

Previous research has not established whether standard violence measures are appropriate for assessing violence in same-sex relationships. This study, therefore, evaluated the structure of an expanded version of the physical violence scale of the Conflict Tactics Scales (CTS) in a sample of randomly selected gay and bisexual men. In total, 284 gay and bisexual men reported on perpetration and receipt of 14 violent acts in their same-sex relationships, both at any time in the past and during the past 12 months. Item response theory analyses indicated that the violence items were unidimensional for receipt ever in the past and for perpetration both ever in the past and in the past 12 months. The items were dispersed along the underlying violence continuum, from lower to higher severity of violence, and discriminated well in this range. We discuss implications for scoring the CTS.

Measuring Physical Violence in Male Same-Sex Relationships

An Item Response Theory Analysis of the Conflict Tactics Scales

**KATHERINE V. REGAN
KIM BARTHOLOMEW
DOUG ORAM**

Simon Fraser University

MONICA A. LANDOLT
University of British Columbia

A growing body of empirical literature suggests that rates of violence in male same-sex relationships are at least as high as those in opposite-sex relationships (e.g., Kelly & Warshafsky, 1987; Landolt & Dutton, 1997). Unfortunately, no studies have systematically examined the measurement of violence in same-sex relationships. Standard measures of violence, such as the Conflict Tactics Scales (CTS) (Straus, 1979), have been developed on heterosexual samples, and it is not clear whether it is appropriate to apply them to same-

Authors' Note: This project was supported by a grant to the second author from the Wayne F. Placek Fund of the American Psychological Foundation. We would like to thank Michael Maraun and Ray Koopman for their statistical advice and Michael Maraun, Stephen Hart, and two anonymous reviewers for their feedback on previous drafts of this article. Correspondence concerning this article should be addressed to Katherine Regan or Kim Bartholomew, Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6. Electronic mail may be sent to zach@sfu.ca or bartholo@sfu.ca.

JOURNAL OF INTERPERSONAL VIOLENCE, Vol. 17 No. 3, March 2002 235-252
© 2002 Sage Publications

sex relationships. Consequently, the goal of the present study was to assess whether an expanded version of the physical assault scale of the CTS (CTS2) (Straus, Hamby, Boney-McCoy, & Sugarman, 1996), the most commonly used measure of partner violence, is appropriate for measuring violence in male same-sex relationships. The CTS was administered to a randomly selected sample of gay and bisexual men, and item response theory (IRT) models were used to assess how well the items of the CTS measure physical violence in male same-sex relationships.

Violence in Male Same-Sex Relationships

Research on violence in male same-sex relationships conducted to date has typically been qualitative in nature, involving in-depth interviews (e.g., Cruz & Firestone, 1998), case studies (e.g., Island & Letellier, 1991), and descriptive studies of small samples (e.g., Farley, 1996; Merrill & Wolfe, 2000). Of the few studies to employ survey methodology to examine violence in male same-sex relationships, most have used variants of the CTS to assess physical violence (Kelly & Warshafsky, 1987; Landolt & Dutton, 1997; Waterman, Dawson, & Bologna, 1989). These studies have, however, relied on convenience methods of recruiting (e.g., newspaper advertisements), and none have assessed the structure of the CTS. Therefore, it is important to determine whether measures designed to assess violence between heterosexual partners are also appropriate for assessing violence between same-sex partners.

The Conflict Tactics Scales

The most commonly used measure of physical violence is the CTS. Questions about violence are presented to respondents as potential ways that couples resolve differences and conflict. The CTS can be used to study perpetration or receipt of various acts, or both. Although the CTS2 (the newest version of the CTS) contains a number of subscales (e.g., a psychological aggression scale), in this study, we focused exclusively on physical violence items.

Straus's (1979) initial analyses of the CTS indicated a single physical violence factor. When the more serious violent acts of using a knife or gun were added to the scale, however, these formed a factor separate from the less serious acts (Straus, 1979). Straus (1990) therefore argued that the CTS physical violence items fall into "minor" and "severe" categories based on their presumed risk of injury (see Table 1). This division has been criticized for ignoring the intent behind an act and for assuming that severe items entail a greater

TABLE 1: Physical Violence Items

<i>Physical Violence Item</i>	<i>Corresponding Conflict Tactics Scales-2 (CTS2) Item</i>	<i>Severity Classification</i>
Pushed or shoved	Pushed or shoved	Minor
Slapped	Slapped	Minor
Threw something that could hurt	Threw something that could hurt	Minor
Twisted arm or hair	Twisted arm or hair	Minor
†Grabbed or held down in anger	Grabbed	Minor
†Scratched or bit	(Not in CTS2)	Minor
†Hit with something that could hurt	Punched or hit with something that could hurt	Severe
†Punched	Punched or hit with something that could hurt	Severe
Slammed against a wall	Slammed against a wall	Severe
Kicked	Kicked	Severe
Beat up	Beat up	Severe
Choked	Choked	Severe
Used a knife or gun	Used a knife or gun	Severe
Burned or scalded on purpose	Burned or scalded on purpose	Severe

† Items that have been changed or added in the current study.

risk of injury than do minor items (e.g., Dobash, Dobash, Wilson, & Daly, 1992; see also Straus, 1990). For example, a man's shoving his partner (a minor item) could result in greater harm than a man's punching his partner (a severe item).

