



# Validity of a common measure of intimate partner violence perpetration: Impact on study inference in trials in low- and middle-income countries

Cari Jo Clark<sup>a,\*</sup>, Irina Bergenfeld<sup>a</sup>, Abbie Shervinskier<sup>a</sup>, Erin R. Johnson<sup>a</sup>, Yuk Fai Cheong<sup>b</sup>, Nadine J. Kaslow<sup>c</sup>, Kathryn M. Yount<sup>a,d</sup>

<sup>a</sup> Hubert Department of Global Health, Rollins School of Public Health, Emory University, USA

<sup>b</sup> Department of Psychology, Emory University, USA

<sup>c</sup> Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, USA

<sup>d</sup> Department of Sociology, Emory University, USA

## ARTICLE INFO

### Keywords:

Intimate partner violence  
Perpetration  
Prevention  
Measurement  
Measurement invariance  
Validity

## ABSTRACT

**Background:** In lower-and middle-income countries (LMICs), studies of interventions to reduce intimate partner violence (IPV) perpetration are expanding yet measurement equivalence of the construct has not been established. We assessed the measurement equivalence of physical and sexual IPV perpetration used in recent trials in LMICs and tested the impact of non-invariance on trial inference.

**Methods:** With data from three recent intervention trials among men (sample size 505–1537 across studies), we calculated tetrachoric correlations among items and used multiple-group confirmatory factor analysis to assess invariance across arms and over time. We also assessed treatment effects adjusting for covariate imbalance and using inverse probability to treatment weights to assess concordance of invariant measures with published results, where warranted.

**Findings:** The average correlation among items was high and increased over time with several items in two studies showing correlations  $\geq 0.85$  at endline. Increases in correlation for physical IPV were concentrated in the treatment arm in two of the studies. The increase in correlation in sexual IPV differed by arm across studies. Across all studies, a correlated two-factor solution was the best fitting model according to the EFAs and CFAs. One study demonstrated measurement invariance across arms and over time. In two of the studies, longitudinal measurement non-invariance was detected in the intervention arms. In post hoc testing, one study attained invariance with a one-factor model and study inference was concordant with published findings. The other study did not attain even partial invariance.

**Conclusion:** Common measures of physical and sexual IPV perpetration cannot be used for valid effect estimation without further refinement. The study highlights the need for an expanded item set, content validity assessments, further measurement invariance testing, and then consistent use of the item sets in future intervention trials to support accurate inference on the effectiveness of IPV perpetration prevention interventions.

## 1. Introduction

Globally, an estimated 27% of women aged 15 years and older report experiencing some form sexual or physical intimate partner violence (IPV) (World Health Organization, 2021). Men disproportionately engage in IPV against their male and female partners (Barker, 2016; World Health Organization and London School of Hygiene and Tropical Medicine, 2010) and the high burden of morbidity and mortality among victims elevates IPV to a critical public health issue (Oram, &

FisherMinnisSeedatWalbyHegartyRoufAngénieuxCallardChandra, 2022; Stubbs & Szoeko, 2022). The development and evaluation of prevention interventions to reduce men's perpetration of IPV is growing globally (Alsina, & BrowneGielkensNoormande Wit, 2024). However, assessing the effectiveness of interventions requires the accurate measurement of IPV perpetration. Our ability to compare estimates of IPV perpetration across study arms and intervention trials requires "cross-group" measurement invariance, meaning that respondents perceive and respond similarly to the scale items regardless of the "group" to

\* Corresponding author.

E-mail address: [cari.j.clark@emory.edu](mailto:cari.j.clark@emory.edu) (C.J. Clark).

<https://doi.org/10.1016/j.ssmph.2024.101683>

Received 4 February 2024; Received in revised form 21 May 2024; Accepted 22 May 2024

Available online 23 May 2024

2352-8273/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

which they belong (Yount, Cheong et al. 2022). Similarly, our ability to compare IPV perpetration over time requires “cross-time” invariance, such that respondents perceive and respond similarly to the scale items over time (Yount et al., 2022a). Cross-group, cross-trial, and cross-time measurement invariance ensures that comparisons of the effects of primary prevention interventions on IPV perpetration can be made.

Despite the public-health importance of IPV perpetration and growth in research on primary prevention (Alsina, & Browne-GielkensNoormande Wit, 2024), comprehensive assessment of the measurement invariance of commonly used scales of IPV perpetration has been limited. Cross-sectional studies have identified non-invariance (i.e., inadequate model fit with or without modifications) (Putnick & Bornstein, 2016) or partial invariance (i.e., adequate model fit only after modification) (Putnick & Bornstein, 2016) of perpetration scales across genders (O’Hara, & PerkinsTeheeBeck, 2018; Shorey, & AllanCohenFiteStuartTemple, 2019; Wareham, & WagersRodriguezNeighbors, 2022), language of administration (Connolly, & NewtonAarons, 2005), and race/ethnicity (Shorey, & AllanCohenFiteStuartTemple, 2019). Collectively, research suggests a potential lack of comparability in the construct of IPV perpetration across intervention and control groups, over time, and across different populations, which are preconditions for accurate comparisons for inference. A study of measurement invariance in the context of assessing the effectiveness of interventions to prevent men’s perpetration of IPV is needed.

The present study addresses this gap. First, employing data from three recent IPV perpetration prevention trials in low- and middle-income countries (LMICs), we aim to explore the factor structure of a scale derived from the Conflict Tactics Scale-Revised (CTS2). Second, we rigorously test whether this scale demonstrates measurement invariance across arm, across time, and across studies.

## 2. Background

The Conflict Tactics Scale (CTS) and its revised and adapted forms are among the most frequently administered scales to assess IPV in community samples (Costa & Barros, 2016). The CTS has been used in its original and adapted forms in a host of cross-sectional multi-country surveys, such as the International Men and Gender Equality Surveys (IMAGES) administered in 32 countries (Equimundo, 2022) and the United Nations Multi-Country Study on Men and Violence in six countries in Asia and the Pacific. In its original or adapted forms, the CTS also has been employed as an outcome in evaluations of perpetration prevention interventions in settings worldwide (Graham, & Embry-YoungMacyMoraccoReyesMartin, 2021; DeHond, & BradyKalokhe, 2023).

