Assessing the physical violence component of the *Revised Conflict Tactics Scales* when used in heterosexual couples: an item response theory analysis

Avaliando o componente de violência física da *Revised Conflict Tactics Scales* quando usado em casais heterossexuais: uma análise via teoria de resposta ao item

Michael E. Reichenheim [1]
Ruben Klein [2]
Claudia Leite Moraes [1]

Abstract

[1] Instituto de Medicina
Social, Universidade do
Estado do Rio de Janeiro,
Rio de Janeiro, Brasil.
[2] Fundação Cesgranrio,
Rio de Janeiro, Brasil.

Correspondence
M. E. Reichenheim
Instituto de Medicina Social,
Universidade do Estado do
Rio de Janeiro.
Rua São Francisco Xavier 524,
7º andar, Rio de Janeiro, RJ
20559-900, Brasil.
michael@ims.uerj.br

Although there are psychometric evaluations of the Revised Conflict Tactics Scales (CTS2) when applied to heterosexual relationships, none has used item response theory (IRT). To address this gap, the present paper assesses the instrument's physical violence subscale. The CTS2 was applied to 764 women who also responded for their partners. Single dimensionality assumption was corroborated. A 2-parameter logistic IRT model was used for estimating location and discriminating power of each item. Differential item functioning and item information pattern along the violence continuum were assessed. Gender differences were detected in 3 out of 12 items. Item coverage of the latent trait spectrum indicated little information at the lower ends, while plenty in the middle and upper ranges. Still, depending on gender, some item overlaps and regions with gaps could be detected. Despite some unresolved problems, the analysis shows that the items form a theoretically coherent information set across the continuum. Provided the user is aware of possible drawbacks, using the physical violence subscale of the CTS2 in heterosexual couples is still a sensible option.

Domestic Violence; Questionnaires; Reliability and Validity

Introduction

The increase in family violence has generated a growing body of research on the development of measurement tools, including the class of instruments known as *Conflict Tactics Scales* [1,2,3,4,5,6,7]. These are based on 'conflict theory' and aim to identify strategies used by individuals to solve disagreements and impasses [8,9]. The first instrument in the series (CTS1) was proposed in the late 1970s and sought to address any kind of violent relationship [10]. Despite its encouraging evaluations over the years and successful use in at least 20 countries [3,11,12,13,14,15], a more refined instrument was subsequently developed for dealing exclusively with intimate partner violence. The *Revised Conflict Tactics Scales* (CTS2) was presented in 1996 and consists of 39 items forming five scales, namely negotiation (6 items), psychological aggression (8 items), physical violence (12 items), sexual coercion (7 items), and injury (6 items) [13].

The CTS2 has been studied since its release. Results show acceptable reliability, validity, and a factor structure in tune with the underlying theory [13,16,17,18,19]. However, when focusing on intimate partner violence in heterosexual couples, such psychometric evaluations have exclusively used classical test theory [20]. In order to address this gap, the aim of the present paper is to provide a more in-depth item analysis than is possible with traditional psychometric procedures.

Because item response theory techniques offer more insight than those of classical test theory, more refined issues may be addressed such as examining the properties of individual items in order improve the scale's efficiency and reliability in mapping the full continuum of intimate partner violence [20,21]. Although there are two other studies about the Conflict Tactics Scales using item response theory, one evaluates an adapted edition of the CTS2 to measure physical violence in the context of male same-sex relationships [22], while the other relates exclusively to the CTS1 [23]. Moreover, little is known about the relative contribution and efficiency of the CTS2 constituent items in capturing the range of violence severity according to specific subgroups of interest. This issue also deserves attention since it is desirable for a scale to be group-independent in order to have breadth and general expediency. From an empirical perspective, insofar as possible the role played by each component item should be invariant with respect to age, gender, and so on.

As a first development of this more detailed scrutiny of the CTS2 applied to heterosexual couples, the present article focuses solely on the physical violence subscale. Three main issues are examined: (1) gender differences, assessing whether the items display similar properties regardless of the perpetrator (the woman versus her male partner); (2) the ability of items to differentiate between individuals/couples possessing different levels of the violence trait, both in terms of their location along the construct continuum and their discriminating capacity; and (3) the extent to which items provide useful and reliable information about intimate partner violence along the violence continuum, which implies searching for possible item gaps or redundancies along the spectrum. An important assumption underlying the appropriateness of the item response theory models used here is also addressed, i.e., to replicate previous studies indicating that the scale on physical violence is effectively measuring a single dimension [24]. However, unlike previous studies, full information factor analysis is employed to deal with the dichotomous nature of test items [25].

Methods

The present study is subsidiary to a hospital-based case-control study exploring the relationship between violence within families of pregnant women and premature childbirth. Further details of the study population and field procedures may be found in Reichenheim & Moraes [26]. The effective sample used in the present analysis includes 764 couples. The physical violence subscale is part of a formally adapted Portuguese version of the CTS2 [27,28]. Items refer to the respondent (woman) and by proxy in relation to her male partner or ex-partner. The recall period covers 6 to 9 months, depending on gestational age. Item contents are presented in the first table of the *Results* section. Complete wordings can be found in Straus et al. [13]. Readers are referred to Moraes et al. [27] for the complete Portuguese version.