If any measurement instrument is to be widely used in research, how well it measures the construct of interest needs to be established. One way to understand a scale is to examine its structure. This can be done in several ways, including assessing the dimensionality of a scale using either linear factor analysis (e.g., Moffitt et al., 1997) or nonlinear factor analysis, that is, IRT (e.g., Schafer, 1996). Although the IRT model is mathematically equivalent to the linear factor analytic model, it is more appropriate when dealing with dichotomous items (McDonald, 1981; Thissen, Steinberg, Pyszczynski, & Greenberg, 1983). IRT also allows for the assessment of how individual items are functioning relative to one another.

Studies that have assessed the factor structure of the CTS in heterosexual samples using linear factor analysis have generally found that a single factor underlies the physical violence items, rather than minor and severe factors, as proposed by Straus (1979). Unidimensional structures have been confirmed in samples of married couples (Barling, O'Leary, Jouriles, Vivian, & MacEwan,

1987), young adults (Moffitt et al., 1997), incarcerated men (Cook, 1998), and high school students in dating relationships (Cascardi, Avery-Leaf, O'Leary, & Slep, 1999). In contrast, a two-factor structure was found in a sample of about 20,000 U.S. military personnel (Pan, Neidig, & O'Leary, 1994). This study, however, used exploratory rather than confirmatory factor analysis, and it may be that the findings would not stand up to the more stringent test of confirmatory factor analysis.

In sum, linear factor analysis of the CTS suggests that there is one factor underlying the physical violence items, rather than separate factors for minor and severe violence. This does not mean, however, that the physical violence items of the CTS all assess the same severity of physical violence. In addition to testing the unidimensionality of a measure, IRT models show how well the items in a scale are functioning relative to one another and where on the continuum of the construct of violence, from low to high severity, each item lies (e.g., Schafer, 1996).

IRT models are mathematical expressions of the relation between individuals' responses on individual items and an underlying construct that is presumed to give rise to these responses (Crocker & Algina, 1986). An IRT analysis generates a probabilistic rank ordering of the items in a measure along an underlying dimension, in this case severity of violence. An item characteristic curve (ICC) represents how the probability of a response on the item varies with the level of the underlying construct, which is standardized with a mean of 0 and a standard deviation of 1.

Two characteristics of ICCs are important, their slope and their position on the underlying construct. The slope (a), which indicates how well an item distinguishes between individual levels of the construct, is called a discrimination parameter. Larger values of a indicate steeper curves of the ICC. Items with steeper curves are useful for distinguishing between individual levels of the latent variable (i.e., the probability of a yes response, plotted on the y-axis, changes quickly from a low to a high value within a brief range of the construct). The position (b , or location parameter) of the item in relation to the underlying construct indicates at what level of the construct the item differentiates between a yes and a no response. The location parameter is equal to the latent score at which half the respondents answer yes to the item. When two items are located relatively close to one another, this indicates that they are assessing similar levels of the underlying construct. Ideally, a measure has a reasonable dispersion of items.

Only one study that we are aware of has applied an IRT analysis to the CTS. Schafer (1996) used IRT to assess the ordering of the CTS violence items on severity. Using data from a sample of 533 college students reporting on self-to-partner violence, he found that for both men and women, the rank

ordering of the items followed a pattern from generally less severe to more severe violence. For women, most items had fairly steep curves, indicating good differentiation between different levels of violence and showed fairly good dispersion along the continuum of severity of violence. For men, the less severe items had fairly steep curves and were well dispersed, but the more severe items had much flatter slopes and were clustered together at the more violent end of the scale. Overall, the IRT model showed fairly good fit for women, indicating unidimensionality, but not for men.

One other study has used this methodology to examine a different measure of violence in a sample of 250 male prisoners randomly selected from Scotland's largest prison (Michie & Cooke, 2000). Participants were asked whether they had been subjected to particular violent behaviors and if they had engaged in those behaviors themselves since the age of 18. The items in the measure of violence, with the exception of a question about slapping, had fairly steep slopes and discriminated effectively at different points on the underlying violence continuum. The items were not, however, ordered as expected in terms of severity of violence. For example, hitting someone with a fist or beating someone discriminated at relatively low levels of the construct, whereas throwing something at someone fell in the middle range of severity.