### 2.1. Factor structure of the CTS

The CTS is grounded in conflict theory, which conceptualizes violence as “an inevitable part of all human association” (p.75) (Straus, 1979). The scale measures tactics (18 acts) used by and experienced by each intimate partner in the context of a conflict or disagreement. Tactics measured in the CTS include reasoning (later termed negotiation), verbal aggression (later termed psychological aggression), and violence (later termed physical assault), which when tested with principal factors analysis formed three factors (Straus, 1979). When items assessing threats and use of a weapon (knife or gun) were added to the physical assault subscale, Straus reported a fourth factor denoting lethal or “serious” physical assault (1979). However, this factor was comprised of only two items, which is below the minimum recommended threshold of three items to identify a factor (Kenny, & KashyBolger, 1998). A review of studies using the CTS, including additional empirical findings (Ballinger III, 2000), confirmed the presence in many, but not all studies, of a four-factor solution (reasoning, psychological aggression, physical assault, severe physical assault) and broad similarities in the number of factors for men’s and women’s reports of perpetration. This review also

revealed frequent instances of differences in item functioning by gender and a case where the items assessing mild physical assault loaded with psychological aggression for men, whereas for women, items measuring psychological aggression and physical assault loaded on separate factors. Also notable were findings of a unidimensional model as the best fit in one study for women, given the high correlation among the factors, but for men, both one- and two-factor models had poor fit when only psychological aggression and physical assault were modeled.

In a revised version of the CTS (CTS2), the addition of two subscales capturing sexual coercion and injury from physical assault resulted in five sub-scales (Straus, & HambyBoney-McCoySugarman, 1996). These five factors and to some extent, additional factors denoting levels of severity, have been replicated in some, but not all studies, although most prior investigations employing factor analysis have been conducted on samples of women (Jones, & BrowneChou, 2017).

### 2.2. Measurement invariance testing of the CTS

Studies of the measurement invariance of IPV perpetration have primarily included adolescent and young adult samples. In a study of US college students, a sub-set of psychological and physical IPV perpetration items from the CTS2 demonstrated configural and metric invariance and partial scalar invariance across gender in a two-factor model (Wareham, & WagersRodriguezNeighbors, 2022). In a study of adolescents in Canada and Italy using a modified version of the CTS2 physical assault sub-scale, invariance of a unidimensional model was found between males in Canada and Italy and between females in Canada and Italy, but not between genders within the same country (Nocentini, & MenesiniPastorelliConnollyPeplerCraig, 2011). Another study in Chile tested a correlated two-factor model of physical assault (moderate and severe) among college students, which was found to be invariant between men and women (Viejo, & RincónOrtega-Ruiz, 2018). Finally, examination of a four-factor model (moderate psychological, severe psychological, moderate physical, severe physical) of the Modified-CTS Spanish language version demonstrated gender and age invariance among adolescents and young adults in Northern Spain (Ortuño-Sierra, Marugán Garrido et al., 2023). Collectively, findings suggest that that a correlated factor structure is the most commonly identified model like that originally proposed by the authors of the CTS, forming separate, but correlated constructs. All the studies involving men or adolescent boys were cross-sectional, and none included an assessment of sexual violence.

### 2.3. Measurement invariance testing of the CTS in IPV prevention intervention research

The only measurement-invariance study of a CTS-derived scale used in the context of an intervention tested cross-study arm and cross-time invariance using data from four studies of women’s victimization (Clark, & BergenfeldCheongKaslowYount, 2023). When assessing the factor structure, three sexual violence items were dropped, two in one study and one in another, due to poor model fit. Confirmatory factor analyses of the remaining items suggested that a correlated two-factor model (physical and sexual IPV) was the best fitting model among the three studies that included both physical and sexual IPV. All studies demonstrated strong fit and full invariance by study arm. In invariance analyses over time, all models demonstrated strong fit but only one study demonstrated full invariance. In the remaining three studies, a parameter had to be freed to establish partial invariance including the loading for ‘hit’ in one study, the threshold for ‘hit’ in another study, and the threshold for ‘slap’ in another study; however, these modifications did not meaningfully affect latent means or effect-size estimates.

To date, comparable studies that assess the measurement invariance of measures for men’s IPV perpetration are lacking. We begin to fill this gap by conducting measurement invariance testing using data from three randomized trials of interventions designed to prevent men’s

perpetration of IPV. Specifically, we 1) examine the factor structure of the items used to assess men’s perpetration of physical and sexual IPV, 2) conduct tests of measurement invariance across study arms and over time, and 3) assess whether any identified non invariance biases study inferences. Answering these three research questions will provide information about whether the version of the CTS used in the study samples is sufficiently invariant so that valid comparisons can be made between intervention and control arms, the fundamental comparison needed to establish intervention effectiveness.

### 3. Methods

#### 3.1. Study sample

Intervention studies measuring IPV perpetration by men as a primary outcome were identified through the *What Works to Prevent Violence Against Women and Girls Consortium*—a flagship program of the United Kingdom’s Department for International Development (renamed the Foreign and Commonwealth Development Office). The program was designed to test interventions to prevent IPV and other forms of violence against women (Crawford, & Lloyd-LaneyBradleyAthertonByrne, 2020). Studies from this program were chosen since participating investigators were encouraged to use the same outcome measure for IPV, studies were fielded at approximately the same time, and studies were of approximately the same duration. The latter two points are important, given secular changes in the prevalence of IPV. Three studies met the present study’s inclusion criteria including having: a) a panel design, b) repeated measurement of IPV perpetration at the individual level, and c) a control group. All three studies that met the inclusion criteria were cluster randomized trials administered between 2015 and 2019 with at least a baseline measurement and a follow-up measurement 24 months apart. Details of each of these studies are below.

**Indashyikirwa (IND) trial.** IND was undertaken in rural locations in seven districts in Rwanda among married or cohabitating couples participating in village savings and loans associations (VSLA). The study invited participants to volunteer, so no participation rate was provided. The study enrolled 4143 participants, 3153 of whom also were measured at endline resulting in a 76% retention rate. The intervention consisted of 21 sessions or approximately 60 h of group-based training focused on gender, violence, conflict resolution, power dynamics, social norms, economic empowerment, and community change (Stern, & HeiseMcLean, 2018). Those in the control condition participated in the usual VSLA without receiving the couples’ curriculum.

**Stepping Stones and Creating Futures (SSCF) trial.** SSCF was undertaken in urban informal settlements in eThekweni Municipality, KwaZulu-Natal Province, South Africa among men and women 18–30 years old, resident in the community, and not working or in education. The study invited participants to volunteer, so no participation rate was provided. The study enrolled 1357 participants, 1050 of whom also were measured at endline resulting in a 77% retention rate. The intervention consisted of 21 sessions or approximately 63 h of group education for men and women on gender norms, violence, sexual health, communication, conflict resolution, and livelihood development (Gibbs, Washington et al. 2017). Those selected for the control condition were waitlisted to receive the intervention programming at the end of follow-up.

**One Man Can (OMC) trial.** OMC was deployed in neighborhoods in a semi-formal settlement near Johannesburg, South Africa among community-dwelling men aged 18–40 years. The study invited participants to volunteer, so no participation rate was provided. The study enrolled 2406 participants, 1458 of whom were also measured at endline resulting in 63% retention rate. The intervention used community mobilization, peer education, and advocacy of varied frequencies to reduce IPV perpetration (Christofides et al., 2020). Those selected for the control condition received no intervention programming.