The study was formally approved by the Research Ethics Committee of the Rio de Janeiro Municipal Health Department in conformance to the principles embodied in the Declaration of Helsinki. Participation in the study followed free informed consent. Confidentiality of information was guaranteed. All the women received orientation on where to seek help if they felt it was necessary.

Statistical analysis

As a first step, one-factor and two-factor full information factor analyses [25] were conducted to determine the extent to which a single latent dependence factor accounted for a common variance among items. The solution used tetrachoric matrices and was implemented in TESTFACT [29].

Differential item functioning was evaluated to address how the items behaved between genders [24]. The method used here follows Thissen et al. [30]. The procedure uses a 1-parameter logistic (1PL) item response theory model and aims to detect interactions between gender and the item location parameters (b_i). These indicate at what level of the latent variable an item is able to differentiate between a positive and negative response, and equals the latent score (θ_s) at which half the respondents answer positively to the item. The 1PL model predicting the probability of endorsement for person on item is defined as

$$P(X_{is} = 1 \mid \theta_s, b_i) = \frac{1}{1 + \{\exp (b_i - \theta_s)\}} \quad (1)$$

Recognizing that the present analysis takes women's responses as "reference" and regards partners as the "focus" group [31], the procedure involves calculating $d = \overline{b}_{i(p)} - \overline{b}_{i(w)}$, $d = b^*_{i(p)} - b_{i(p)}$ and $\Delta_i = b^*_{i(p)} - b_{i(w)}$, where $\overline{b}_{i(w)}$ and $\overline{b}_{i(p)}$ are the means of all location estimates for women and partners, respectively; $b^*_{i(p)}$ are the individual adjusted estimates regarding the second group; and Δ_i are the differential item functioning estimates obtained by subtracting the effectively estimated parameters for women from the adjusted $b^*_{i(p)}$ values. Standard normal test statistics were used to assess the significance of Δ_i.

Once the items involving differential item functioning were uncovered, a 2-parameter lo-

gistic (2PL) model was considered for the main item response theory analysis. The 2PL model is given by

$$P(X_{is} = 1 \mid \theta_s, b_i) = \frac{1}{1 + \{\exp (a_i (b_i - \theta_s))\}} \quad (2)$$

where b_i and θ_s are defined in equation (1) and a_i represents item discrimination. Larger values of a_i correspond to steeper item characteristic curves (ICC), indicating that an item is discriminating increasingly well between levels of the construct. Items with differential item functioning include separate b_i and a_i estimates for each gender group, while those not involving differential item functioning are common to both. The choice of a 2PL model was based on a formal comparison with the 1PL model using a likelihood ratio test of statistical significance [30]. Generally, the difference between the -2log(likelihood)s associated with each model is given by $G^2_{diff} = G^2_{(1PL)} - G^2_{(2PL)}$, which is distributed as a chi-square with $df = df_{(1PL)} - df_{(2PL)}$. Since in the present data $G^2_{diff} = 5386.070 - 5333.460 = 52.61$ and $df = 14$, $p = 0.0001$.

The root mean square-fit statistic was used to explore item fit [24,32]. This statistic is given by $\text{RMS}(z_i^2) = \sqrt{\sum_{s=1}^{N} (z_{si})^2 / N}$, where z_{si} is the standardized residual defined as $z_{si} = (x_{si} - p_i(\theta_s)) / \sqrt{V[x_{si}]}$ and, in turn, x_{si} is the observed status of subject s on item i (either 0 or 1). The variance of x_{si} is $V[x_{si}] = p_i(\theta_s)(1 - p_i(\theta_s))$, $p_i(\theta_s)$ being the model-derived conditional probability of a subject s endorsing an item i calculated from the parameters estimated in Equation 2. Values of $\text{RMS}(z_i^2)$ outside the range of -2 to +2 may indicate that the overall item fit is questionable.

The third main goal, namely to examine the amount of item information along the violence continuum, was addressed through the test information [24]. In a 2PL model, this is given by $TI(\theta_s) = \sum_{i=1}^{I} I(\theta_s)$, which is a function of the information curve for item i at trait level θ_s, i.e., $I(\theta_s) = a_i^2 P_i(\theta_s)(1 - P_i(\theta_s))$. Note that, since $TI(\theta_s) = 1/(SE(\theta_s))^2$ [24], this is also an indicator of the instrument's precision at θ_s.

The item response theory analysis was carried out in BILOG-MG [31]. Item response curves shown in Figures 1 and 2 were obtained in Stata 8.2 [33].

Results

Table 1 provides the factor structure of the subscale. Irrespective of perpetrator, the full information factor analysis indicates that almost all loadings are above 0.6 in model A and in the first factor of the two-factor B model. Focusing on the woman as perpetrator, nearly two-thirds of the total variance is accounted for by the first factor, whereas the second "explains" less than 7%. An even more striking pattern is found for partners.

