Item Response Theory Analysis of Intimate-Partner Violence in a Community Sample

Anita Jose
Stony Brook University

Thomas M. Olino
University of Pittsburgh

K. Daniel O'Leary
Stony Brook University

The Revised Conflict Tactics Scale (CTS2; Straus, Hamby, Boney-McCoy, & Sugarman, 1996) is a widely used measure of physical, psychological, and sexual aggression as well as injury and negotiation between partners. In this study, we analyzed male-to-female and female-to-male physical aggression using data from 453 community couples. We used item-response theory (IRT) to assess the range and precision of physical aggression severity in the past year, as evaluated by the CTS2. Our analyses support a single dimension of physical aggression with quantitative, rather than qualitative, differentiation between minor and severe physical aggression items. Surprisingly, test information curves revealed that male perpetration and victimization items provided as much or more information as their counterpart female victimization and perpetration items over the same range of aggression severity. Finally, the data suggests that CTS2 items best assessed moderate-to-severe levels of physical aggression in the previous 12 months. However, virtually no information was assessed by male or female reports of perpetration or victimization items below or around the mean of aggression (i.e., theta; $\theta = 0$). Suggestions for improved item coverage and implications for the assessment of aggression are discussed.

Keywords: conflict tactics scale, intimate partner violence, item-response theory, relationship aggression

Over 10% of women and men in a nationally representative United States sample have reported victimization by their partners within the prior 12 months (Straus & Gelles, 1990), reflecting the high prevalence of intimate partner violence in this country. Similar rates have been found in other nationally representative samples (e.g., Schafer, Caetano & Clark, 1998). Further, physical aggression has been identified as a frequent occurrence in community samples. For instance, a longitudinal study of young couples indicated that 31% of men and 44% of women reported perpetrating aggression against their partners in the year prior to marriage (O'Leary et al., 1989).

To date, a wealth of psychological research has relied on the original (CTS; Straus, 1979, 1990) and revised (CTS2; Straus et al., 1996) Conflict Tactics Scale to better understand aggression in relationships. Conventionally, the CTS2 physical assault scale includes two subscales, one reflecting minor aggression (e.g., slapping) and one reflecting severe aggression (e.g., punching or kicking). According to Straus and colleagues (Straus et al., 1996), factor analyses of the original CTS (Straus, 1979) have supported this bidimensional conceptualization of violence. However, other research suggests that physical aggression is unidimensional. For instance, Barling and colleagues (Barling, O'Leary, Jouriles, Vivian, & MacEwen, 1987) factor-analyzed data from couples 6 months postmarriage and individuals seeking couple therapy using a modified version of the CTS. Their results indicated that almost all physical aggression items loaded onto one factor, which explained 36% of the variance in clinic husbands' responses, and 34% of the variance in community couples' responses. Thus, all items reflected a common aggression factor, without indicating the need for qualitatively different subfactors to explain shared variance among the items. Barling and colleagues' (1987) data also indicated that certain items (e.g., "Threatened to use a knife or a gun," "Beat him/her up") loaded onto the physical aggression factor in the clinic sample but not the community sample, suggesting that these very low base-rate CTS items are not commonly endorsed in the general public.

Work on the CTS and CTS2 has often relied on a single informant of violence, with either male or female partners contributing data about the level of aggression in the entire dyad, with the implicit assumption that the items assessing perpetration and victimization are functioning similarly across informant. However, research findings have demonstrated only moderate agreement between partners for various relationship behaviors (e.g., Christensen, Sullaway, & King, 1983). Thus, items assessing perpetration and victimization may provide different kinds of data about

aggression. Addressing this possibility requires the use of nonclassical measurement techniques.

Measuring Relationship Aggression Using Item Response Theory (IRT)

IRT is based on the understanding that a respondent's score on each individual item reflects his or her *ability* level (theta; θ) on an underlying construct. IRT models the level of this underlying trait (e.g., male-to-female aggression) as a function of two item characteristics: Discrimination and difficulty (Embretson & Reise, 2000). Difficulty (*b*) is the level of the measured construct at which there is a 50% chance of endorsing the item of interest in the keyed direction ($p(\theta) = 0.5$). Curves that are further right along the *x*-axis reflect more difficult items (e.g., the respondent's level of the measured construct must be higher to endorse the keyed item). Discrimination (*a*) is an index of how rapidly the probability of the keyed endorsement changes at the trait level corresponding to the item difficulty, and therefore, how well an item differentiates between levels of the measured construct. Steeper item-characteristic curve (ICC) slopes indicate greater discrimination. Results of IRT analyses provide estimates about the position of the latent trait (θ) where each item or collection of items provides the most information. Results can be summarized visually by producing item- or test-information functions, which display regions of the latent trait that best assess psychometric information. These functions are standardized to have a mean of zero and a standard deviation of one, permitting comparisons across multiple inventories (Reise & Henson, 2003).

There are only three published reports examining physical aggression using IRT methods, of which we are aware. Schafer (1996) analyzed the eight original CTS (Straus, 1979, 1990) Form N items separately for college-aged men's (*n* = 263) and women's (*n* = 270) self-reported perpetration. Here, goodness-of-fit of the IRT model differed across genders, with good model fit for women's reports but poor model fit for men's reports. Visual inspection indicates adequate discrimination for all women's items except possibly the "Used a knife or a gun" item, but adequate discrimination for only four of the eight men's items. Schafer's results also indicated that CTS items tended to assess higher levels of the latent trait of violence, with little coverage for aggression at lower levels of θ.