#### 3.2. Data

The *What Works* items measuring physical and sexual IPV perpetration were adapted from the World Health Organization’s (WHO) Multi-country Study on Women’s Health and Domestic Violence Against Women (Garcia-Moreno, & JansenEllsbergHeiseWhatts, 2005). Adaptations included the combination of two of the severe WHO physical IPV items into one and some additional minor adjustments to wording. The WHO scale itself was adapted from items in the Conflicts Tactics Scale (CTS) Revised version (Straus, & HambyBoney-McCoySugarman, 1996; Garcia-Moreno, & JansenEllsbergHeiseWhatts, 2005).

All studies measured IPV perpetration in the prior 12 months with the same five physical IPV items and between three and four sexual IPV items, one of which was identically worded across studies (Table 1). Given the scarcity of data in the higher frequency categories, response options were collapsed to be dichotomous (ever vs never in past 12 months). Items were administered in Kinyarwanda in the IND trial, in English, isi-Zulu, or isiXhosa for the SSCF trial, and in English, isiZulu, Xitsonga or Sepedi in the OMC trial.

**Table 1**  
English language wording for items designed to measure physical and sexual IPV perpetration, by study.

	In the past 12 months, how many times have you ...	IND	SSCF	OMC
SLAP	slapped a current or previous girlfriend or wife or thrown something at her which could hurt her?	Yes	Yes	Yes
PUSH	pushed or shoved a current or previous girlfriend or wife?	Yes	Yes	Yes
HIT	hit a current or previous girlfriend or wife with a fist or with something else which could hurt her?	Yes	Yes	Yes
KICK	kick, drag, beat, choke, or burn a previous or current girlfriend, partner, or wife?	Yes	Yes	Yes
WEAPON	threaten to use or actually use a gun, knife, or other weapon against a previous or current girlfriend, partner, or wife?	Yes	Yes	Yes
FORCE SEX	physically forced your current or previous girlfriend or wife to have sex with you when she did not want to?	Yes	Yes	Yes
COERCED SEX	used threats or intimidation to get your current or previous partner, girlfriend or wife to have sex when she did not want to?	Yes	Yes	added "but not physical force"
OTHER SEX	forced your current or previous girlfriend or wife to do something sexual that she did not want to do?	defined forced as "physical force or threats"	Yes	Yes
PORN	forced your current or previous girlfriend or wife to watch pornography when she didn't want to?	NA	Yes	Yes

Notes: Answer options for all three studies included “Never” (0), “One time” (1), “A few times” (2), “Many times” (3). NA = not available.

### 3.3. Analysis

Univariate analyses of all items were conducted by study arm and time (see Table 2 for item prevalences). Tetrachoric correlations were estimated to assess the associations among the items at each time point for each study. Average correlations among items were also computed for each study by type of IPV, study arm, and time.

Given the lack of evidence on the factor structure of the short item set used in this study, we used incremental exploratory factor analyses (EFA) of one to two factors, which is the largest number of theoretically relevant factors that could be fit, given the small number of items available for analysis to assess the factor structure of the physical and sexual IPV items (Research Question 1). We fit these models using a random split-half sample and accounted for the complex sampling design of each study. Model fit was assessed using well-established metrics and cut points of good fit, including the Root Mean Square Error of Approximation (RMSEA <0.06), Comparative Fit Index, and Tucker-Lewis Index (CFI, TLI >0.95) (Cheung & Rensvold, 2002; Hu and Bentler 1995, 1999; Putnick & Bornstein, 2016; Vandenberg & Lance, 2000). Using the best fitting model for each study, we fit models using confirmatory factor analysis (CFA) in the second random split half sample to confirm the fit of this factor structure.

We performed measurement invariance testing (Research Question 2) using multiple group confirmatory factor analysis (MG-CFA) for dichotomous indicators (Millsap, 2012) with weighted least squares mean and variance adjusted (WLSMV) estimators to appropriately handle the categorical nature of the binary indicators. We compared the fit of models with and without factor loadings and thresholds constrained (Davidov, & MeulemanCiecuchSchmidtBilliet, 2014). Model fit was assessed using the criteria described above with non-invariance being defined by a worsening in RMSEA, CFI, and TLI of more than 0.01 (Liu, & MillsapWestTeinTanakaGrimm, 2017).

For studies showing any non-invariance, we used maximum likelihood estimation to determine whether non-invariance arose primarily from loadings or thresholds. Where there was a lack of support for invariance through comparison of configural versus metric models, modification indices from weighted least square mean and variance adjusted (WLSMV) estimation were used to identify potential constraints to relax. We had planned to estimate the severity of the impact of non-invariance on cross-arm comparisons by estimating the standardized differences of latent IPV means between models with and without equality restrictions and calculate the change in the difference-in-difference estimates as a proportion of the standard deviation (Research Question 3). However, as will be seen below, one of the studies demonstrated scalar invariance, not requiring further testing, one achieved only configural invariance, precluding further analysis, and one achieved scalar invariance in post hoc analyses. For this study, the analytic approach described in its impact paper was replicated using

the invariant form of the IPV perpetration variable. The reported impact analysis was a cluster-level comparison of endline perpetration scores, adjusting for baseline perpetration and unbalanced baseline covariates and weighting with inverse probability to treatment weights to account for loss-to-follow up (Christofides et al., 2020).

### 4. Results

The prevalence of physical IPV in the prior 12-months ranged from 24.3% (IND) to 48.0% (SSCF) at baseline and 15.6% (IND) to 40.4% (SSCF) at endline (Table 2). Slap and push were the most frequently reported acts of physical IPV, and weapon was the least frequently reported act across studies and time. The prevalence of sexual IPV in the prior 12-months ranged from 21.0% (IND) to 32.3% (SSCF) at baseline and 13.7% (IND) to 27.7% (SSCF) at endline, with greater variability in the most and least frequently reported items across studies and time.

Table 3 displays the tetrachoric correlations among the items by IPV type, study arm, and assessment wave for each study. The average correlation among items measuring physical IPV ranged from 0.68 to 0.73 across the studies at baseline and 0.72 to 0.88 at endline. The average correlation among items measuring sexual IPV ranged from 0.54 to 0.75 across the studies at baseline and from 0.58 to 0.93 across studies at endline. The average correlation among all items followed a similar increasing trend over time. The average increase in correlation among physical IPV items ranged from 0.03 (SSCF) to 0.15 (OMC). These increases were in the treatment arm in all studies, although the OMC trial also saw equally large increases in the correlation among these items in the control arm. The average increase in correlation among the sexual IPV items ranged from 0.07 (IND) to 0.17 (OMC). These increases were influenced by the treatment arm in the IND trial, the control arm in the SSCF trial, and both arms in the OMC trial. Across studies, the correlation of items within IPV types was generally higher than across types. In the SSCF and OMC trials, several items correlated with other items ≥0.85 within the same type for physical and sexual IPV at endline (SSCF: ‘kick,’ ‘coerced sex,’ ‘other sex’), with all items showing this trend in the OMC trial including most sexual abuse items, which correlated over 0.90. Item-level correlations for each study are available in the online supplement.