Overall, these two studies indicate that IRT analysis can be used to assess the structure of measures of violence. It appears that IRT would be a useful way of providing an assessment of the structure of violence in male same-sex relationships.

Overview of Current Study

We examined the structure of a modified version of the CTS, which was used to assess physical violence in male same-sex relationships, in a randomly chosen sample of gay and bisexual men living in the West End of Vancouver. IRT analyses were used to determine whether the physical violence items form a unidimensional construct and, if so, where and how well they discriminate along the severity of violence continuum. To confirm that this continuum represents severity of violence, we looked at the relationship between receipt of violence and degree of injury.

METHOD

The present study is based on the West End Relationship Project (WERP) conducted in Vancouver, British Columbia, Canada. WERP was designed to

document the experiences of gay and bisexual men in same-sex relationships. WERP involved two phases, a telephone survey and an in-person interview session. The present study focused on the first phase of WERP.

Participants

A total of 1,176 men living in the West End of Vancouver participated in the survey. Of these, 300 gay and bisexual men composed the target sample for this project. About 91% of the participants in the target sample identified themselves as gay, and 9% identified themselves as bisexual. Their age range was between 20 and 79 years, with a mean age of 39 years ($SD = 10.32$). The ethnic breakdown of the target sample was as follows: British (41%), other European (25%), Chinese/East Asian (3%), Canadian (19%), and Other (8%). Some respondents did not identify a specific ethnic background (4%). The distribution of personal income was: less than \$20,000 (18%), \$20,000 to \$29,999 (19%), \$30,000 to \$39,999 (24%), \$40,000 to \$49,999 (15%), and \$50,000 or more (25%). The relationship status of the target sample was as follows: living with a romantic partner (26%), involved in an ongoing romantic or sexual relationship (22%), and had dated or had been in an ongoing relationship in the past 12 months (34%). Sixteen men reported no same-sex relationship experience and were excluded from the present analyses.

Because Canadian Census data does not include sexual orientation, we were unable to assess the representativeness of the target sample. Therefore, we compared the demographic composition of the complete sample of 1,176 men with 1996 Canadian Census data for the West End of the City of Vancouver (Statistics Canada, 1999). The current sample was roughly comparable with the population in terms of ethnicity and age. Our sample, however, overrepresented men with higher incomes (40% earned \$40,000 or more versus 24% in the population) and underrepresented those with lower incomes (22% earned less than \$20,000 versus 46% in the population). Furthermore, it appears that our target sample overrepresented men in their thirties and early forties (59% versus 38% in the population) and underrepresented men in their twenties (17% versus 24% in the population) and men older than 60 years of age (5% versus 14% in the population).

Telephone Survey

Vancouver's West End was chosen as the appropriate site for the telephone survey because relatively high concentrations of gay and bisexual men reside in the area. The sample was drawn from the latest telephone listing for West End exchanges, and a constant was added to the last digit of each listing to

increase the chance of reaching newly listed or unlisted respondents. All households were first screened for men 18 years or older. If more than one adult male resided in the household, a potential respondent was chosen, using the most recent birthday method. Surveyors ensured that respondents were able to complete the survey in private. The overall response rate, calculated as the number of interviews completed divided by the number of known eligible respondents, was 49%. All men were asked about their sexual orientation. Men indicating that they were heterosexual ($n = 876$) were given a brief form of the survey assessing demographics. Gay and bisexual men ($n = 300$) were asked to complete the full version of the survey, which took 15 to 20 minutes.

After demographics were assessed, gay and bisexual respondents were asked about acts of psychological, physical, or sexual aggression that they might have been the perpetrators or recipients of in their same-sex relationships. Subsequent questions assessed injury, experiences of family violence during childhood, and health-related issues.

Measure of physical violence. A 14-item measure (see Table 1) was developed from the CTS (Straus, 1979) and the CTS2 (Straus et al., 1996) to assess participants' reports of receipt and perpetration of physically violent acts. This measure included all the items from the original CTS, using the revised wording of the CTS2 where these items have been changed. One item, *punched or hit with something that could hurt*, was expanded into two items: *punched* and *hit with something that could hurt*. Another item, *grabbed*, was changed to *grabbed or held down*, and an item relating to scratching or biting was added. For each item, respondents were first asked if they had ever engaged in that behavior toward a male partner. If they indicated that they had, they were asked how often they had done so in the past year. Next, they were asked if a male partner had ever directed the same behavior toward them. Again, if they indicated that a partner had done so, they were asked how many times in the past year this had occurred. Participants' responses for the past year were recoded to indicate whether or not the behavior had occurred.