In addition, a recent paper by Regan and colleagues (Regan, Bartholomew, Kwong, Trinke, & Henderson, 2006) used a 14-item measure of relationship violence, selecting items from the CTS (Straus, 1979, 1990) and CTS2 (Straus et al., 1996) for use in a Canadian community sample of 531 men and 561 women. These men and women came from independent couple-dyads and so were not in a relationship with each other. Although annual prevalence rates of physical aggression in this sample were too low for analysis, IRT analyses of lifetime aggression indicated that all physical aggression items demonstrated relatively high discrimination, and a range of difficulty levels. Furthermore, perpetrators and victims consistently rated the severity of aggression items similarly. However, some heterogeneity among minor and severe item difficulty was observed. This suggests that some items that are considered minor aggression may in fact be reflective of a higher level of the latent trait of physical aggression toward the partner. For this reason, the authors suggest physical aggression should not be differentiated as minor or severe, but rather that the construct reflects a single continuum.

Another recent article used IRT techniques to measure physical aggression in same-sex male couples (Regan, Bartholomew, Oram, & Landolt, 2002). In this research, 284 gay and bisexual men completed a 14-item measure based on the CTS (Straus, 1979, 1990) and CTS2 (Straus et al., 1996) where they reported physical aggression perpetration and victimization. Concurrent with Regan and colleagues' (2006) findings, these results indicated that lifetime perpetration and victimization, and victimization in the past 12 months had relatively high discrimination, and a range of item difficulty. Once again, Regan and colleagues (2002) found that there was some heterogeneity in the placement of minor and severe items along the *x*-axis.

The Present Study

The present study adds to the existing literature on the CTS2 assessment of physical assault, long considered the "gold standard" for assessing relationship violence, in important ways. We conduct IRT analyses of male-to-female and female-to-male relationship violence data to assess the quality and coverage of CTS2 physical assault items. The novelty of this study lies in the following:

First, we use rates of aggression within the past year. Although Regan and colleagues' (2006) research on opposite sex aggression only provided results on lifetime aggression, rates within the last 12 months are likely less susceptible to memory decay and therefore more clinically relevant. Similar findings about the effects of time on accurate recall of various events have been reported in the literature (e.g., Jenkins, Earle-Richardson, Slingerland, & May, 2002; Mock, Acheampong, Adjei, & Koepsell, 1999). Further, analyzing past year rates will also maximize the likelihood that both partners are reporting on aggression within the current relationship.

Second, the IRT studies described above used older versions of the CTS, or modified scales incorporating items from both the CTS and CTS2. However, in clinical practice, we anticipate that only one measure will be used, and that it will likely be the current version of the Conflict Tactics Scale. To this end, we use the complete physical assault subscale of the CTS2 in our analyses.

Finally, perhaps the most novel aspect of this study is that we use reports from both partners to compare the range and precision of information obtained from perpetrator reports and their partners' victimization reports over the same time period. This provides an opportunity to compare the breadth and depth of data provided across informants within the same dyad. Such analyses have not yet been presented in the literature.

Male-to-female violence is most commonly studied since it is associated with significant negative consequences for the victim. For instance, Kar and O'Leary (2010) found that physically victimized women reported more fear of their partners than physically victimized men. In addition, research suggests elevations in depressive, anxious, and trauma-related symptoms in women who are victims of physical aggression (e.g., Campbell, 2002; Cascardi & O'Leary, 1993; Coker, Smith, Bethea, King, & McKeown, 2000). Finally, women are more likely to be injured and unable to work than men as a consequence of relationship aggression (e.g., Archer, 2000; Tjaden & Thoennes, 2000). However, since the

literature has also consistently documented higher rates of female-to-male physical aggression than the converse (e.g., Archer, 2000), the present study addresses both male-perpetrated physical aggression (and concomitant female victimization) and aggression in the opposite direction (female-perpetrated physical aggression and male victimization).

Based on previous results, we anticipate that the CTS will demonstrate quantitative, as opposed to qualitative, distinctions between items. Specifically, we hypothesize that our item pool will be unidimensional, but that individual items will exhibit a range of difficulty levels across the latent trait. As individuals in a United States community-based sample may not frequently endorse items reflecting very severe aggression, we anticipate that the CTS2 will best assess information at moderate levels of aggression for both victimization and perpetration items. Reflecting convention in the relationship violence literature, we also hypothesize that victimization items, regardless of whether men or women are the victims, will consistently provide more information across the range of aggression when compared to perpetration items.

Method

The present study utilizes data from a large, multivariate data set including 453 couples with young children in Suffolk County, NY (O'Leary, Slep, & O'Leary, 2007). The original study was approved by the institutional review board of Stony Brook University, and informed consent was obtained from all participants. The current analyses examine the range of severity of physical relationship aggression assessed by the CTS2 using IRT methods.