In the EFA, the best fitting model was a two-factor solution for all studies (Fig. 1). Of the two-factor models, the OMC study was the ‘cleanest’ with no cross-loadings and all physical IPV items loaded on one factor together while all sexual IPV items loaded together on the other factor. The SSCF model had two cross loadings (‘weapon’ and ‘porn’) while the IND study had one cross loading (‘weapon’). The EFAs with item-level detail are available in the online supplement.

On theoretical grounds we opted not to drop cross-loading items but to proceed with the theoretically aligned two-factor CFAs. Across all studies, the two-factor CFAs had very strong fit (Table 4). Chi-square

**Table 2**  
Prevalence rate of prior-Year physical and sexual IPV items by study, study arm, and time.

ITEM	IND (N = 1537)				SSCF (N = 505)				OMC (N = 1460)			
	Baseline		Endline		Baseline		Endline		Baseline		Endline	
	N = 763	N = 774	N = 763	N = 774	N = 237	N = 268	N = 237	N = 268	N = 746	N = 714	N = 746	N = 714
	Trt %	Ctl %	Trt %	Ctl %	Trt %	Ctl %	Trt %	Ctl %	Trt %	Ctl %	Trt %	Ctl %
SLAP	15.0	11.5	7.0	6.8	33.3	30.7	25.3	28.4	24.0	21.8	15.1	15.6
PUSH	19.9	15.7	11.0	12.6	31.6	33.0	25.7	27.6	24.7	24.1	16.9	16.0
HIT	9.1	7.6	5.1	4.0	25.3	22.5	21.1	22.0	19.6	17.5	12.1	12.5
KICK	2.2	3.2	1.6	1.2	18.6	17.6	15.6	19.4	16.2	13.5	10.7	11.4
WEAPON	0.9	1.8	1.3	0.9	13.5	11.2	13.9	13.8	11.9	11.3	6.7	9.7
FORCED SEX	13.1	11.1	5.8	7.4	16.9	18.7	15.2	17.2	17.0	15.2	10.6	11.0
COERCED SEX	17.2	15.5	8.5	12.8	19.0	17.6	12.7	16.0	15.8	15.4	10.1	9.4
OTHER SEX	2.5	3.1	2.6	2.5	16.9	20.6	15.2	19.8	18.0	16.1	9.8	10.5
PORN	na	Na	na	na	14.8	15.5	14.8	17.2	15.7	14.0	9.8	11.0

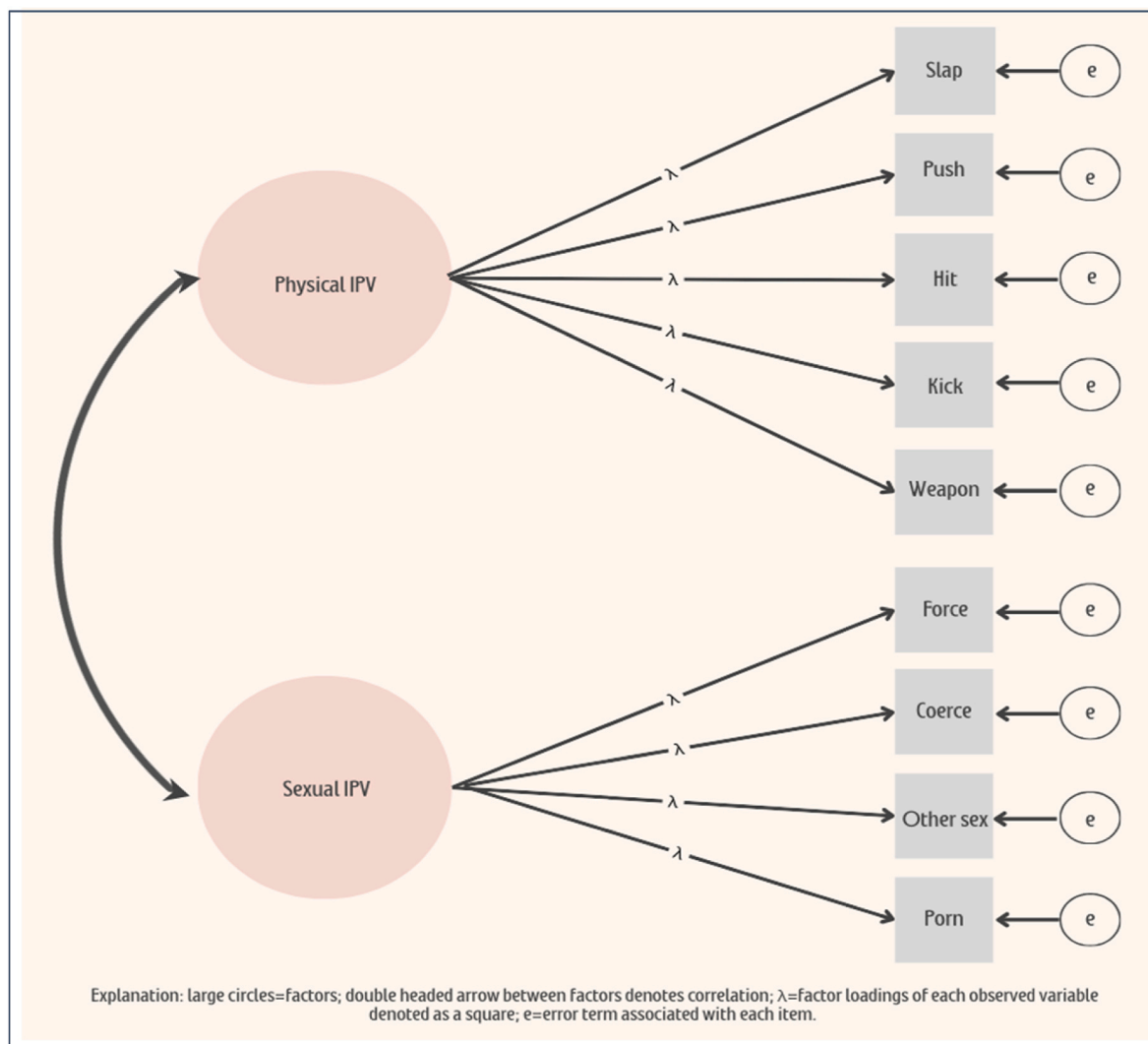
Notes: IND: Indashyikirwa; SSCF: Stepping Stones and Creating Futures; OMC: One Man Can.