Measure of physical injury. Five items assessed physical injuries that resulted from receipt of physically violent acts. Respondents were asked about both severe injuries, such as broken bones or a concussion, and less severe injuries, such as sprains or bruises. They were also asked whether they had ever seen a doctor because of an injury resulting from conflict, if they had needed to do so but had not, and whether they had felt physical pain that still hurt the next day as a result of conflict with a male partner. The injury items were combined to form an injury index ($\alpha = .82$).

RESULTS

IRT models for the violence items were estimated separately for receipt and perpetration, ever in the past and in the past 12 months. Items were excluded from the analysis when they were not endorsed at all or when only one participant had either been the recipient or perpetrator of that act. Multilog 6.0 (Thissen, 1991) was employed to fit the two-parameter logistic model to the item responses. To assess the goodness-of-fit of the model, we examined the standardized residuals, which were computed for each response option of each item. Thus, for the 14 items, there were 28 standardized residuals. If the model is correct, then these statistics will have an approximately standard normal distribution with roughly 95% being less than 1.96 in magnitude (Hambleton & Swaminathan, 1985). The results for all analyses are shown in Table 2 and Figure 1. The numbers in Figure 1 refer to the items listed in Table 1.

Receipt Ever in the Past

The two-parameter logistic model showed good fit for receipt of violence ever in the past, with all 28 standardized residuals less than 1.96 in magnitude.

Table 2 and Figure 1 show that the 14 items discriminated well on the underlying severity of violence continuum. The items were generally ordered according to their presumed level of severity. There were, however, several exceptions to the expected ordering of items. *Punched* and *slammed against a wall*, considered severe items (see Table 1), both had low *b* parameters, indicating that they discriminated best at lower levels of the continuum. In addition, *scratched or bit* and *twisted arm or hair*, considered minor items, had fairly high *b* parameters, indicating that they discriminated best at high levels of the continuum.

Although covering a fairly wide range on the underlying continuum, the violence items near the middle of the distribution tended to cluster together. *Pushed or shoved* and *punched* were fairly separate from the rest of the items at the lower end of the continuum, as was *burned or scalded* at the upper end of the continuum. The rest of the items tended to form several smaller clusters in between the extremes.

Receipt During the Past 12 Months

Burned or scalded was excluded from this analysis because only one individual reported being the recipient of this act during this time. The two-

TABLE 2: Estimates of Item Response Theory Parameters

<i>Physical Violence Item</i>	<i>Receipt Ever</i>		<i>Perpetration Ever</i>		<i>Perpetration Past 12 months</i>		
	a	(SE) b	(SE) a	(SE) b	(SE) a	(SE) b	
Pushed or shoved	3.25	(1.00) 0.40	(0.08) 2.44	(0.38) 0.74	(0.10) 6.22	(1.37) 1.12	(0.11)
Slapped	2.08	(0.39) 1.21	(0.16) 2.31	(0.48) 1.50	(0.18) 3.42	(1.84) 2.01	(0.36)
Threw something that could hurt	2.54	(0.45) 1.32	(0.14) 1.81	(0.59) 2.21	(0.44) 2.62	(1.17) 2.21	(0.57)
Twisted arm or hair	2.45	(0.59) 1.59	(0.19) 2.51	(0.85) 2.11	(0.29) 2.62	(1.57) 2.21	(0.40)
Grabbed or held down in anger	1.97	(0.47) 1.35	(0.19) 2.78	(0.58) 1.29	(0.13) 3.69	(1.15) 1.57	(0.16)
Scratched or bit	2.30	(0.67) 1.78	(0.20) 1.90	(0.52) 2.31	(0.42) 2.75	(1.29) 2.26	(0.65)
Hit with something that could hurt	2.83	(0.80) 1.62	(0.20) 2.60	(1.05) 2.38	(0.39) 9.63	(—) 2.17	(4.71)
Punched	3.14	(0.70) 0.86	(0.11) 2.71	(0.54) 1.36	(0.14) 6.07	(—) 1.73	(3.18)
Slammed against a wall	2.74	(0.57) 1.18	(0.14) 3.24	(0.67) 1.35	(0.12) 4.27	(1.91) 1.66	(0.17)
Kicked	2.47	(0.80) 1.81	(0.22) 2.65	(0.83) 2.07	(0.30) 3.26	(2.25) 2.35	(0.48)
Beat up	4.49	(1.46) 1.51	(0.11) 2.31	(0.86) 2.28	(0.36) 3.58	(2.63) 2.45	(0.47)
Choked	2.55	(0.71) 1.87	(0.26) 2.82	(0.96) 2.03	(0.26) 2.40	(2.29) 2.50	(0.82)
Used a knife or gun	2.08	(0.71) 2.10	(0.29)				
Burned or scalded on purpose	1.57	(0.64) 3.23	(0.86)				

NOTE: Dashes indicate the standard error was not estimated.

parameter logistic model was not a good fit for the data for receipt of violence in the past 12 months. Only 2 of 26 standardized residuals were less than 1.96 in magnitude, indicating that this model was not a good representation of the data. Because the model did not fit the data, the parameter estimates are not reported here.