The differential item functioning results are in Table 2. Three items present differential functioning (i-1, i-8, and i-9) as judged from the statistical significance of the Δ_i estimates. Note that negative values emerge when the adjusted location estimates for the "focus" group ($b^*_{i(p)}$) are lower than those for the "reference" group ($b^*_{i(w)}$), a situation following a lower endorsement frequency in the latter. Thus, from a psychometric standpoint, a negative value indicates that a given act captures a higher level of the trait (violence) when perpetrated by women. Concentrating on the statistically significant differential item functioning items, this is the case for *beating up* (i-8) and *grabbing the other* (i-9). Conversely, *throwing something that may hurt* (i-1) is less common in men than in women, which entails a higher θ_s when carried out by the former.

Acknowledging three items with differential item functioning, the main item response theory estimates are shown in Table 3. Standard errors are relatively small for both types of estimates, even for item 11 that had relatively few subjects endorsing it and the least satisfactory fit (last column). All but item 11 and item 1 for partners present discriminations (a_i) above 2.0.

The "steepness" of each item may be better examined in Figure 1, which shows the ICC obtained from the fitted estimates. In addition, the curves convey the item coverage of the latent trait spectrum, indicating that there is very little, if any, information below $\theta_s = -1$. This is expected since the mean of θ_s is approximately zero and there are not many positive responses to the items.

A similar picture is expressed by the location estimates presented in Table 3 (second column) extending from $b_3 = 1.060$ to $b_{11} = 3.503$, and more markedly, by the gender specific information functions (curves) shown in Figure 2. Notice that information peaks at $\theta_s = 1.6$ for women and $\theta_s = 1.7$ for partners.

Figure 2 also displays the positioning of items along the violence continuum, where item overlaps and regions with gaps may be best inspected. Irrespective of gender, there is some sparseness at the upper spectrum between i-6 and i-11, which adds to the previously alluded absence of information taking place at the lower ends. Conversely, items 5, 7, and 12 tend to cluster in a very narrow region, precisely where information is most abundant.

Besides those "common" aspects, some features also emerge as a consequence of differential item functioning. Focusing on the three items, b_1 and b_8 are much farther apart than b_9, the first

Table 1

One-factor (Model A) and two-factor (Model B) full information factor analysis. Results are given by gender
(type of perpetrator) and sorted by appearance in the original form.

| | Women | | | Partners | | |
| | Model A | Model B | | Model A | Model B | |
Item *	f1	f1	f2	f1	f1	f2
i-1: Threw something that could hurt (7/8)	0.746	0.745	0.193	0.744	0.743	0.143
i-2: Twisted arm or hair (9/10)	0.619	0.619	0.167	0.846	0.845	0.084
i-3: Pushed or shoved (17/18)	0.819	0.832	0.361	0.863	0.864	0.176
i-4: Used a knife or gun (21/22)	0.725	0.727	–0.182	0.854	0.865	–0.365
i-5: Punched/Hit with something (27/28)	0.796	0.792	0.033	0.946	0.945	0.064
i-6: Choked the other (33/34)	0.890	0.890	–0.173	0.941	0.939	–0.014
i-7: Slammed against the wall (37/38)	0.833	0.829	0.034	0.814	0.817	0.204
i-8: Beat up the other (43/44)	0.826	0.832	0.279	0.940	0.937	0.012
i-9: Grabbed the other (45/46)	0.639	0.637	–0.105	0.809	0.807	–0.035
i-10: Slapped the other (53/54)	0.850	0.846	0.012	0.908	0.905	0.018
i-11: Burned or scalded on purpose (61/62)	0.569	0.604	–0.637	0.637	0.660	–0.502
i-12: Kicked the other (73/74)	0.790	0.789	–0.136	0.893	0.892	0.117
Eigenvalue	7.013	7.068	0.773	8.749	8.782	0.505
Percent of variance	59.9	61.8	6.4	74.1	76.6	4.1

* Ordering follows the appearance in the English edition of the CTS2 in which items on physical violence are randomly placed among those of the other
subscales and items pertaining to respondents and partners as perpetrators receive consecutive numbers [13]. The original numbering system is provided
in brackets.

Table 2

Differential item functioning estimates (Δ_i). Items sorted by appearance in the original form.

Item	Δ_i *	z **	p value
i-1: Threw something that could hurt	0.727 (0.220)	3.305	0.001
i-2: Twisted arm or hair	-0.462 (0.256)	1.805	0.071
i-3: Pushed or shoved	0.146 (0.102)	1.431	0.152
i-4: Used a knife or gun	0.270 (0.383)	0.705	0.481
i-5: Punched/Hit with something	0.012 (0.140)	0.086	0.932
i-6: Choked the other	-0.108 (0.363)	0.297	0.766
i-7: Slammed against the wall	-0.145 (0.163)	0.890	0.374
i-8: Beat up the other	-0.644 (0.226)	2.850	0.004
i-9: Grabbed the other	-0.290 (0.116)	2.500	0.012
i-10: Slapped the other	-0.076 (0.104)	0.731	0.465
i-11: Burned or scalded on purpose	0.614 (0.713)	0.861	0.389
i-12: Kicked the other	-0.042 (0.171)	0.246	0.806

* In brackets: standard errors (s.e. (Δ_i));

** z = |Δ_i / s.e. (Δ_i)|

two estimates clearly occupying different areas of the spectrum according to gender. Further gaps or overlaps are thus especially discernible among women. In contrast to the distribution among partners (men), i-1 and i-9 tend to occupy the same area forming a new cluster along with i-10; i-8 now clutters the already crowded region filled in by i-4 and i-6; while two gaps are left open, one between i-12 and i-2, and another between the latter and i-4.

