructions. However, we found evidence that a sizable proportion of participants were aware of the directions they received. One of the few significant effects of instruction condition showed that more participants in the Joking condition (50%) than the Intentional (11%) or Standard (7%) conditions referred to the instructions when explaining why they did not endorse an act that occurred in a playful context on the cognitive interview.

For many participants, the impact or intent of the behavior were critical factors in deciding whether to endorse it. If an act caused harm, participants were highly likely to report it even if it were viewed as playful or unintentional. For example, one participant explained why they endorsed an item by stating, “There have been a few times where they were kidding but I didn’t take it as kidding. It hurt and I was offended by it.” Similarly, the lack of harm or partner intent were frequently given as reasons for not reporting behaviors that occurred while joking around or roughhousing. However, there also were occasions in which the respondents’ perception of intent differed from their partners’, and in these cases, the respondents’ perception guided their decision to report or not report it. Taken together, these data suggest that aggressive behaviors that occur while joking or roughhousing are not necessarily “false positives”; if the victim was hurt by the behavior or perceived it as intentional, “playful” behavior may have the same impact as a behavior that was intended to be harmful. The reason that some participants referred to a lack of intent or harm as reasons not to report behaviors while others reported items despite viewing them as unintentional or harmful is not clear. Further investigation of how respondents weigh information about intent and impact is needed to better understand their decision-making.

Overall, we found very few significant sex differences in these data. There were no differences between men and women on *CADRI* scores or responses on the think aloud task, but there were two on responses to the cognitive interview. Specifically, men were less likely than women to state they would endorse an act that occurred when their partner was playing around, joking, or roughhousing and to report playful behaviors and behaviors that hurt but were perceived to be unintentional. These findings may reflect the possibility that men view aggressive behaviors by women as less likely to cause serious harm than vice versa. However, the fact that a large majority of men still indicated they would endorse these items suggests that men do not generally discount aggressive behavior by women. We also found no significant interactions between sex and instruction conditions on any variables. Sex differences have been reported inconsistently in prior studies testing the impact of different instructions on measures of violence victimization. For example, Sargent et al. (2020) found very few sex differences across the four studies described in that article; in two of the studies men reported more victimization than women and in one women reported more than men, but these differences did not vary across instruction conditions. In contrast, Hamby (2016a) found that women reported higher rates of victimization than men when instructed not to report joking or horseplay behaviors. The most powerful statistical test of sex differences was reported by Sargent et al. (2020), who conducted a within-study meta-analysis that combined the samples from the four studies (*n* = 2,350) and did not find a significant main effect of sex or an interaction between instruction condition. It may be that differences in responding between men and women are subtle and vary across samples and measures, but additional research utilizing large samples is needed to draw a more definitive conclusion.

# Limitations

The limitations to the study should be considered when interpreting the results. First, the sample consisted of predominantly White university students, and the findings may not generalize to more racially and ethnically diverse populations, to older or younger individuals, or to clinical populations. Second, nearly all of the participants were involved in casual dating relationships, and individuals who are in long-term committed relationships may perceive their partners’ intent or the impact of a behavior differently. Third, it is possible there were small effects that we did not have sufficient power to detect. Interaction effects, in particular, require more power, and so we cannot interpret the lack of significant interactions as evidence that no such effects exist. As noted above, the within-sample meta-analysis conducted by Sargent et al. (2020; Study 5) also did not find significant effects of sex, instruction condition, or their interaction on reports of violence in a much larger sample, but it will be important to replicate the findings concerning participants’ reasons for endorsing or not endorsing particular items. Finally, the think aloud task and cognitive interview relied on participants’ willingness to describe their thoughts and beliefs about a potentially emotionally difficult topic and it is possible there were factors influencing their decisions that they did not report.

## Research Implications

Definitions of violence in psychological research typically have emphasized the intent of the behavior for distinguishing aggressive and nonaggressive behavior, and attempts to improve self-report measures of relationship violence have focused on reducing reports of acts that are not intended to harm. However, the present findings indicate that behaviors that occur in a playful context can be perceived as intentional and harmful by the victim and consequently cannot be assumed to be “false positives” or “overreporting” (Ackerman, 2018). Although our findings show that respondents do attend to their partners’ intent, they suggest that measures of relationship violence may more effectively distinguish violent from nonviolent behaviors by focusing on the impact of the behavior. This could be assessed by adding checkboxes after each item that respondents could mark if they had been hurt physically or psychologically by a particular act. It would then be important to empirically test whether summing the number of acts marked as harmful improves the validity of the measure relative to the total number of behaviors endorsed or the number of behaviors that were endorsed but not marked as harmful. This approach also would provide potentially important information about behaviors that participants endorsed simply because they happened. Another approach to assessing the impact of aggressive behavior is reflected in the *Revised Conflict Tactics Scale* (Straus et al., 1996), which added a subscale (Injury) assessing whether the respondent was injured by their partner (e.g., were they cut or bleeding, did they need to seek medical care). However, the Injury subscale rarely is reported in studies of intimate partner violence, perhaps because it increases the length of the questionnaire or primarily assesses fairly serious physical injuries that are less commonly found in community or college samples. Including a checkbox for each item likely would add less time than having participants complete an additional subscale and would enable respondents to define “harm” more subjectively and broadly.