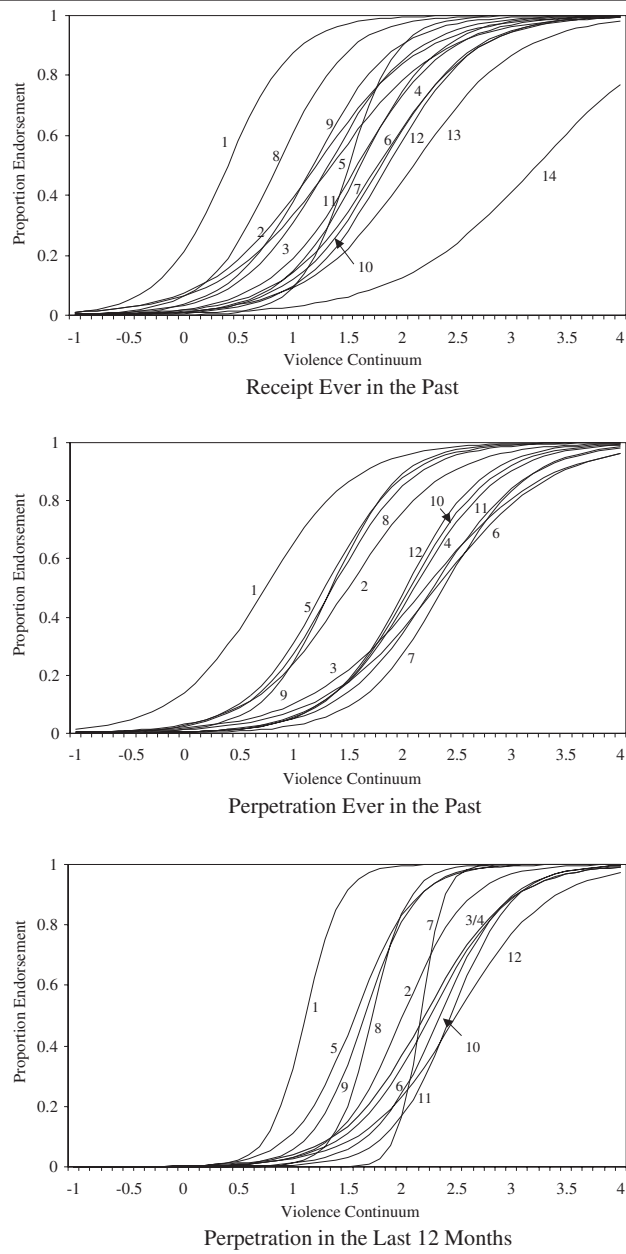


Figure 1: Item Characteristic Curves for Receipt and Perpetration of Violence

Perpetration Ever in the Past

Burned or scalded was excluded from this analysis because no respondents reported ever perpetrating this act. Also, *used a knife or gun* was excluded from this analysis because only one respondent reported ever perpetrating this act. The two-parameter logistic model fit the data for perpetration of violence ever in the past very well, as evidenced by the fact that all 24 standardized residuals were less than 1.96 in magnitude.

Table 2 and Figure 1 show that the 12 items discriminated well on the underlying severity of violence continuum. In this analysis, there was greater overlap in the items with regard to their presumed level of severity than there was in the analysis for receipt ever in the past. As before, *slammed against a wall* and *punched* had lower *b* parameters than expected. In addition, *scratched or bit* and *twisted arm or hair* had higher *b* parameters than expected. The primary difference in this analysis was that *choked* and *kicked*, two severe items, were located lower on the underlying continuum than were other, more minor items, whereas for receipt ever in the past, they were located at higher values of the continuum. As before, the items formed several distinct clusters. *Pushed or shoved* was, again, fairly separate from the rest of the items at the lower end of the continuum.

Perpetration During the Past 12 Months

Burned or scalded was again excluded from this analysis because of no endorsement, as was *used a knife or gun*. The two-parameter logistic model showed good fit for the data for perpetration of violence in the past 12 months, with all 24 standardized residuals less than 1.96 in magnitude.