Discussion

The present study supports a single-factor solution for the physical abuse scale of the CTS2.

Table 3

2-parameter logistic (2PL) model estimates, including separate b_i and a_i for three items presenting gender differentials (DIF). Sorted by increasing intensity of violence (b_i).

Item *	b_i **	a_i **	RMS(z_i^2) ***
i-3: Pushed or shoved	1.060 (0.038)	3.012 (0.182)	0.574
i-9: Grabbed the other (P)	1.133 (0.085)	2.184 (0.193)	0.589
i-9: Grabbed the other (W)	1.325 (0.058)	2.692 (0.246)	0.697
i-1: Threw something that could hurt (W)	1.409 (0.069)	2.360 (0.223)	0.735
i-10: Slapped the other	1.429 (0.039)	3.847 (0.354)	0.471
i-5: Punched/Hit with something	1.686 (0.059)	3.236 (0.310)	0.516
i-7: Slammed against the wall	1.821 (0.066)	3.695 (0.400)	0.426
i-12: Kicked the other	1.857 (0.079)	2.513 (0.224)	0.629
i-8: Beat up partner (P)	2.018 (0.109)	3.246 (0.488)	0.346
i-2: Twisted arm or hair	2.204 (0.119)	2.032 (0.196)	0.710
i-1: Threw something that could hurt (P)	2.366 (0.206)	1.722 (0.224)	0.703
i-4: Used a knife or gun	2.567 (0.161)	2.273 (0.285)	0.576
i-8: Beat up the other (W)	2.592 (0.211)	3.500 (0.797)	0.313
i-6: Choked the other	2.759 (0.190)	2.965 (0.532)	0.445
i-11: Burned or scalded on purpose	3.503 (0.409)	1.968 (0.384)	0.923

* Items with DIF identified according to perpetrator. In brackets: W – women; P – partners (men);

** In brackets: standard errors;

*** Item root mean square fit statistic.

Even if this finding agrees with several previous reports [16,17,18,22,23,28,34,35,36], it has the strength of being based on a full information factor analysis. Note that this result is central to what follows, since a single dimensionality of a scale being scrutinized is essential to render item response theory interpretable [24].

All item slopes were reasonably large, indicating that the items discriminate well between levels of the construct. Most ICC tend to cover about one unit interval of the q_s (violence continuum). This is indicative that the majority of subjects endorsing a given item are quite similar in their underlying latent-trait. The same trend has been found in other studies on the CTS using item response theory [22,23].

This is also true for the mapping of items along the violence continuum [22,23], which, without further qualification, seems to be quite reasonable. Still, a few points require some thought, since this mapping does not happen along the entire range, given that information on violence is confined exclusively to the upper half of the spectrum. As a result, moderately violent subjects may go unnoticed in practice. Failure to detect violence-positive individuals can have public health consequences if incipient cases that could escalate in the future are missed. In addition, from an epidemiological research perspective, misclassifying couples as "negatives" could lead to the attenuation of risk measures if the variable is the target exposure [37,38] or to

residual confounding if included in a model as a confounder [39,40]. Further research could identify items to fill this gap.

Several gaps and item overlaps also occur along the continuum where information is available. Besides, the item distribution pattern is somewhat distorted by differential item functioning, especially for women. While not markedly influencing the information curves (Figure 2), some redundancies are still found. One may thus argue whether certain items might need refinement or replacement or should even be withdrawn altogether. Of course, precisely which items is a matter for debate and future investigations.

Between partners, the effect of differential item functioning is more on the ordering of items than on distribution along the continuum. While there is a fairly even spread, an apparent inconsistency vis-à-vis the theoretically expected item build-up may be spotted, which, incidentally, is preserved among women. Anticipated at the lower end, item 1 (*threw something that could hurt*) now occupies the middle range of the spectrum. The same "misplacement" was found by Regan et al. [22] studying same-sex male partners. One may conjecture that, although representing a more "trivial" act for women, the item has another meaning for men, tending to be performed mostly by those "bearing" higher violence levels in this group. Another idea to explore is that the act is culturally "feminine" in essence and is simply less used by men (whether violent or not).