## Clinical Implications

The present findings underscore the importance of understanding individuals’ subjective experience of behaviors that occur in relationships. In clinical or health settings where screening questionnaires are used to identify intimate partner violence, it is important to follow up on any endorsed items with open-ended questions that assess individuals’ perceptions of the intent and effect of the behavior, even if they occurred in a playful context. Individuals who experience abusive behavior on a frequent basis may be more likely to perceive ambiguous or ostensibly “joking” behavior as intentional, and chronic victimization in turn increases the risk for experiencing greater mental health difficulties. In contrast, some victims may downplay abusive behavior if their partner claims to be “kidding” or describes it as accidental, and so it is critical to ask about the victim’s perception and any harm that resulted.

# [References](https://web.p.ebscohost.com/ehost/detail/detail?vid=2&sid=3a90ade6-fe92-4847-9030-d499777f5be4%40redis&bdata=JkF1dGhUeXBlPXNzbyZzaXRlPWVob3N0LWxpdmUmc2NvcGU9c2l0ZQ%3d%3d#toc)

Ackerman, N. (2018). Assessing Conflict Tactics Scale validity by examining intimate partner violence overreporting. ***Psychology of Violence***, *8*(2), 207–217. 10.1037/vio0000112

Arriaga, X. B. (2002). Joking violence among highly committed individuals. *Journal of Interpersonal Violence*, *17*(6), 591–610. 10.1177/0886260502017006001

Cascardi, M., Blank, S., & Dodani, V. (2019). Comparison of the CADRI and CTS2 for measuring psychological and physical dating violence perpetration and victimization. *Journal of Interpersonal Violence*, *34*(16), 3466–3491. 10.1177/0886260516670182

DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship*, *39*(2), 155–164. 10.1111/j.1547-5069.2007.00161.x

DeWall, C. N., Anderson, C. A., & Bushman, B. J. (2011). The general aggression model: Theoretical extensions to violence. ***Psychology of Violence***, *1*(3), 245–258. 10.1037/a0023842

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, *87*(3), 215–251. 10.1037/0033-295X.87.3.215

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

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*(3), 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*(5), 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*(8), 1019–1023. 10.1007/s10896-016-9878-z

Hamby, S. (2016a). Self-report measures that do not produce gender parity in intimate partner violence: A multi-study investigation. ***Psychology of Violence***, *6*(2), 323–335. 10.1037/a0038207

Hamby, S. (2016b). Advancing survey science for intimate partner violence: The Partner Victimization Scale and other innovations. ***Psychology of Violence***, *6*(2), 352–359. 10.1037/vio0000053

Hamby, S., & Grych, J. (2013). Evidence-based interventions need to be more systematic, not more disruptive. *American Psychologist*, *68*(6), 476–477. 10.1037/a0032884

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238–247. 10.1037/1040-3590.7.3.238

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277–1288. 10.1177/1049732305276687

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 & Neglect: The International Journal*, *33*(7), 451–460. 10.1016/j.chiabu.2008.11.005

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*(4), 477–490. 10.1037/a0037404

Sargent, K. S., Yule, K., Bridges, K., Jouriles, E. N., & Grych, J. H. (2020). Do instructions intended to reduce false positives improve the measurement of physical partner violence victimization among adolescents and young adults? ***Psychology of Violence***, *10*(2), 152–161. 10.1037/vio0000246

Sillito, C. (2012). Gendered physical and emotional health consequences of situational couple violence for heterosexual married and cohabiting couples. *Feminist Criminology*, *7*(4), 255–281. 10.1177/1557085111431695

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. 10.1177/019251396017003001

Willis, G. B. (2005). Cognitive interviewing in practice: Think-aloud, verbal probing and other techniques. *Cognitive interviewing: A tool for improving questionnaire design* (pp. 42–63). Sage Publications. 10.4135/9781412983655.n4

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*(2), 277–293. 10.1037/1040-3590.13.2.277

World Health Organization. (2002). *World report on violence and health: Summary*.

Zhang, Y., & Wildemuth, B. M. (2009). Unstructured interviews. In B. M. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science* (pp. 222–231). Libraries Unlimited.