Participants

This study used random-digit dialing to recruit participants in the Suffolk County, NY area. Adult respondents were screened to determine that they met criteria for the study. As the original study for which these data were collected focused on relationship and parenting variables, eligible respondents had been living as a couple for the previous year and had a 3–7-year-old child that was the biological child of at least one partner in the dyad. In addition, participants needed to be able to complete assessments in English.

Those who met all of the above criteria were interviewed on the telephone about family functioning. Of the 1,815 couples who met eligibility criteria and expressed interest in participating, data were collected from 453 couples.

Measures

Conflict Tactics Scale, Revised (CTS2;Straus et al., 1996). This commonly used 78-item questionnaire assesses behaviors, including negotiation, psychological aggression, physical assault, injury, and sexual coercion, during relationship conflict within the past year. Thirty-nine items ask about perpetration and 39 items ask about victimization. Each scale, except negotiation, includes subscales for minor and severe aggression. Coefficient alpha was high for all scales in Straus and colleagues' (1996) original study (ranging from 0.79 to 0.95). Only the physical assault scale was used in this study, for which reliability analyses were conducted using SPSS v.17. For male-to-female violence, coefficient alpha was 0.73 for male perpetration and 0.82 for female victimization. Female-to-male violence was analyzed without counting item pair 61/62 due to lack of variance in either female or male reports; alpha was 0.77 for female perpetration and 0.83 for male victimization.

Procedure

Participants in this study completed the CTS2 independently of their partners. A more extensive description of the study procedure and study methods in general can be obtained from other sources (e.g., O'Leary et al., 2007; Slep & O'Leary, 2007). The annual prevalence of aggressive behaviors is depicted in Table 1.

We dichotomized participants' CTS2 responses (presence vs. absence of aggression, keyed such that the presence of aggression was the outcome of interest) when conducting the following analyses, as many of the frequency-based response options were not endorsed. We considered collapsing some options together; however, this would yield a different number of response options for different items and would make item comparisons unbalanced. Consistent with this rationale, across male perpetration and victimization and female perpetration and victimization, there were a

Table 1
Annual Prevalence of Partner Aggression

Item	Male Report (%)		Female Report (%)	
	Perpetration	Victimization	Perpetration	Victimization
Threw something that could hurt	7.5	15.5	13.7	10.6
Twisted arm or hair	8.2	8.8	4.6	7.1
Pushed or shoved	17.2	21.2	26.0	18.5
Used a knife or gun	0.4	0.7	0.7	0.2
Punched or hit with something that could hurt	2.9	7.7	6.2	3.1
Choked	2.0	1.8	0.4	1.1
Slammed against a wall	4.4	3.8	1.5	3.3
Beat up	0.7	1.8	0.7	1.3
Grabbed	16.6	16.6	14.9	13.2
Slapped	2.9	7.5	8.4	3.5
Burned or scalded	1.1	1.1	0.0	0.0
Kicked	2.0	5.5	6.4	3.3
Any Physical	25.4	28.5	34.2	29.8

total of 1377 endorsements of past-year incidents, regardless of the frequency of the behaviors. Of these instances, 671 (48.7%) were of individuals indicating that a behavior occurred on one occasion. Thus, much of the variance in past-year responses comes from individuals reporting only one instance of specific aggressive behaviors in the past year.

Results

Data Analytic Plan

As detailed above, participants' responses were transformed into dichotomous variables reflecting *occurrence* (1) and *nonoccurrence* (0) of aggression in the past year. The sample was randomly split in two, and exploratory factor analyses (EFA) were conducted on one half of the cases, while confirmatory factor analyses (CFA) were conducted on the other half. This was done separately for male reports (perpetration and victimization), female reports (perpetration and victimization), and couple-level data (male-to-female and female-to-male). All factor analyses were conducted to assess unidimensionality of the scales (a prerequisite for conducting IRT analyses), using Mplus v.6 (Muthén & Muthén, 2010). EFA results supporting unidimensionality are indicated by at least a 4 to 1 ratio of the first to second eigenvalues (Embretson & Reise, 2000). CFA results supporting unidimensionality are indicated by CFI > .90 indicating adequate fit and CFI > .95 indicating good fit, and root mean square error of approximation (RMSEA) values between .05 and .10 indicating acceptable fit (Steiger, 1990) and values less than .05 indicating good fit (McDonald & Ho, 2002).

IRT models were estimated for male-to-female aggression, and separately, for female-to-male aggression in the 12 months preceding data collection. Analyses were conducted with Multilog v.7.03 (Thissen, Chen, & Bock, 2003), using a two parameter logistic (2PL) model. Finally, information curves were created using the open source R v.2.11.1 statistical program (Ihaka & Gentleman, 1996).

Descriptive Information

Data on male-to-female aggression and female-to-male aggression were collected and analyzed from 453 heterosexual couples. The average age of the male partner was 37.25 ($SD = 6.02$) years, and he reported having 14.24 ($SD = 2.34$) years of education. The average age of the female partner was 35.06 ($SD = 5.00$) years, and she reported having 14.28 ($SD = 2.26$) years of education. Couples had an average of 2.36 ($SD = 1.00$) children in the household. Table 1 provides the annual prevalence reported for each physical aggression item in the analysis. Overall, 25.4% of men and 28.5% of women in this sample reported the occurrence of any male-to-female physical aggression in the past year. When female-to-male aggression was measured, 34.2% of women and 29.8% of men reported it had occurred in the past year. These rates are higher than self-reported annual aggression rates in representative samples, which tend to hover around 10% (Jose & O'Leary, 2009), perhaps because of the young age of participants in this study, compared to a nationally representative sample (O'Leary & Woodin, 2005).