Figure 1

Common item characteristic curves (ICC) for items without differential item functioning (DIF) (a);
gender related ICC for items with DIF (b-d).
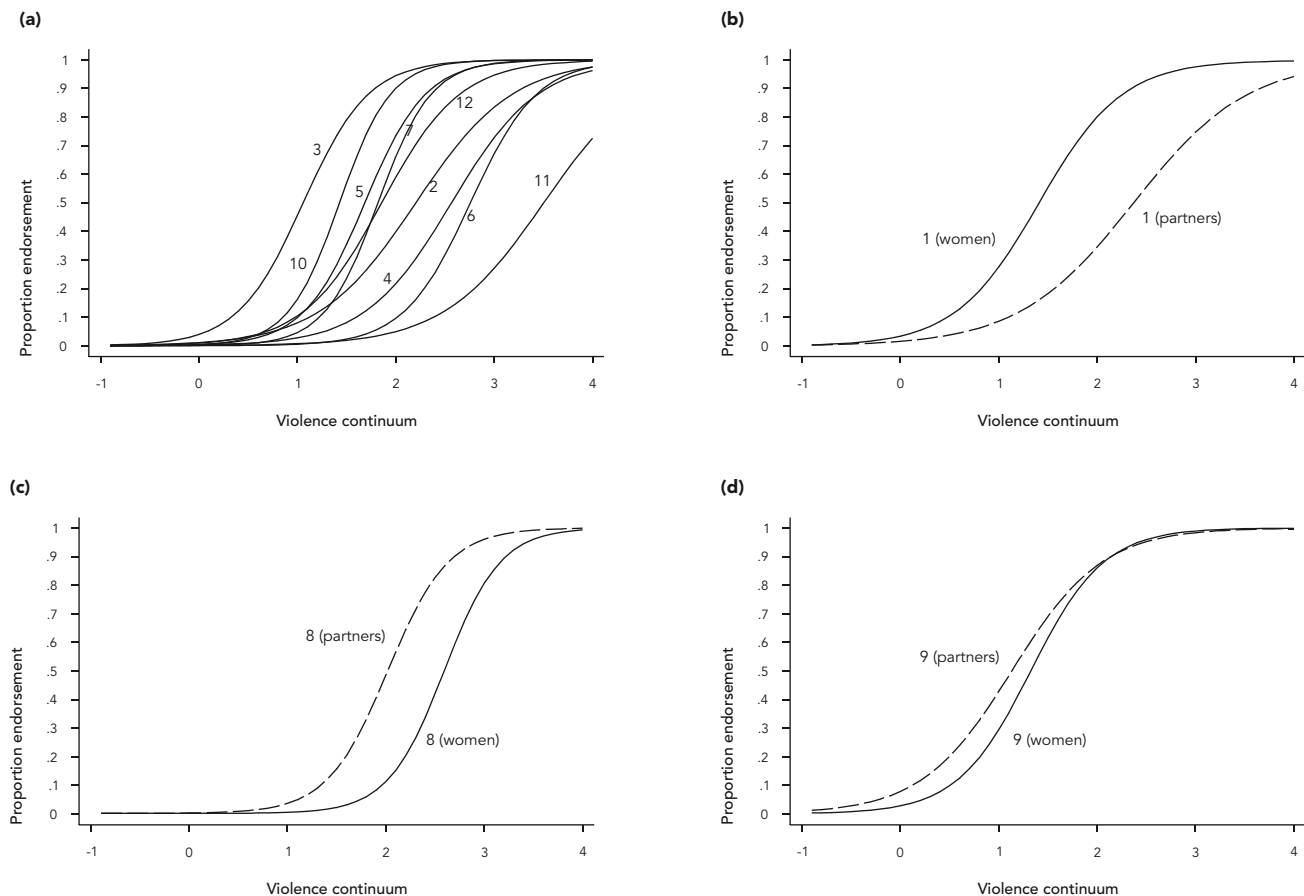
At any rate, according to the present study the component items are definitely not gender-invariant. While acknowledging that the acts allegedly perpetrated by partners were obtained by proxy responses and that spousal consensus on the CTS scales may not always be optimal [41,42,43,44], the very presence of differential item functioning can raise concern. Depending on gender, the same enactment may indicate different levels of violence, whether factual or related to a given item idiosyncrasy. Regardless, the ramification is that an item may not be grasping the same intensity of violence. This is particularly important if the use of raw scores is foreseen in practice because a given count may not actually correspond to identical levels of the construct being tapped. From a psychometric perspective, such a discrepancy detracts from the scale's "universality", which recognizably is a quite desirable (if not essential) property of any measurement tool.

Several issues are worth considering when assessing the present findings. First, one can contend that the results emerge in a different cultural milieu than that of the instrument's original development (USA), thus potentially limiting the generalizations. However, as indicated in the methods section, the CTS2 version employed here underwent a careful cross-cultural adaptation [27,28] based on strict procedural guidelines [45,46]. Even so, given the instrument's widespread use, it would be interesting to have more studies using item response theory models replicated on the scales of the original CTS2 in English, as well as on other versions besides Portuguese.