**Table 3**  
Correlations among items by study, IPV type, study arm, and time.

	Baseline			Endline			Difference (Endline – Baseline)		
	Mean	Tx	Ctl	Mean	Tx	Ctl	Mean	Tx	Ctl
IND: Physical	0.68	0.53	0.76	0.72	0.74	0.69	0.04	0.21	-0.07
IND: Sexual	0.54	0.43	0.63	0.61	0.58	0.63	0.07	0.16	0.00
IND: Physical and Sexual	0.55	0.40	0.63	0.61	0.62	0.60	0.07	0.22	-0.03
SSCF: Physical	0.73	0.72	0.74	0.76	0.81	0.71	0.03	0.09	-0.03
SSCF: Sexual	0.73	0.83	0.63	0.82	0.84	0.80	0.09	0.01	0.17
SSCF: Physical and Sexual	0.65	0.69	0.61	0.70	0.75	0.66	0.05	0.05	0.05
OMC: Physical	0.73	0.74	0.73	0.88	0.88	0.89	0.15	0.14	0.16
OMC: Sexual	0.75	0.79	0.70	0.93	0.93	0.92	0.17	0.14	0.22
OMC: Physical and Sexual	0.66	0.67	0.65	0.81	0.82	0.80	0.15	0.15	0.16

**Notes:** IND: Indashyikirwa; SSCF: Stepping Stones and Creating Futures; OMC: One Man Can. Tx: Treatment; Ctl: Control.



**Fig. 1.** Best fitting model, correlated two-factor solution, baseline.

difference testing between a one- and a two-factor solution confirmed this finding. The CFAs with item-level detail are available in the online supplement. The two-factor model for each study was subjected to invariance testing by study arm and time (see Table 5 for summary, see Table 6 for detailed results).

The baseline and endline cross-study arm tests all demonstrated scalar invariance, suggesting that the scale had similar measurement properties across arms at these two time points. The SSCF trial demonstrated invariance across time and within each study arm. Both the IND

and OMC trials demonstrated scalar non-invariance in the intervention arms over time, suggesting that the measurement properties of the perpetration scale differed over time in the intervention arms. In the maximum likelihood models for each study, the invariance was concentrated in the thresholds denoted by the likelihood ratio tests comparing 1) the configural to the metric model and 2) the metric to the scalar models. In the OMC trial, there were no suggested threshold or loading modifications that would improve the model fit. In the IND trial, the ‘slap’ threshold was the only threshold identified along with its

**Table 4**  
Exploratory and confirmatory factor analysis by study, baseline.

	Exploratory Factor Analysis		Confirmatory Factor Analysis
	1 Factor	2 Factor	2 Factor
<b>IND</b>	N = 768		N = 769
Chi Square	98.343	16.859	29.558
Degrees of Freedom	20	13	19
Chi Square P-Value	0.000	0.206	0.058
RMSEA	0.071	0.020	0.027
RMSEA 90%CI	(0.058, 0.086)	(0.000, 0.043)	(0.000, 0.045)
CFI	0.943	0.997	0.989
TLI	0.921	0.994	0.984
<b>SSCF</b>	N = 253		N = 251
Chi Square	117.063	25.152	24.764
Degrees of Freedom	27	19	26
Chi Square P-Value	0.000	0.156	0.532
RMSEA	0.115	0.036	0.000
RMSEA 90%CI	(0.094, 0.137)	(0.000, 0.070)	(0.000, 0.047)
CFI	0.952	0.997	1.000
TLI	0.936	0.994	1.000
<b>OMC</b>	N = 728		N = 731
Chi Square	68.204	26.964	31.130
Degrees of Freedom	27	19	26
Chi Square P-Value	0.000	0.106	0.224
RMSEA	0.046	0.024	0.016
RMSEA 90%CI	(0.032, 0.059)	(0.000, 0.043)	(0.000, 0.035)
CFI	0.988	0.998	0.998
TLI	0.984	0.996	0.997

**Notes:** IND: Indashyikirwa; SSCF: Stepping Stones and Creating Futures; OMC: One Man Can.

loading. Freeing the threshold in subsequent models to establish partial scalar invariance did not result in adequate model fit, nor did freeing both its threshold and factor loading and no other loadings or thresholds were identified as modifications that would improve model fit. In post hoc testing of a more parsimonious one-factor model, the OMC study achieved measurement invariance across study arms and time ( $\Delta$ RMSEA,  $\Delta$ CFI, and  $\Delta$ TFI <0.01 for all models); the IND trial did not. The measurement of IPV was noninvariant over time in the intervention arm ( $\Delta$ RMSEA: 0.003,  $\Delta$ CFI: 0.026, and  $\Delta$ TFI: 0.015). In further post hoc testing on the IND trial data to attempt to reach partial scalar invariance, we made several other modifications including dropping the ‘coerced sex’ item in a one-factor model due to its very high correlation with ‘forced sex’, dropping the ‘other sex’ item due to it having the lowest loading in the EFA model, and adding a correlation among the ‘forced sex’ and ‘coerced sex’ items at endline and at both baseline and endline in one and two factor models. The results were the same across all attempts, namely evidence of scalar non-invariance in the intervention arm over time (see the online supplement for results).

**Table 5**  
Measurement invariance testing summary by study.

Level of Measurement Invariance Established ...	Two-Factor Model			One-Factor Model		
	IND	SSCF	OMC	IND	SSCF	OMC
Baseline Cross-Arm	Scalar	Scalar	Scalar	Scalar	Na	Scalar
Endline Cross-Arm	Scalar	Scalar	Scalar	Scalar	Na	Scalar
Cross-time Overall	Scalar	Scalar	Scalar	Scalar	Na	Scalar
Cross-Time Treatment Arm	Configural	Scalar	Configural	Configural	Na	Scalar
Cross-Time Control Arm	Scalar	Scalar	Scalar	Scalar	Na	Scalar

**Notes:** IND: Indashyikirwa; SSCF: Stepping Stones and Creating Futures; OMC: One Man Can; Na: Not applicable.

#### 4.1. Concordance with published study results

Given the invariance in the SSCF trial no re-examination of trial results was needed. The inability to identify a measurement model in the IND trial that was invariant precluded further analysis, as the scale is not measuring the same construct over time in the treatment and control arms. For the OMC trial, we replicated the trial’s endline analysis using the one-factor invariant model finding similar null results (estimate 0.02, p-value 0.42).

### 5. Discussion

The CTS2, and scales derived from it, are the most widely used measures of men’s IPV perpetration for surveillance and impact assessment of primary prevention studies. The measurement invariance of these scales, however, remains unknown. Valid comparison of measures for IPV perpetration in cross-group surveillance and impact assessment require that these measures achieve measurement equivalence.

This novel study addresses this gap. We find that the scale cannot be used to assess intervention impact in its current form. Unlike results of a similar measurement invariance study of women IPV’s victimization (Clark, & BergenfeldCheongKaslowYount, 2023), where at least partial scalar invariance was found within and across trials and the identified non-invariance had negligible impact on inferences regarding intervention effects, the findings among men suggest considerable caution is needed. Only one of the studies demonstrated measurement invariance across study arms and time. Only one study demonstrated measurement invariance after considerable post hoc testing and was therefore modeled differently than the trial impact paper, and one never attained measurement invariance in the intervention arm over time, precluding comparison of change in IPV perpetration between the intervention and control groups, which is foundational to assessing the impact of a prevention intervention.