Item unidimensionality. Although some studies have suggested that physical aggression is multidimensional (e.g., Pan, Neidig, & O'Leary, 1994; using a large military sample), other research suggests that the construct is comprised of one factor (e.g., Barling et al., 1987; Regan et al., 2006; Schafer, 1996). For these reasons, we split the sample and conducted both EFA and one-factor CFA analyses to assess unidimensionality.

Mplus v.6 (Muthén & Muthén, 2010) was used to conduct all EFA and CFA analyses; results can be found in Table 2. Although the sample was randomly split in half, low base-rate items (e.g., items with fewer than five endorsements overall) were included in both EFA and CFA models. Although we did this reluctantly, as it introduced nonrandom selection into the analyses, we felt that including these items would enhance the generalizability of these analyses. Item pair 61/62 ("Burned or scalded") was removed from female reports for all analyses (individual and couple-level data) due to lack of variance for the item. Despite the presence of some empty cells, the model terminated normally. For EFA analyses, the ratio of the first to second eigenvalues for all but one analysis (couple reports of female-to-male aggression) exceeded the recommendation of 4:1. The average ratio for all six analyses was well beyond the threshold, at 5.5:1. CFA analyses assumed a 1-factor solution a priori, and they were conducted using the WLSMV (mean-and-variance-adjusted weighted least squares) estimator, which provides valid fit statistics for CFA models using dichotomous observed variables (Flora & Curran, 2004). As seen in Table 2, all one-factor models demonstrated good fit to the data.

Table 2
Factor Analytic Results

	EFA (1-Factor)		CFA (1-Factor)			
	N	1st:2nd Eigenvalue	*N*	*CFI*	*TLI*	*RMSEA*
Male-to-Female Aggression						
Male Perpetration	230	8.11:1.07	230	.99	.99	.04
Female Victimization	227	6.31:0.99	228	.99	.98	.06
Couple Level Data	232	14.94:3.45	233	.95	.94	.06
Female-to-Male Aggression						
Female Perpetration	230	7.40:1.56	230	.98	.97	.05
Male Victimization	230	8.38:1.21	232	.99	.99	.04
Couple Level Data	234	11.81:3.89	230	.90	.89	.07

IRT Analyses

The data were examined with a 2PL IRT model, using Multilog 7.03 (Thissen et al., 2003). Results are presented in Tables 3 and 4, which model both partners' reports of each behavior. Consistent with the analysis above, female reports of item pair 61/62 ("Burned or scalded") were removed from both analyses during the factor analysis step due to lack of variance (i.e., the item was not endorsed by any participant). Item 34 ("Choked") was also removed from female reports of male-to-female aggression (see Table 3) after a preliminary IRT calibration revealed that it had an extreme value on the difficulty parameter.

Male-to-female aggression. Our results were based on calibrating the full 12-item scale for male perpetration and the 10-item scale for female victimization simultaneously (without female victimization items 34 or 62). A similar pattern of results was found when male perpetration items were limited to the 10 items that corresponded to the 10 female victimization items, thus, the results are not due to an unbalanced number of items.

Discrimination values over 1.00 are typically viewed as adequate in IRT analyses. For the most part, item discrimination was high in this sample; averaging 2.75 across all analyses (the range was $0.78 \leq a \leq 4.22$), indicating that most items differentiated well between individuals at higher and lower levels of aggression severity. Overall, male perpetration items (mean $a = 3.06$; range: $0.78 \leq a \leq 4.22$) demonstrated greater average discrimination but also a larger range, compared to female victimization items (mean $a = 2.37$; range: $1.87 \leq a \leq 2.90$). Additionally, discrimination was generally slightly higher for minor aggression items compared to severe aggression items. This pattern was clearest when comparing minor (mean $a = 3.42$; range: $2.05 \leq a \leq 4.22$) and severe (mean $a = 2.80$; range: $0.78 \leq a \leq 4.16$) male perpetration items. Item 61 seemed to drive the larger range and lower mean discrimination of the severe aggression items.

The average difficulty level (severity location on the x-axis) is fixed at the mean ($\theta = 0$) and, as expected, minor aggression items generally exhibited a lower level of difficulty for both male and female participants. Minor aggression item difficulty ranged from 1.04 to 2.09 (mean $b = 1.53$) for males, and 1.11 to 2.09 (mean $b = 1.62$) for females. As expected, difficulty was typically greater

for severe aggression items: male perpetration items ranged from 1.95 to 6.12 (mean $b = 2.92$), and female victimization items ranged from 2.10 to 4.01 (mean $b = 2.68$). However, one minor perpetration item ("Slapped") was rated as higher in difficulty than two severe aggression items ("Slammed against a wall" and "Punched or hit with something that could hurt") for males. Although there was no overlap between minor and severe aggression items for female victimization, the minor aggression item "Slapped" had a difficulty level of just 0.01 units fewer than the severe aggression item "Slammed against a wall" for females.