Figure 2

Positioning of items along the violence continuum and information region by gender (type of perpetrator).
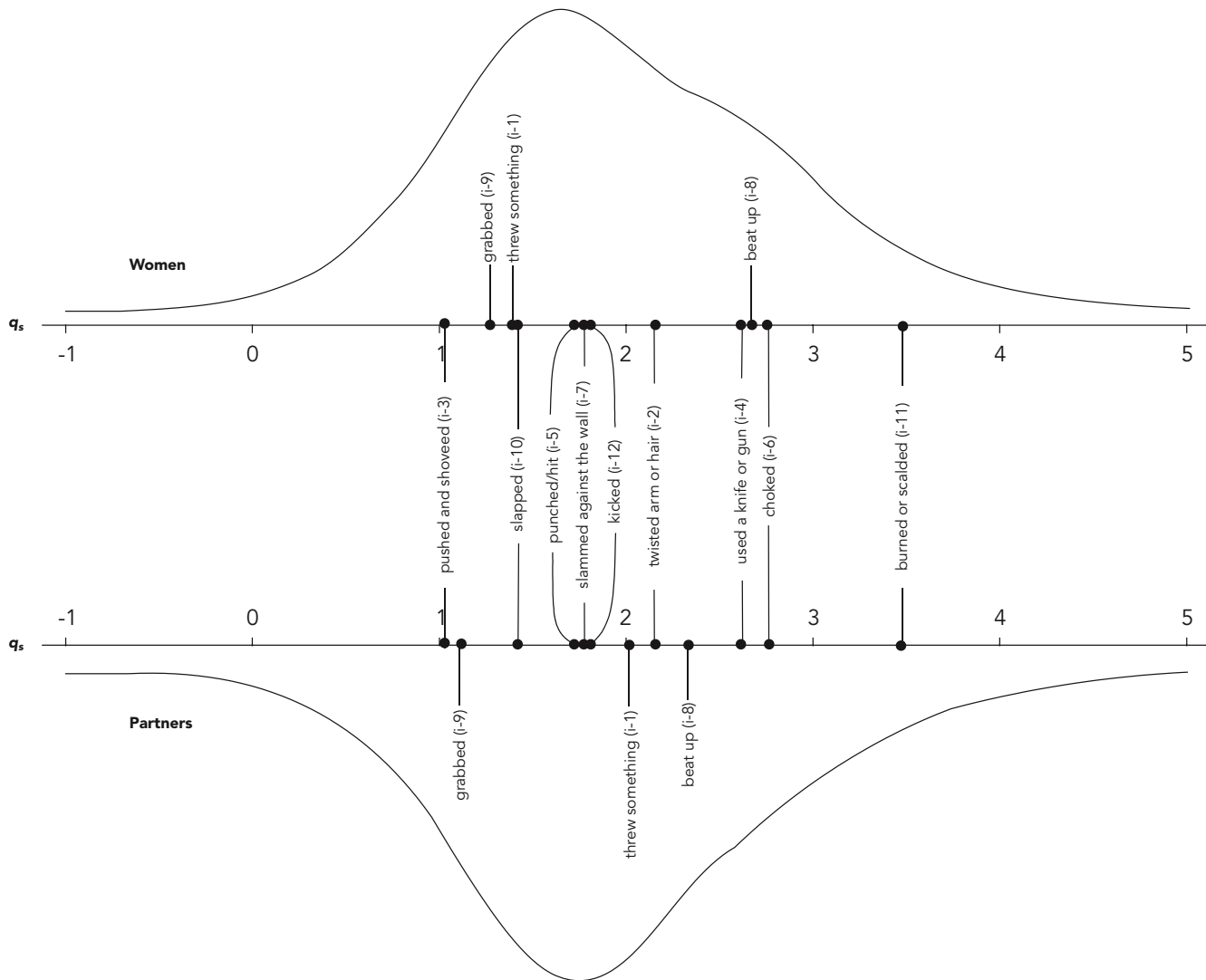
Another important point when generalizing the findings is that the present study was restricted to pregnancy. This not only implies a shorter recall period than the usual 12 months, but may also be portraying an atypical situation. Although this still has to be confirmed, one can envisage certain events (items) happening more often outside of pregnancy, and as a consequence, item response theory results showing better coverage of the instrument along the violence continuum. It should be born in mind that b_i is also a function of the endorsement proportion, the more frequent the item, the lower the respective estimate [24]. Again, the research program on the CTS2 would gain from further studies using item response theory models applied to data from a general heterosexual population, which thus far has surprisingly never been done.

In addition to the suggestions made so far, an interesting development would be to conduct a more refined evaluation of the CTS2 using item responses coded into more than two categories. Noting that each level bears its own discriminating power and position in ordered-response item

response theory models [47], identifying those with good psychometric properties could help reduce the total number of component items and ultimately improve the instrument's operational capability. Readers should be reminded that if used in its current full edition (78 items), application of the CTS2 in one sitting is tiresome for all concerned, even if responses are limited to dichotomous answers.

As emphasized previously, this study is the first in a series aiming to explore the component scales of the CTS2 applied to heterosexual couples using item response theory methods. Thus, further studies are needed to examine the other subscales, particularly psychological aggression. A more in-depth scrutiny of this subscale might shed light on the violence spectrum missed by the physical abuse subscale and effectively serve as a useful complement rather than merely "one more" tool in the early detection of intimate partner violence.