Table 2 and Figure 1 show that the items were very discriminating on the underlying severity of violence continuum. Multilog was not able to estimate the standard errors for the *a* parameters for *hit with something* and *punched*, and the standard errors for the corresponding *b* parameters were quite large. These parameter estimates should be interpreted with caution. The items were for the most part ordered according to their expected level of severity. There were, again, several exceptions to the expected ordering of items. As before, *slammed against a wall* and *punched*, both severe items, had low *b* parameters. In addition, *scratched or bit*, *twisted arm or hair*, and *thrown something*, all minor items, had fairly high *b* parameters. For this analysis, although the items did not span a large region on the underlying continuum, they were more evenly spaced out over that range. Nonetheless, two items had the same value of *b*: *thrown something* and *twisted arm or hair* ($b = 2.21$).

Summary of IRT Findings

In terms of discrimination, the slopes of the items in all three analyses reported here (excluding receipt in the past 12 months) were fairly steep, indicating that they discriminate well between lower and higher severity of violence. Also, the values of b were generally well dispersed on the severity of violence continuum.

For receipt ever in the past and perpetration both ever in the past and in the past 12 months, *pushed or shoved* had the lowest b parameter, indicating that it was the item associated with the lowest level of severity of violence. For the receipt analysis, the ordering of the items was generally as expected, with minor items located at lower levels of the continuum and severe items located at higher levels of the continuum. For perpetration, the items were not ordered as neatly, with minor and severe items located over all levels of the continuum. Consistently, across all three analyses, *slammed against a wall* and *punched*, both severe items, were located lower on the severity of violence continuum than expected. Furthermore, *scratched or bit* and *twisted arm or hair*, both minor items, were located higher on the severity of violence continuum than expected.

Composite Violence Score

Because the model fit the data well for the three analyses discussed above, indicating that the items were unidimensional, it was appropriate to form a composite of the items (Thissen et al., 1983). When dealing with a unidimensional construct, the simplest method of scoring a measure is to add up the number of items that are endorsed by any given individual to form a total score (Thissen et al., 1983). In this case, we added the number of different acts that participants reported either receiving or perpetrating, with each act receiving equal weight. The model estimated reliability (i.e., precision of measurement) of these composites was high (receipt ever in the past, $r^2 = .95$; perpetration ever in the past, $r^2 = .93$; perpetration in the past 12 months, $r^2 = .95$). This type of scoring has been used in the past with the physical violence items of the CTS, and the total scores have been referred to as variety scores (e.g., Moffitt et al., 1997). Variety scores are reported to be highly correlated with total frequency scores and have the advantage that their distributions are less skewed than distributions of frequencies (Moffitt et al., 1997). In our sample, the variety scores for perpetration in the past 12 months were highly correlated with total frequencies for perpetration in the past 12 months ($r = .70$). Variety scores also avoid a major problem with frequency scores of violent acts. That is, when items are summed into a frequency score, more

weight is given to acts that are less serious but more common, whereas less weight is given to acts that are serious but infrequent. For example, if someone pushed a partner several times, he would obtain a higher violence frequency score than someone who choked his partner twice.

To confirm that the variety scores did in fact represent the underlying construct of severity of violence, we looked at the relationship between each participant's variety score and his Multilog-estimated score on theta, the underlying dimension of severity of violence. The relationship was linear and very strong for all three analyses: receipt ever in the past ($r = .95$), perpetration ever in the past ($r = .95$), and perpetration in the past 12 months ($r = .91$).

We also looked at how well the physical violence items were functioning as a measure of severity of violence by examining the relationship between receipt of violence ever in the past and degree of injury. To do this, we looked at the relationship between the injury index and individuals' variety scores and Multilog-estimated theta scores for respondents who had been the recipients of physically violent acts. The relationship between injury and theta scores was linear and strong ($r = .71$), as was the relationship between injury and variety scores ($r = .71$).

DISCUSSION

This study examined the structure of the physical violence items of an expanded version of the CTS in a randomly selected sample of gay and bisexual men. IRT analyses indicated that the physical violence items form a unidimensional construct for receipt ever in the past as well as for perpetration ever in the past and in the past 12 months. The one-factor structure does not, however, contradict the idea that the items represent different levels of severity of violence, as was demonstrated by the item-specific results. In fact, the strong associations between injury and individuals' theta and variety scores indicate that the items form a measure of severity of violence.

Although the two-parameter logistic model fit the data well for three of the analyses, it did not fit for receipt in the past 12 months. Apparently, the relatively simple model we used to represent this data was not adequate to explain the complex interaction between the item characteristics and the characteristics of the sample. It is likely that some other model might represent the data better, but it is unclear what this other model might be. We conducted follow-up analyses using linear factor analysis, which indicated the presence of a dominant primary factor. Thus, it would appear that the items are also unidimensional for receipt in the past 12 months.