A graph illustrating the informational coverage of male perpetration and female victimization items can be found in Figure 1. Information for overall male-to-female aggression (the solid black line) is best assessed at the peak of the curve, located at approximately two standard deviations above the mean, or $\theta = 2$. At and below the mean, male and female reports seem to provide similar (and limited) levels of information. However, across most points above the mean, male perpetration reports provide much more information about aggression than female victimization reports, meaning that male perpetration items provide more precise data relative to female victimization items, about the severity of male-to-female aggression.

Female-to-male aggression. Our results were based on calibrating the 11 female perpetration items (without female perpetration Item 61, "Burned or scalded", which was not endorsed in the sample) and the 12 male victimization items simultaneously. Discrimination and difficulty parameters are presented in Table 4. Item discrimination averaged 2.51 across all analyses (the total range was $0.82 \leq a \leq 4.29$), indicating adequate overall item discrimination. Male victimization items demonstrated a higher average discrimination and also a larger range of values (mean $a = 3.05$; $0.82 \leq a \leq 4.29$) compared to female perpetration items (mean $a = 1.92$; range: $1.41 \leq a \leq 2.49$). Mean item discrimination was generally higher for minor (mean $a = 3.29$; range: $2.89 \leq a \leq 4.19$) and severe (mean $a = 2.87$; range: $0.82 \leq a \leq 4.29$) male victimization items compared to female perpetration, at both the minor (mean $a = 1.95$; range: $1.62 \leq a \leq 2.46$) and severe (mean $a = 1.90$; range: $1.41 \leq a \leq 2.49$) levels. As with male-

Table 3

Discrimination and Difficulty Parameters for Male-to-Female Physical Aggression in the Past Year

Type			Perpetrator Report (Male Report)		Victim Report (Female Report)	
Minor or Severe	CTS2 Pair	Item	Discrimination	Difficulty	Discrimination	Difficulty
M	7/8	Threw something that could hurt	2.05	1.88	1.87	1.70
M	9/10	Twisted arm or hair	3.70	1.53	2.52	1.78
M	17/18	Pushed or shoved	4.22	1.04	2.40	1.11
S	21/22	Used a knife or gun	3.08	2.91	1.92	4.01
S	27/28	Punched or hit with something that could hurt	3.51	2.05	2.37	2.29
S	33/34	Choked	2.19	2.61	—	—
S	37/38	Slammed against a wall	2.90	1.95	2.90	2.10
S	43/44	Beat up	4.16	2.52	2.82	2.54
M	45/46	Grabbed	3.82	1.09	2.16	1.43
M	53/54	Slapped	3.29	2.09	2.78	2.09
S	61/62	Burned or scalded	.78	6.12	—	—
S	73/74	Kicked	3.00	2.31	1.98	2.44

Table 4

Discrimination and Difficulty Parameters For Female-to-Male Physical Aggression in the Past Year

Type			Perpetrator Report (Female Report)		Victim Report (Male Report)	
Minor or Severe	CTS2 Pair	Item	Discrimination	Difficulty	Discrimination	Difficulty
M	7/8	Threw something that could hurt	1.83	1.52	3.11	1.20
M	9/10	Twisted arm or hair	1.62	2.48	2.89	1.59
M	17/18	Pushed or shoved	1.96	0.87	4.19	0.91
S	21/22	Used a knife or gun	1.80	3.54	2.93	2.77
S	27/28	Punched or hit with something that could hurt	2.26	1.95	2.13	1.64
S	33/34	Choked	1.84	3.71	2.95	2.37
S	37/38	Slammed against a wall	2.49	2.59	3.16	2.00
S	43/44	Beat up	1.58	3.85	4.29	2.16
M	45/46	Grabbed	1.89	1.46	3.04	1.15
M	53/54	Slapped	2.46	1.70	3.23	1.64
S	61/62	Burned or scalded	—	—	.82	5.87
S	73/74	Kicked	1.41	2.43	2.81	1.87

to-female aggression items, female-to-male discrimination only fell lower than 1.00 for the burning or scalding item.

In terms of item difficulty, severe aggression items were, on average, reflective of higher levels of the latent trait compared to minor aggression items. For female perpetration items, minor aggression items ranged in difficulty from 0.87 to 2.48 (mean $b = 1.61$) and severe aggression items were generally more difficult, with values ranging from 1.95 to 3.85 (mean $b = 2.85$). However, Item 9 ("Twisted arm or hair") had a difficulty level of 2.48 (within the severe aggression range) despite being a minor aggression item. For male victimization items, the mean difficulty level for minor aggression was 1.30 (range: $0.91 \leq b \leq 1.64$) while severe aggression difficulty averaged 2.67 (range: $1.64 \leq b \leq 5.87$). Minor aggression item 54 ("Slapped") and severe aggression item 28 ("Punched or hit with something that could hurt) both had identical difficulty levels.

A graph illustrating the informational coverage of male victimization and female perpetration items can be found in Figure 2. Accounting for both partners' reports together (the solid black line), information is best assessed around two standard deviations above the mean ($\theta = 2$). However, male victimization items provide more information about underlying aggression than female perpetration items. For both male victimization and female perpetration items, aggression is best assessed between one and three standard deviations from the mean. Similar to male-to-female aggression, almost no information is obtained at or below the mean ($\theta \leq 0$).