Across trials, the best fitting exploratory and confirmatory model, a correlated two-factor solution, is aligned with prior research on the CTS (Straus, 1979, Straus, & HambyBoney-McCoySugarman, 1996; Ballinger III, 2000; Jones, & BrowneChou, 2017). However, this model was invariant in only one trial (SSCF), with a more parsimonious one-factor solution needed for invariance in another trial (OMC). In this trial (OMC), the high correlations within physical and sexual IPV might have driven the performance of the one-factor model. The very high correlation among items across studies, especially for sexual IPV, suggests that respondents may not be differentiating among the items. If the items are being perceived to be asking almost the same question, then measurement is inefficient, and the construct lacks content validity.

The increasing correlation among the items over time, especially in the IND trial, and the presence of measurement invariance in the treatment arms in the IND and OMC trials, might reflect the impact of the intervention such that individuals exposed to the intervention may have changed the way they interpreted the questions or other methodological issues such as homogenization of responses, common experience effect, or social desirability bias. However, the OMC trial, the only trial to attempt exposure among community members instead of focused

**Table 6**  
Measurement invariance testing of IPV perpetration measure by intervention study.

IND								
Model	$\chi^2$	d.f.	RMSEA	CFI	TLI	$\Delta\chi^2$ P-value	$\Delta$ RMSEA	$\Delta$ CFI
<b>Baseline cross-arm (n = 1537)</b>								
Configural	53.010	38	0.023	0.995	0.992	–	–	–
Scalar	57.369	42	0.022	0.995	0.993	–	–	–
Configural v scalar	5.294	4	–	–	–	0.258	–0.001	0.000
<b>Endline cross-arm (n = 1534)</b>								
Configural	70.667	38	0.033	0.984	0.977	–	–	–
Scalar	70.816	42	0.030	0.986	0.982	–	–	–
Configural v scalar	3.064	4	–	–	–	0.547	–0.003	0.002
<b>Cross-time (B/E) (n = 1537)</b>								
Configural	161.033	98	0.020	0.981	0.977	–	–	–
Scalar	211.886	112	0.024	0.971	0.968	–	–	–
Configural v scalar	55.972	14	–	–	–	0.000	0.004	–0.010
<b>Cross-time treatment (B/E) (n = 763)</b>								
Configural	113.985	98	0.015	0.989	0.987	–	–	–
Scalar	171.886	112	0.026	0.959	0.956	–	–	–
Configural v scalar	63.570	14	–	–	–	0.000	0.011	–0.030
<b>Cross-time control (B/E) (n = 774)</b>								
Configural	116.099	98	0.015	0.991	0.989	–	–	–
Scalar	140.918	112	0.018	0.986	0.985	–	–	–
Configural v scalar	27.548	14	–	–	–	0.016	0.003	–0.005
<b>SSCF</b>								
Model	$\chi^2$	d.f.	RMSEA	CFI	TLI	$\Delta\chi^2$ P-value	$\Delta$ RMSEA	$\Delta$ CFI
<b>Baseline cross-arm (n = 504)</b>								
Configural	112.336	52	0.068	0.987	0.982	–	–	–
Scalar	114.362	57	0.063	0.988	0.985	–	–	–
Configural v scalar	1.519	5	–	–	–	0.911	–0.005	0.001
<b>Endline cross-arm (n = 505)</b>								
Configural	70.774	52	0.038	0.995	0.994	–	–	–
Scalar	74.119	57	0.034	0.996	0.995	–	–	–
Configural v scalar	2.911	5	–	–	–	0.714	–0.004	0.001
<b>Cross-time (B/E) (n = 505)</b>								
Configural	163.682	129	0.023	0.993	0.992	–	–	–
Scalar	178.144	145	0.021	0.994	0.993	–	–	–
Configural v scalar	20.007	16	–	–	–	0.220	–0.002	0.001
<b>Cross-time treatment (B/E) (n = 237)</b>								
Configural	155.835	129	0.030	0.992	0.991	–	–	–
Scalar	173.098	145	0.029	0.992	0.991	–	–	–
Configural v scalar	22.530	16	–	–	–	0.127	–0.001	0.000
<b>Cross-time control (B/E) (n = 268)</b>								
Configural	137.403	129	0.016	0.996	0.995	–	–	–
Scalar	151.387	145	0.013	0.997	0.997	–	–	–
Configural v scalar	14.882	16	–	–	–	0.533	–0.003	0.001
<b>OMC</b>								
Model	$\chi^2$	d.f.	RMSEA	CFI	TLI	$\Delta\chi^2$ P-value	$\Delta$ RMSEA	$\Delta$ CFI
<b>Baseline cross-arm (n = 1459)</b>								
Configural	72.430	52	0.023	0.996	0.995	–	–	–
Scalar	74.410	57	0.020	0.997	0.996	–	–	–
Configural v scalar	2.282	5	–	–	–	0.809	–0.003	0.001
<b>Endline cross-arm (n = 1458)</b>								
Configural	72.283	52	0.023	0.999	0.999	–	–	–
Scalar	76.781	57	0.022	0.999	0.999	–	–	–
Configural v scalar	4.109	5	–	–	–	0.534	–0.001	0.000
<b>Cross-time (n = 1460)</b>								
Configural	170.113	129	0.015	0.998	0.998	–	–	–
Scalar	265.176	145	0.024	0.994	0.994	–	–	–
Configural v scalar	113.499	16	–	–	–	0.000	0.009	–0.004
<b>Cross-time treatment (n = 746)</b>								
Configural	143.291	129	0.012	0.999	0.999	–	–	–
Scalar	204.284	145	0.023	0.996	0.996	–	–	–
Configural v scalar	75.949	16	–	–	–	0.000	0.011	–0.003
<b>Cross-time control (n = 714)</b>								
Configural	137.955	129	0.010	0.999	0.999	–	–	–
Scalar	171.750	145	0.016	0.997	0.996	–	–	–
Configural v scalar	43.896	16	–	–	–	0.000	0.006	–0.002

**Notes:** IND: Indashyikirwa; SSCF: Stepping Stones and Creating Futures; OMC: One Man Can.

intervention on individual men, reported limited intervention exposure, suggesting that factors other than the treatment are likely affecting measurement invariance over time. For example, the participants may have experienced other unanticipated exposures in the treatment and control arms, given the change in correlation among both arms.