In terms of individual item characteristics, some consistent findings emerged for the three analyses (receipt ever in the past, perpetration ever in the past, and perpetration in the past 12 months). Across analyses, the slopes of the items were large and the distribution of the item locations was reasonable. The results of the IRT analyses indicate, however, that the suggested ordering of the physical violence items of the CTS may not be appropriate when examining violence in male same-sex relationships. Although more minor items were generally located at lower levels of the continuum than more severe items, there were several inconsistencies in the distinction between minor and severe items. There were also consistencies between the current and previous IRT analyses of violence measures. Notably, the placement of *pushed and shoved* at the low end of the continuum in all three analyses is consistent with expectations and findings of previous studies (Michie & Cooke, 2000; Schafer, 1996).

Although *punched* was expected to be a severe item, it tended to be intermixed with minor items at the lower end of the continuum. In their study of male prisoners' violence (including violence between men), Michie and Cooke (2000) also found that an item relating to hitting with a fist or object or beating someone was located lower than expected on the severity of violence continuum. The similarity in terms of this finding could be because both their study and the present study looked at violence between men. In contrast, Schafer (1996) looked at violence between men and women and found hitting to be higher on the severity of violence continuum. It may be that when two men engage in a violent incident, punching is one of the first acts that occurs, with less serious consequences. Conversely, when a man and a woman engage in a violent incident, punching may tend to be a more serious act with more serious consequences, especially when done by a man toward a woman.

Another item that was consistently located lower than expected along the continuum of severity of violence was *slammed against a wall*. Perhaps this item is similar enough to *pushed and shoved* that it may not seem particularly severe. In contrast, two items were consistently located higher than expected along the severity of violence continuum: *twisted arm or hair* and *scratched*. These acts require that partners be in close proximity to one another. Perhaps, men are only likely to use these acts when they are engaging in a fairly serious violent incident in which they are closely involved with one another. Overall, the results suggest that the severity of various violent acts may differ depending on between whom the violence is happening. Extrapolating from the current findings, we expect that the relative severity of violent acts may also differ in female same-sex relationships. Therefore, assessing the CTS in lesbian relationships should be a priority for future research.

Several characteristics of the sample limit its generalizability. First, respondents were gay and bisexual men living in a large urban center, and these findings cannot be generalized to gay and bisexual men living in rural areas. A related issue is that the sample was drawn from an area of the city that is known to have a high proportion of gay and bisexual residents. Thus, the men in this sample were more likely to be “out” in terms of their sexual orientation than men who live in other areas of the city or in rural areas. In addition, our sample was predominantly of British and European origin, although this ethnic breakdown is comparable to that of the West End as a whole. Compared to the census data, our sample did underrepresent lower income men and both younger and older gay men. Unfortunately, we have no way of assessing whether our sample is representative of gay men living in the West End of Vancouver or whether our demographic breakdown may have resulted from selection biases.

In this study, we looked at the validity of an expanded version of the CTS as a measure of physical violence in male same-sex relationships by assessing the structure of the measure. We also confirmed that severity of violence as assessed by this measure is strongly related to injury level. There are, however, other ways to assess validity that should be pursued in future research. For example, Straus (1990) suggests examining interpartner agreement about levels of violence in a relationship to assess concurrent validity of the CTS (for an example of this approach, see Moffitt et al., 1997).

Our study focused exclusively on physical violence. Other forms of abuse, notably psychological and sexual abuse, are important in their own right, and their role in male same-sex relationships needs to be examined. For example, research with heterosexual samples suggests that psychological and physical violence are highly correlated (e.g., Moffitt et al., 1997) and that psychological abuse may have as harmful an effect as physical violence (e.g., Follingstad, Rutledge, Berg, Hause, & Polek, 1990). It is important to keep in mind, however, that within the criminal justice system, the physical violence acts assessed in this study would constitute assault.

The results of this study have implications for scoring the physical violence subscale of the CTS. The present study and previous research indicate that the physical violence items form a unidimensional construct. Thus, it is more appropriate to speak of physical violence overall, rather than to discuss minor and severe physical violence as separate constructs. The results of the IRT analyses further indicated that, when looking at male same-sex relationships, minor violence items were not necessarily located lower on the severity of violence continuum than were severe violence items. This suggests that using the traditional approach to defining minor and severe physical violence may misrepresent what is actually happening in these relationships. Given

that the physical violence items of the CTS are unidimensional, it is appropriate to use a scoring method that adds together the number of different acts endorsed to form a total (variety) score. The fact that the variety scores were strongly related to individuals' theta scores and to the likelihood of injury for receipt of violence provides support for the use of variety scores as a method of scoring the CTS.

This study investigated whether the physical violence items of an expanded version of the CTS are appropriate for use with male same-sex relationships. IRT analyses indicated that these items represent a single latent dimension. The analyses also indicated that the items fall on a continuum of severity of violence. Future research needs to address why the ordering of the items along this continuum appears to be different for male same-sex relationships than for heterosexual relationships. The present study does indicate, however, that the use of the physical violence items of the CTS to assess the severity of violence in male same-sex relationships is reasonable.