Discussion

Overall, IRT results suggest that the CTS2 physical assault scale assesses physical aggression at moderate to high levels for both
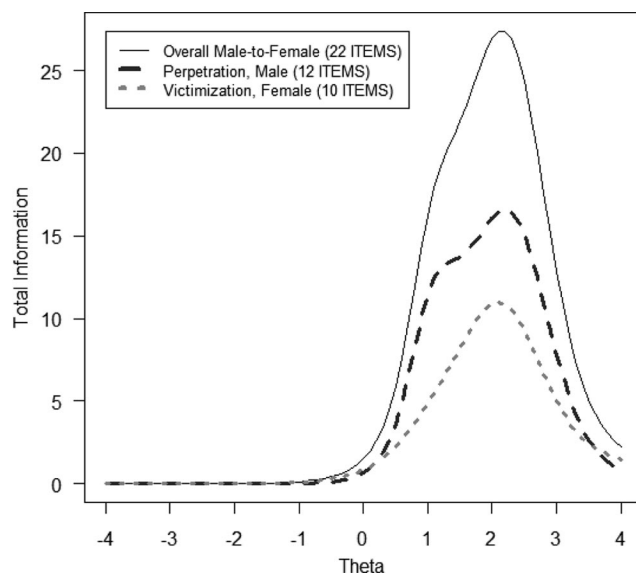

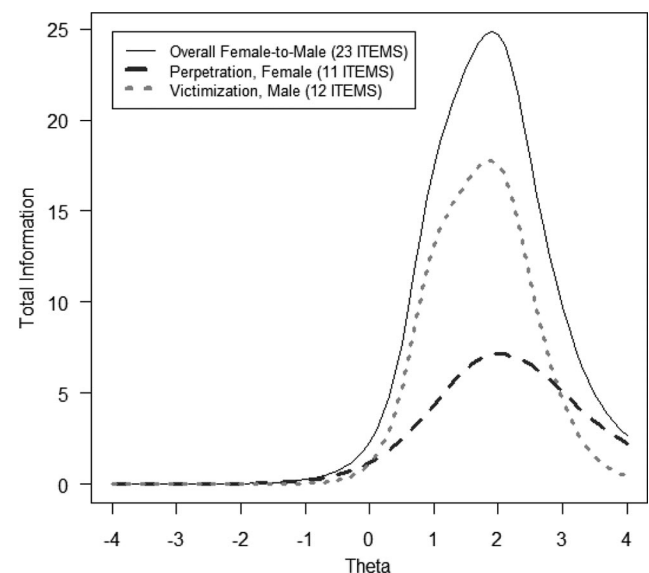*Figure 1.* Information: Male-to-Female Physical Aggression in the Past Year.

Figure 2. Information: Female-to-Male Physical Aggression in the Past Year.

male-to-female and female-to-male aggression, but is limited in its ability to capture information about lower level aggression, especially at or below mean levels. The results also revealed, surprisingly, that greater information is generally provided by male reported items for both perpetration and victimization. We consider it important to note that our analyses reflect the ability of the CTS *items* to assess aggression severity, rather than assessing differences in male and female *levels* of aggression.

Unlike other IRT research on this topic, this study examined data collected from men and women from the same dyad. As is evident from the information curves presented (Figures 1 and 2), using both partners' item responses provided more information than using only one partner's responses when reporting male-to-female and female-to-male aggression during the prior year. Surprisingly, male reported items provided a greater amount of information than female items for both perpetration and victimization, at all except the lowest and highest levels of aggression. Although this data should be interpreted cautiously due to the relatively small sample size, our results suggest that if both partners' reports are available, items assessing male reports may provide more information with greater precision than female reports, especially for moderate levels of aggression in the year prior. Of course, although male items provide more *information* than female reports, these analyses do not speak to gender differences in the *prevalence* of aggression.

Most individual items differentiated at least fairly well between higher and lower levels of physical aggression, based on their discrimination values. As a precondition of our IRT analyses, factor analytic results indicate that CTS2 physical assault items lie on a single continuum, reflecting quantitative differences, rather than reflecting two qualitatively distinct subtypes of aggression (minor and severe). Thus, examining associations for "minor" and "severe" aggression and important outcomes will likely be biased due to statistical (e.g., restriction of range) and conceptual (e.g., limited construct measurement; reifying artificial distinctions) considerations.

Based on the IRT analyses, the CTS2 physical assault scale assesses aggression trait levels from roughly one standard deviation to three standard deviations above the mean for both male-to-female and female-to-male aggression, with the most precise information being assessed at approximately two standard deviations above the mean. However, coverage is limited at lower levels of the construct. Aggression may be underestimated as behaviors at lower trait levels are not assessed by this scale. Based on this finding, it is important to develop and include additional items that reflect lower levels of the latent trait. For instance, new items (e.g., items reflecting nonpartner or nonfamily aggression), or possibly high-level psychological relationship aggression items may be used in conjunction with physical aggression items in an attempt to increase the coverage of this construct. Although analyzing the role of psychological aggression items in the trait of physical aggression is beyond the scope of the present study, the inclusion of items that provide data for the lower levels of the construct can provide a clearer picture of aggression for individuals in the general public.