The implication of this study’s findings on published reports of trial

outcomes suggests that considerable caution is needed when using this scale to assess intervention impact since it measured the same construct equivalently across arms and over time in only one trial (SSCF) supporting the study’s inference that the intervention reduced men’s perpetration of IPV (Gibbs, Washington et al., 2020). For the OMC trial, the use of a single factor form of the outcome variable (the only form that was invariant) did

not appreciably alter the study's null finding (Christofides, & HatcherRebomboMcBrideMunshiPinoAbdelatifPeacockLevinJewkes, 2020). The difference between the treatment and control arm was so small that using an invariant form of the outcome variable made no difference. For the IND trial, the reported significant reduction in men's reports of perpetration compared to the control arm (Dunkle, & SternChatterjiHeise, 2020) is not justified given the lack of measurement invariance in the treatment arm over time. However, the IND trial also reported a reduction in women's reports of IPV victimization, and the measure was found to be partially invariant (threshold for slap had to be freed to establish invariance over time), but this minor modification did not affect study inference (Clark, & BergenfeldCheongKaslowYount, 2023). If there had been no measurement of women's victimization, the implications of this study's findings on trial inference would potentially be severe.

### 5.1. Limitations and strengths

The findings must be considered in light of study limitations. We have relied on the English translations of the items used in the studies. The surveys were administrated in up to four different languages and the underlying quality of the translations is unknown. The one study in which invariance was detected also had the smallest sample size and therefore the most limited power. However, this study relied on alternative fit indices (RMSEA, CFI, TLI) for model fit determination, which are less sensitive to sample size (Cheung & Rensvold, 2002). The small number of studies and limited geographic scope limits generalizability of our findings beyond these studies. Despite these limitations, the studies were diverse in terms of preventive intervention programming, which offers a more realistic test of measurement invariance across study-design contexts. These studies also utilized the most common items in LMICs for surveillance and evaluations, providing a strong basis to begin to assess the utility of these items for impact assessments.

### 5.2. Research and policy implications

As has been documented among invariance tests of women's reports of IPV victimization in intervention trials (Clark, & BergenfeldCheong-KaslowYount, 2023) and national surveys (Yount, Bergenfeld et al. 2022; Yount, Cheong et al. 2022), several lessons can be drawn, which may be even more important for measuring men's IPV perpetration. First, there is a lack of consensus on the definition and scope of the construct of IPV perpetration. The basis of the scales used in this study were developed in the 1970s and revised in the 1990s to measure relationship conflict in high-income settings in the Global North. Despite the CTS being designed to measure men's and women's victimization and perpetration, the findings from this study suggest that the scale performs more poorly among men than prior research among women using identical or nearly identical items (Clark, & BergenfeldCheong-KaslowYount, 2023), warranting caution when using the item set to assess the impact of an intervention to reduce perpetration among men.

Second, despite an attempt by the *What Works* Consortium to standardize measurement of IPV across studies, in some studies, items were modified. While pooled testing was not warranted due to variation in factor structures across studies, the items were designed to measure the same act. Modifications in the items increase the likelihood that they are interpreted differently across the studies leading to non-invariance. The effort to modify the sexual IPV items is likely due, in part, to their lack of specificity, which allows for a very large range of interpretations, the opposite of what a survey item is designed to elicit.

Third, the small number of items limits the investigator's ability to drop items should they show item-specific non-invariance across arms and time. While very large item sets are impractical, the current very small item set leaves little flexibility for dropping non-invariant items. The small item set also limits an assessment of severity, as prior research on the CTS and CTS2 are suggestive of separate factors by levels of severity, which if assessed, would offer additional insights into the

impact of interventions across these sub-domains. The limited number of items and very high correlations also suggest inefficient measurement with limited content validity.

Fourth, further research is needed to ascertain if the non-invariance is due to differences in the meaning of the construct across socio-demographic-cultural settings, the methods used (e.g. sampling, non-response, retention, mode of administration, type of intervention), or item-level differences arising from potential differences in translation or wording differences, which are well-established causes of non-invariance in cross-cultural research (Davidov, & MeulemanCieciuchSchmidtBilliet, 2014). A large number of studies with data on each of these domains is needed to answer this question.

Fifth, in the short-term, further measurement-focused research on a larger number of studies across geographic settings and intervention types is needed to ascertain the extent of the problem of non-invariance and whether certain intervention content causes cognitive shifts in interpretation or propensity to report IPV perpetration. In the longer-term, research is also needed that eventually leads to the development of scales that can be compared across groups, studies, and time and relied upon without further measurement invariance testing.

Finally, the lack of invariance across the trials has implications for policy makers. Evidence-based decision-making requires confidence in study findings. The lack of consistency in the performance of the perpetration items suggests that caution is needed when identifying effective interventions for replication and scale-up and has knock-on effects on the veracity of cost-effectiveness assessments—critical considerations when choosing how to invest scarce resources.

## 6. Conclusion

Taken together, the findings suggest caution when using the versions of the CTS used in the sample to assess intervention effectiveness without further refinement of the scale. Until such time, measurement invariance testing is needed to ensure accurate study inferences are drawn from IPV perpetration trials when these item sets are the outcome (s) of interest.

### Ethics statement

The Emory University Institutional Review Board determined that this project does not require IRB review because it is not research with "human subjects", nor is it a "clinical investigation" as defined by federal regulations. It relies entirely on de-identified, publicly available data.

### Declarations of interest

None.

### Funding

Eunice Kennedy Shriver National Institute of Child Health and Human Development (1R01HD099224). The sponsor had no role in the study design, analysis, interpretation, report writing, and the decision to submit the article for publication.

### CRediT authorship contribution statement

**Cari Jo Clark:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Irina Bergenfeld:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation. **Abbie Shervinskis:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation. **Erin R. Johnson:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **Yuk Fai Cheong:** Writing – review & editing, Supervision, Resources,



Methodology, Funding acquisition, Conceptualization. **Nadine J. Kaslow:** Writing – review & editing. **Kathryn M. Yount:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

## Data availability

The data are freely available at <http://medat.samrc.ac.za/index.php/catalog/WW>

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ssmph.2024.101683>.