Despite the limitations noted above and some still untapped questions, the item response theory analysis reported here shows that the CTS2 items on physical violence define a one-dimensional trait and form a theoretically coherent information set across the continuum, at least where there is actual coverage. Provided the user is aware of some possible gender discrepancies and that not all empirically negative cases are necessarily void of violence, use of the physical violence CTS2 subscale in heterosexual relationships is still a sensible option.

Resumo

Ainda que existam avaliações psicométricas da Revised Conflict Tactics Scales (CTS2) quando aplicada a casais heterossexuais, nenhuma usou o método de Teoria de Resposta ao Item (TRI). Preenchendo essa lacuna, este estudo examina a subescala de violência física do instrumento. A CTS2 foi aplicada a 764 mulheres que também respondiam pelos seus companheiros. A assunção de unidimensionalidade foi corroborada. Utilizou-se um modelo logístico de 2 parâmetros para estimar a localização e capacidade discriminante dos itens. Avaliou-se o funcionamento diferencial destes, bem como padrão de informação ao longo do contínuo de violência. Detectou-se diferenciais de gênero em três dos 12 itens. A cobertura dos itens relativa ao espectro do traço latente indicou pouca informação nos segmentos inferiores, mas plena no centro e nos superiores. Todavia, dependentes de gênero, algumas superposições e lacunas puderam ser detectadas. A despeito de alguns problemas detectados, mostrou-se que os itens formam um conjunto coerente ao longo do contínuo. Desde que o usuário esteja ciente dos possíveis problemas, usar a subescala de violência física da CTS2 em casais heterossexuais continua sendo uma opção sensata.

Violência Doméstica; Questionários; Confiabilidade e Validade

Contributors

M. E. Reichenheim participated in all phases of the study (fundraising/administration; theoretical development; field work, data processing, and analysis; and drafting the paper as first author). R. Klein was involved in designing the paper's theoretical and statistical framework, as well as collaborating in the analysis and final versions. C. L. Moraes participated in the fundraising/administration, fieldwork, data processing, and drafting of the paper.

Acknowledgements

References

1. Kashani JH, Daniel AE, Dandoy AC, Holcomb WR. Family violence: impact on children. J Am Acad Child Adolesc Psychiatry 1992; 31:181-9.

2. Heise L, Pitanguy J, Germain A. Violence against women: the hidden health burden. Washington DC: The World Bank; 1994.

3. Straus MA, Gelles RJ. Physical violence in American families: risk factors and adaptations to violence in 8145 families. New Brunswick: Transaction Publisher; 1995.

4. Gelles RJ. Intimate violence in families. London: Sage Publications; 1997.

5. Carlson BE. Children exposed to intimate partner violence: research findings and implications for intervention. Trauma Violence Abuse 2000; 1: 321-42.

6. Krug EG, Dahlberg LL, Mercy JA, Zwi AB, Lozano R. World report on violence and health. Geneva: World Health Organization; 2002.

7. Waters H, Hyder A, Rajkotia Y, Basu S, Rehwinkel JA, Butchart A. The economic dimensions of interpersonal violence. Geneva: Department of Injuries and Violence Prevention, World Health Organization; 2004.

8. Simmel G. Conflict and the web of group affiliations. New York: The Free Press; 1955.

9. Adams BN. Coercion and consensus theories: some unresolved issues. Am J Sociol 1965; 71:714-6.

10. Straus MA. Measuring intra-familiar conflict and violence: the conflict tactics (CT) scales. J Marriage Fam 1979; 41:75-88.

11. Archer J. Assessment of the reliability of the conflict tactics scales: a meta-analytic review. J Interpers Viol 1999; 14:1263-89.

12. Straus M. Handbook for the Conflict Tactics Scales (CTS). Durham: Family Research Laboratory, University of New Hampshire; 2000.

13. Straus MA, Hamby SL, Boney-McCoy S, Sugarman DB. The revised Conflict Tactics Scales (CTS2): development and preliminary psychometric data. J Fam Issues 1996; 17:283-316.

14. Barnett OW, Miller-Perrin CL, Perrin R. Family violence across the lifespan: an introduction. London: Sage Publications; 1997.

15. Hasselmann MH, Reichenheim ME. Adaptação transcultural da versão em português da *Conflict Tactics Scales Form R* (CTS-1), usada para aferir violência no casal: equivalências semântica e de mensuração. Cad Saúde Pública 2003; 19:1083-93.

16. Newton RR, Connely CD, Landsverk JA. An examination of measurement characteristics and factorial validity of the Revised Conflict Tactics Scale. Educ Psychol Meas 2001; 61:317-35.

17. Lucente SW, Fals-Stewart W, Richards HJ, Goscha J. Factor structure and reliability of the Revised Conflict Tactics Scales for incarcerated female substance abusers. J Fam Viol 2001; 16:437-50.