In the context of couple interventions, the CTS2 is a valuable instrument to assess aggression at intake, as it will provide important details about the level of ongoing aggression. However, as the sensitivity of measurement is modest at lower levels of aggression,

this instrument may need to be supplemented with additional items that are not currently assessed by the CTS2. Reflective of the truism in psychology that the best predictor of future behavior is past behavior, identifying past incidents at low levels of the latent trait (physical aggression) may allow greater predictability about future aggression. This in turn may allow clinicians to better intervene with couples at higher risk for physical aggression.

Limitations and Future Directions

Our findings should be interpreted in light of the fact that these data come from a community-based sample of parents. Despite the relatively high overall prevalence of aggression found in this sample, participants only infrequently reported the occurrence of certain severely aggressive behaviors. Specifically, although "Choked" was included in the analysis of female-to-male violence, it was excluded from IRT analysis of female victimization as it was an outlier. Similarly, there was no variance (due to a lack of any endorsements) in female reports of perpetration or victimization for the "Burned or scalded" item. We believe these items may be better captured in a sample exhibiting more severe violence (e.g., individuals in couples therapy, individuals in batterer programs or arrested for domestic violence).

As the participants in the current study do not report engaging in aggression as frequently or as severely as participants in clinical samples (e.g., Cascardi, Langhinrichsen, & Vivian, 1992; Simpson & Christensen, 2005), IRT analyses conducted using data from samples at higher risk for relationship violence can supplement the results of the present study. Studying samples with higher levels of aggression may also facilitate increased assessment precision as CTS items may be examined using the frequency categories, as opposed to dichotomous (e.g., present vs. absent) coding.

Further, these data are from a sample of parents of young children. As expected with younger couples, the prevalence of physical aggression is higher than in a sample that would be representative of the full age range of the population (O'Leary & Woodin, 2005). Future research may benefit from random sampling to obtain a sample that spans the full age range.

Future research may also include incorporating items to test a broader range of the latent trait measured here. This may mean including items reflecting higher levels of psychological aggression or other types of aggression (e.g., nonfamily aggression), as well as items reflecting intensity of the aggressive behavior. If factor analyses on an item pool enlarged in this way continue to suggest unidimensionality, and IRT results demonstrate that these items successfully cover lower levels of the latent trait, such findings would provide clarification about the relationship between physical aggression, psychological aggression, and injury.

References

Archer, J. (2000). Sex differences in aggression between heterosexual partners: A meta-analytic review. *Psychological Bulletin, 126,* 651–680. doi:10.1037/0033-2909.126.5.651

Barling, J., O'Leary, K. D., Jouriles, E. N., Vivian, D., & MacEwen, K. E. (1987). Factor similarity of the Conflict Tactics Scales across samples, spouses, and sites: Issues and implications. *Journal of Family Violence, 2,* 37–54. Retrieved from http://www.springerlink.com/content/0885–7482. doi:10.1007/BF00976369

Campbell, J. C. (2002). Health consequences of intimate partner violence. *The Lancet, 359,* 1331–1336. doi:10.1016/S0140-6736(02)08336-8

Cascardi, M., Langhinrichsen, J., & Vivian, D. (1992). Marital aggression: Impact injury and health correlates for husbands and wives. *Archives of Internal Medicine, 152,* 1178–1184. doi:10.1001/archinte.1992.00400180048007 doi: 10.1001/archinte.152.6.1178

Cascardi, M., & O'Leary, K. D. (1993). Depressive symptomatology, self-esteem, and self-blame in battered women. *Journal of Family Violence, 7,* 249–259. Retrieved from http://www.springerlink.com/content/0885–7482.

Christensen, A., Sullaway, M., & King, C. E. (1983). Systematic error in behavioral reports of dyadic interaction: Egocentric bias and content effects. *Behavioral Assessment, 5,* 129–140.

Coker, A. L., Smith, P. H., Bethea, L., King, M. R., & McKeown, R. E. (2000). Physical health consequences of physical and psychological intimate partner violence. *Archives of Family Medicine, 9,* 1–7. Retrieved from: http://archfami.ama-assn.org/

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum. Retrieved from: Http://www.taylorandfrancis.com/

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9,* 466–491. doi:10.1037/1082-989X.9.4.466

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics, 5,* 299–314. doi:10.2307/1390807

Jenkins, P., Earle-Richardson, G., Slingerland, D. T., & May, J. (2002). Time dependent memory decay. *American Journal of Industrial Medicine, 41,* 98–101. doi:10.1002/ajim.10035

Jose, A., & O'Leary, K. D. (2009). Prevalence of partner aggression in representative and clinic samples. In K. D. O'Leary & E. Woodin (Eds.), *Psychological and physical aggression in couples: Causes and interventions* (pp. 15–35). Washington, DC: American Psychological Association. Retrieved from http://www.apa.org/pubs/books/4317187.aspx

Kar, H. L., & O'Leary, K. D. (2010). Gender symmetry or asymmetry in intimate partner victimization? Not an either/or answer. *Partner Abuse, 1,* 152–168. doi:10.1891/1946-6560.1.2.152