## References

- Alsina, E., Browne, J. L., Gielkens, D., Noorman, M. A., & de Wit, J. B. (2024). Interventions to prevent intimate partner violence: A systematic review and meta-analysis. *Violence Against Women*, *30*(3–4), 953–980.
- Ballinger III, B. C. (2000). *Factor analysis of the partner and stranger versions of the Conflict Tactics Scale*. Texas Tech University.
- Barker, G. (2016). Male violence or patriarchal violence? Global trends in men and violence. *Sexualidad, Salud y Sociedad (Rio de Janeiro)*, *31*, 316–330.
- Cheung, G., & Rensvold, R. (2002). Evaluating goodness-of-fit indices for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.
- Christofides, N. J., Hatcher, A. M., Rebombo, D., McBride, R.-S., Munshi, S., Pino, A., Abdelatif, N., Peacock, D., Levin, J., & Jewkes, R. K. (2020). Effectiveness of a multi-level intervention to reduce men's perpetration of intimate partner violence: A cluster randomised controlled trial. *Trials*, *21*(1), 1–13.
- Clark, C. J., Bergenfeld, I., Cheong, Y. F., Kaslow, N. J., & Yount, K. M. (2023). Impact of measurement variability on study inference in partner violence prevention trials in low-and middle-income countries. *Assessment*, *30*(5), 1339–1353.
- Connelly, C. D., Newton, R. R., & Aarons, G. A. (2005). A psychometric examination of English and Spanish versions of the revised conflict tactics scales. *Journal of Interpersonal Violence*, *20*(12), 1560–1579.
- Costa, D., & Barros, H. (2016). Instruments to assess intimate partner violence: A scoping review of the literature. *Violence & Victims*, *31*(4), 591–621.
- Crawford, D., Lloyd-Laney, M., Bradley, T., Atherton, L., & Byrne, G. (2020). Final performance evaluation of DFID's what Works to prevent violence against women and girls programme'. *DFID's What Works to Prevent VAWG Programme. Surrey, IMC Worldwide*.
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*, 55–75.
- DeHond, A., Brady, F., & Kalokhe, A. S. (2023). Prevention of perpetration of intimate partner violence by men and boys in low-and middle-income countries: A scoping review of primary prevention interventions. *Trauma, Violence, & Abuse*, *24*(4), 2412–2428.
- Dunkle, K., Stern, E., Chatterji, S., & Heise, L. (2020). Effective prevention of intimate partner violence through couples training: A randomised controlled trial of Indashyikirwa in Rwanda. *BMJ Global Health*, *5*(12), Article e002439.
- Equimundo. (2022). *The international men and gender equality survey: A status report on men, women, and gender equality in 15 headlines*. Washington, DC, Equimundo.
- Garcia-Moreno, C., Jansen, H. A. F. M., Ellsberg, M., Heise, L., & Whatts, C. (2005). *WHO multi-country study on women's health and domestic violence against women: Initial results on prevalence, health outcomes and women's responses*. Geneva: World Health Organization.
- Gibbs, A., Washington, L., Abdelatif, N., Chirwa, E., Willan, S., Shai, N., Sikweyiya, Y., Mkhwanazi, S., Ntini, N., & Jewkes, R. (2020). Stepping stones and creating futures intervention to prevent intimate partner violence among young people: Cluster randomized controlled trial. *Journal of Adolescent Health*, *66*(3), 323–335.
- Gibbs, A., Washington, L., Willan, S., Ntini, N., Khumalo, T., Mbatha, N., Sikweyiya, Y., Shai, N., Chirwa, E., & Strauss, M. (2017). The stepping stones and creating futures intervention to prevent intimate partner violence and HIV-risk behaviours in durban, South Africa: Study protocol for a cluster randomized control trial, and baseline characteristics. *BMC Public Health*, *17*(1), 1–15.
- Graham, L. M., Embry, V., Young, B.-R., Macy, R. J., Moracco, K. E., Reyes, H. L. M., & Martin, S. L. (2021). Evaluations of prevention programs for sexual, dating, and intimate partner violence for boys and men: A systematic review. *Trauma, Violence, & Abuse*, *22*(3), 439–465.
- Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Sage Publications, Inc.
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.
- Jones, R. T., Browne, K., & Chou, S. (2017). A critique of the revised conflict tactics scales-2 (CTS-2). *Aggression and Violent Behavior*, *37*, 83–90.
- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In G. D. Todd, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 233–265). New York: Oxford University Press.
- Liu, Y., Millsap, R. E., West, S. G., Tein, J.-Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, *22*(3), 486.
- Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. Routledge.
- Nocentini, A., Menesini, E., Pastorelli, C., Connolly, J., Pepler, D., & Craig, W. (2011). Physical dating aggression in adolescence. *European Psychologist*, *16*(4).
- O'Hara, K. L., Perkins, A. B., Tehee, M., & Beck, C. J. (2018). Measurement invariance across sexes in intimate partner abuse research. *Psychology of violence*, *8*(5), 560.
- Oram, S., Fisher, H. L., Minnis, H., Seedat, S., Walby, S., Hegarty, K., Rouf, K., Angénioux, C., Callard, F., & Chandra, P. S. (2022). The lancet psychiatry commission on intimate partner violence and mental health: Advancing mental health services, research, and policy. *The Lancet Psychiatry*, *9*(6), 487–524.
- Ortuño-Sierra, J., Marugán Garrido, N., Gutiérrez García, A., Ciarreta López, A., & Camara-Pastor, T. (2023). Measuring violence behaviours in the context of romantic relationships during adolescence: New evidence about the modified conflict tactics scale. *Children*, *10*(2), 297.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90.
- Shorey, R. C., Allan, N. P., Cohen, J. R., Fite, P. J., Stuart, G. L., & Temple, J. R. (2019). Testing the factor structure and measurement invariance of the conflict in Adolescent Dating Relationship Inventory. *Psychological Assessment*, *31*(3), 410.
- Stern, E., Heise, L., & McLean, L. (2018). *Working with couples to prevent IPV: Indashyikirwa in Rwanda*. London, UK.
- Straus, M. A. (1979). Measuring intrafamily conflict and violence: The conflict tactics (CT) scales. *Journal of Marriage and Family*, *75*, 75–88.
- Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised Conflict Tactics Scales (CTS2): Development and preliminary psychometric data. *Journal of Family Issues*, *17*(3), 283–316.
- Stubbs, A., & Szoek, C. (2022). The effect of intimate partner violence on the physical health and health-related behaviors of women: A systematic review of the literature. *Trauma, Violence, & Abuse*, *23*(4), 1157–1172.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.
- Viejo, C., Rincón, P., & Ortega-Ruiz, R. (2018). Physical violence in young Chilean couples: Association with the relationship quality. *Children and Young Services Review*, *93*, 217–225.
- Wareham, J., Wagers, S. M., Rodriguez, L. M., & Neighbors, C. (2022). An exploration of measurement invariance across sex in intimate partner violence perpetration. *Victims and Offenders*, *17*(2), 161–181.
- World Health Organization. (2021). *Violence against women prevalence estimates, 2018: Global, regional and national prevalence estimates for intimate partner violence against women and global and regional prevalence estimates for non-partner sexual violence against women*. Geneva: World Health Organization.
- World Health Organization and London School of Hygiene and Tropical Medicine. (2010). *Preventing intimate partner and sexual violence against women: Taking action and generating evidence*. Geneva: World Health Organization.
- Yount, K. M., Bergenfeld, I., Mhamud, N., Clark, C. J., Kaslow, N. J., & Cheong, Y. F. (2022a). Monitoring sustainable development goal 5.2: Cross-country cross-time invariance of measures for intimate partner violence. *PLoS One*, *17*(6), Article e0267373.
- Yount, K. M., Cheong, Y. F., Khan, Z., Bergenfeld, I., Kaslow, N., & Clark, C. J. (2022b). Global measurement of intimate partner violence to monitor Sustainable Development Goal 5. *BMC Public Health*, *22*(1), 1–14.