18. Jones NT, Ji P, Beck M, Beck N. The reliability and validity of the revised conflict tactics scale (CTS2) in a female incarcerated population. J Fam Issues 2002; 23:441-57.

19. Straus MA. Cross-cultural reliability and validity of the Revised Conflict Tactics Scales: a study of university student dating couples in 17 nations. Cross-Cult Res 2004; 38:407-32.

20. Nunnally JCJ, Bernstein I. Psychometric theory. New York: McGraw-Hill; 1995.

21. Wilson M. Constructing measures. An item response modeling approach. Mahwah: Lawrence Erlbaum Associates; 2005.

22. Regan KV, Bartholomew K, Oram D, Landolt MA. Measuring physical violence in male same-sex relationships: an item response theory analysis of the Conflict Tactics Scales. J Interpers Viol 2002; 17:235-52.

23. Schafer J. Measuring spousal violence with the Conflict Tactics Scale. Notes on reliability and validity issues. J Interpers Viol 1996; 11:572-85.

24. Embretson SE, Reise SP. Item response theory for psychologists. Mahwah: Lawrence Erlbaum Associates; 2000.

25. Bock RD, Gibbons RD, Muraki E. Full-information factor analysis. Appl Psychol Meas 1988; 12:261-80.

26. Reichenheim ME, Moraes CL. A comparison between the abuse assessment screen and the Revised Conflict Tactics Scales for measuring physical violence during pregnancy. J Epidemiol Community Health 2004; 58:253-7.

27. Moraes CL, Hasselmann MH, Reichenheim ME. Adaptação transcultural para o português do instrumento "*Revised Conflict Tactics Scales (CTS2)*" utilizado para identificar a violência entre casais. Cad Saúde Pública 2002; 18:163-75.

28. Moraes CL, Reichenheim ME. Cross-cultural measurement equivalence of the Revised Conflict Tactics Scales (CTS2) Portuguese version used to identify violence within couples. Cad Saúde Pública 2002; 18:783-96.

29. Wilson D, Wood R, Gibbons RD. TESTFACT: test scoring, item statistics, and item factor analysis. Mooresville: Scientific Software; 1984.

30. Thissen D, Steinberg L, Wainer H. Detection of DIF using the parameters of item response models. In: Holland PW, Wainer H, editors. Differential item functioning. Hillsdale: Lawrence Erlbaum Associates; 1993. p. 67-113.

31. Zimowski MF, Muraki E, Mislevy R, Bock RD. BILOG-MG for Windows 3.0. Lincolnwood: Scientific Software International; 2002.

32. Masters GN, Wright BD. The essential process in family of measurement models. Psychometrika 1984; 49:529-44.

33. StataCorp. Stata Statistical Software/SE, Release 8.2. College Station: Stata Corporation; 2004.

34. Barling J, O'Leary KD, Jouriles EM, Vivian D, Macewen KE. Factor similarity of the Conflict Tactics Scales across samples, spouses, and sites: Issues and implications. J Fam Viol 1987; 2:37-54.

35. Moffitt TE, Caspi A, Krueger RF, Magdol L, Margolin G, Silva PA, et al. Do partners agree about abuse in their relationship? A psychometric evaluation of interpartner agreement. Psychol Assess 1997; 9:47-56.

36. Cascadi M, Avery-Leaf S, O'Leary KD, Slep AMS. Factor structure and convergent validity of the Conflict Tactics Scale in high school students. Psychol Assess 1999; 11:546-55.

37. Wacholder S. When measurement errors correlate with truth: surprising effects of nondifferential misclassification. Epidemiology 1995; 6:157-61.

38. Dosemeci M, Wacholder S, Lubin JH. Does non-differential misclassification always bias a true effect toward the null value? Am J Epidemiol 1990; 132:746-8.

39. Greenland S, Robins JM. Confounding and misclassification. Am J Epidemiol 1985; 122:495-506.

40. Blettner M, Sauerbrei W. Influence of model-building strategies on the results of a case-control study. Stat Med 1993; 12:1325-38.

41. Browning J, Dutton D. Assessment of wife assault with the Conflict Tactics Scale: using couple data to quantify the differential reporting effect. J Marriage Fam 1986; 48:375-9.

42. Edleson JL, Brygger MP. Gender differences in reporting of battering incidences. J Appl Fam Child Stud 1986; 35:377-82.

43. Jouriles EN, O'Leary KD. Interspousal reliability of reports of marital violence. J Consult Clin Psychol 1985; 53:419-21.

44. O'Leary KD, Arias I. Assessing agreement of reports of spousal abuse. In: Hotaling GT, Finkelhorn D, Kirkpatrick JT, Straus MA, editors. Family abuse and its consequences: new directions in research. Newbury Park: Sage Publications; 1989. p. 218-27.

45. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. Qual Life Res 1998; 7:323-35.

46. Behling O, Law KS. Translating questionnaires and other research instrument: problems and solutions. Thousand Oaks: Sage Publications; 2000.

47. van der Linden WJ, Hambleton RK. Handbook of modern item response theory. New York: Springer; 1996.