McDonald, R. P., & Ho, M. H. (2002). Principles and practice in reporting statistical equation analyses. *Psychological Methods, 7* , 64–82. Retrieved from http://www.apa.org/pubs/journals/met/index.aspx

Mock, C., Acheampong, F., Ajdei, S., & Koepsell, T. (1999). The effect of recall on estimation of incidence rates for injury in Ghana. *International Journal of Epidemiology, 28,* 750–755. doi:10.1093/ije/28.4.750

Muthén, B. O., & Muthén, L. (2010). *Mplus version* 6 [Computer Program]. Los Angeles, CA: University of California Los Angeles. Retrieved from www.statmodel.com

O'Leary, K. D., Barling, J., Arias, I., Rosenbaum, A., Malone, J., & Tyree, A. (1989). Prevalence and stability of physical aggression between spouses: A longitudinal analysis. *Journal of Consulting and Clinical Psychology, 57,* 263–268. doi:10.1037/0022-006X.57.2.263

O'Leary, K. D., Slep, A. M., & O'Leary, S. G. (2007). Multivariate models of men's and women's partner aggression. *Journal of Consulting and Clinical Psychology, 75,* 752–764. doi:10.1037/0022-006X.75.5.752

O'Leary, K. D., & Woodin, E. M. (2005). Partner aggression and problem drinking across the lifespan: How much do they decline? *Clinical Psychology Review, 25,* 877–894. doi:10.1016/j.cpr.2005.03.004

Pan, H. S., Neidig, P. H., & O'Leary, K. D. (1994). Male-female and aggressor-victim differences in the factor structure of the Modified Conflict Tactics Scale. *Journal of Interpersonal Violence, 9,* 366–382. doi:10.1177/088626094009003006

Regan, K. V., Bartholomew, K., Kwong, M. J., Trinke, S. J., & Henderson, A. J. Z. (2006). The relative severity of acts of physical violence in heterosexual relationships: An item response theory analysis. *Personal Relationships, 13,* 37–52. Retrieved from: Http://www.wiley.com/bw/journal.asp?ref=1350–4126

Regan, K. V., Bartholomew, K., Oram, D., & Landolt, M. A. (2002). Measuring physical violence in male same-sex relationships: An item response theory analysis of the Conflict Tactics Scales. *Journal of Interpersonal Violence, 17,* 235–252. Retrieved from http://jiv.sagepub .com/

Reise, S. P., & Henson, J. M. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment, 81,* 93–103. doi:10.1207/S15327752JPA8102_01

Schafer, J. (1996). Measuring spousal violence with the Conflict Tactics Scale: Notes on reliability and validity issues. *Journal of Interpersonal Violence, 11,* 572–585. doi:10.1177/088626096011004008

Schafer, J., Caetano, R., & Clark, C. L. (1998). Rates of intimate partner violence in the United States. *American Journal of Public Health, 88,* 1702–1704. Retrieved from http://ajph.aphapublications.org/

Simpson, L. E., & Christensen, A. (2005). Spousal agreement regarding relationship aggression among treatment-seeking couples. *Psychological Assessment, 17,* 423–432. doi:10.1037/1040-3590.17.4.423

Slep, A. M., & O'Leary, S. G. (2007). Multivariate models of mothers' and fathers' aggression toward their children. *Journal of Consulting and Clinical Psychology, 75,* 739–751. doi:10.1037/0022-006X.75.5.739

Steiger, J. H. (1990). Structural model evaluation and modification. *Multivariate Behavioral Research,* 25, 214–212. Retrieved from http://www.tandf.co.uk/journals/titles/0027–3171

Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family, 41,* 75–88. Retrieved from http://www.wiley.com/bw/journal.asp?ref=0022–2445&site=1

Straus, M. A. (1990). The Conflict Tactics Scale and its critics: An evaluation and new data on validity and reliability. In M. A. Straus & R. J. Gelles (Eds.), *Physical violence in American families: Risk factors and adaptations to violence in 8,145 families* (pp. 49–73). New Brunswick, NJ: Transaction Publishing. Retrieved from http://www.transactionpub.com/merchant2/

Straus, M. A., & Gelles, R. J. (1990). How violent are American families? Estimates from the National Family Violence Resurvey and other studies. In M. A. Straus & R. J. Gelles (Eds.), *Physical violence in American families: Risk factors and adaptations to violence in 8,145 families* (pp. 95–112). New Brunswick, NJ: Transaction Publishing. Retrieved from http://www.transactionpub.com/merchant2/

Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised Conflict Tactics Scale (CTS2): Development and preliminary psychometric data. *Journal of Family Issues, 17,* 283–316. doi: 10.1177/019251396017003001

Thissen, D., Chen, W. H., & Bock, R. D. (2003). *Multilog (version 7.03)* [Computer Program]. Lincolnwood, IL: Scientific Software International. Retrieved from http://www.ssicentral.com/irt/index.html

Tjaden, P., & Thoennes, N. (2000). Prevalence and consequences of male-to-female and female-to-male intimate partner violence as measured by the National Violence Against Women Survey. *Violence Against Women, 6,* 142–161. Retrieved from http://vaw.sagepub.com/