REFERENCES

- Barling, J., O'Leary, K. D., Jouriles, E. N., Vivian, D., & MacEwan, 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.
- Cascardi, M., Avery-Leaf, S., O'Leary, K. D., & Slep, A.M.S. (1999). Factor structure and convergent validity of the Conflict Tactics Scale in high school students. *Psychological Assessment, 11*, 546-555.
- Cook, S. L. (1998). Assessing threats to the valid measurement of violence against women constructs. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 59*, 0909.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt Brace Jovanovich College.
- Cruz, J. M., & Firestone, J. M. (1998). Exploring violence and abuse in gay male relationships. *Violence and Victims, 13*, 159-173.
- Dobash, R. P., Dobash, R. E., Wilson, M., & Daly, M. (1992). The myth of sexual symmetry in marital violence. *Social Problems, 39*, 71-91.
- Farley, N. (1996). A survey of factors contributing to gay and lesbian domestic violence. In C. M. Renzetti & C. H. Miley (Eds.), *Violence in gay and lesbian domestic partnerships* (pp. 35-42). Binghamton, NY: Haworth.
- Follingstad, D. R., Rutledge, L. L., Berg, B. J., Hause, E. S., & Polek, D. S. (1990). The role of emotional abuse in physically abusive relationships. *Journal of Family Violence, 5*, 107-120.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston: Kluwer Academic.
- Island, D., & Letellier, P. (1991). *Men who beat the men who love them: Battered gay men and domestic violence*. Binghamton, NY: Harrington Park Press.

- Kelly, C. E., & Warshafsky, L. (1987, July). *Partner abuse in gay male and lesbian couples*. Paper presented at the Third National Conference for Family Violence Researchers, Durham, NH.
- Landolt, M. A., & Dutton, D. G. (1997). Power and personality: An analysis of gay male intimate abuse. *Sex Roles, 37*, 335-359.
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology, 34*, 100-117.
- Merrill, G. S., & Wolfe, V. A. (2000). Battered gay men: An exploration of abuse, help seeking, and why they stay. *Journal of Homosexuality, 39*(2), 1-30.
- Michie, C., & Cooke, D. J. (2000). *The structure of violence behavior: A hierarchical model of the MacArthur Community Violence Screening Instrument*. Unpublished manuscript, Glasgow Caledonian University, Scotland.
- Moffitt, T. E., Caspi, A., Krueger, R. F., Magdol, L., Margolin, G., Silva, P. A., & Sydney, R. (1997). Do partners agree about abuse in their relationship? A psychometric evaluation of interpartner agreement. *Psychological Assessment, 9*, 47-56.
- 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.
- Schafer, J. (1996). Measuring spousal violence with the Conflict Tactics Scale. *Journal of Interpersonal Violence, 11*, 572-585.
- Statistics Canada. (1999). *Profile of census tracts (1996 Census of Canada, catalogue number 95-213-XPB)*. Ottawa: Industry Canada.
- Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CT) Scales. *Journal of Marriage and the Family, 41*, 75-88.
- Straus, M. A. (1990). The Conflict Tactics Scales 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.
- Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised Conflict Tactics Scales (CTS2): Development and preliminary psychometric data. *Journal of Family Issues, 17*, 283-316.
- Thissen, D. (1991). *Multilog 6.0 user's guide*. Chicago: Scientific Software.
- Thissen, D., Steinberg, L., Pyszczynski, T., & Greenberg, J. (1983). An item response theory for personality and attitude scales: Item analysis using restricted factor analysis. *Applied Psychological Measurement, 7*, 211-226.
- Waterman, C. K., Dawson, L. J., & Bologna, M. J. (1989). Sexual coercion in gay male and lesbian relationships: Predictors and implications for support services. *Journal of Sex Research, 26*, 118-124.

Katherine V. Regan is a student in the Department of Psychology at Simon Fraser University (in Vancouver, British Columbia, Canada). Her research interests relate to gender and relationship violence. Currently, she works as a program development and evaluation consultant for the Government of British Columbia.

Kim Bartholomew received her doctorate in personality psychology from Stanford University in 1989 and subsequently joined the Department of Psychology at Simon Fraser

University. Her research has focused on attachment processes in adult intimate relationships, relationship abuse, and same-sex relationships.

Doug Oram received an M.A. in counseling psychology from Antioch University (in Washington State) in 1995. He currently works as a counselor and is a member of a research team investigating the relationship experiences of gay men.

Monica A. Landolt received her doctorate in social psychology from the University of British Columbia in 1999. Her research interests involve health psychology. Currently, she works as a research consultant for the British Columbia Transplant Society.