Measuring violence perpetration: Stability of teachers' self-reports before and after an anti-violence training in Cote d'Ivoire

Abstract

Background: Epidemiological studies of interpersonal violence commonly use self-reported violence perpetration as an outcome measure, but few studies have investigated the stability of and influences on self-reports.

Objective: To assess changes in teachers' self-reported use of physical violence against students before and after a one-day violence prevention training, and factors associated with changed reports in Cote d'Ivoire.

Methods: Before and after the training, 157 teachers completed surveys containing 32 questions adapted from the ICAST-CI. Changes in physical violence usage were summarized over lifetime, past school term, and past-week timeframes, and the consistency in responses assessed via intraclass correlation coefficients (3,k), percent agreement, and kappa statistics. Factors associated with changed reports were assessed using robust multiple linear regression with 1,000 bootstrapped replications.

Results: Although reports before and after the training should have remained constant, the proportion of teachers reporting 1+ act of violence dropped substantially (lifetime: 73% to 47%). Most teachers (73%) changed 1+ response. Kappa for individual items showed ranging disagreement (lifetime: 0.275-0.795). Variables significantly associated with greater numbers of changed reports included: greater mental health distress (lifetime: beta = 1.061, 95% CI = 0.229, 2.404), older age (past school term: beta = 0.067, 95% CI = 0.018, 0.113); and variables targeted during training, including increasing awareness of consequences of violence (past week: beta =

0.241, 95% CI = 0.046, 0.435) and decreasing acceptance of physical discipline practices in schools (past school term: beta= -0.169, 95% CI= -0.338, -0.045).

Conclusions: Interpreting self-reports of violence perpetration requires caution. Formal investigations into reliability and validity of self-reported violence perpetration and victimization are needed.

Keywords: Violence, measurement, child health, epidemiology, international health

Introduction

Violence is a critical health and human rights issue globally. More than 1 billion children are estimated to experience violence annually (Hillis, Mercy, Amobi, & Kress, 2016), and 1 in 3 women experience non-partner sexual or intimate partner physical or sexual violence in their lifetime (K Devries, Mak, Child, et al., 2013). Experience of violence in childhood, adolescence and adulthood is associated with a host of longer-term adverse health outcomes, including increased risk of suicide and depression, sexually transmitted infections, alcohol and other substance use, and poor educational and employment outcomes (Boden, Horwood, & Fergusson, 2007; K Devries, Child, et al., 2013; K Devries, Mak, Bacchus, et al., 2013; Norman et al., 2012). Large programs of work have begun in diverse global regions to address these experiences ("Gender and adolescence: global evidence (GAGE)," ; "What works to prevent violence"). The overlap between violence against women and children is increasingly prominent on the global health and development agendas, and now features in sustainable development goals 5 and 16 (UN, 2015).

While this focus is welcome, rigorous study into the measurement of violence perpetration and victimization is less well developed. General consensus exists about the necessity of using act-based questions for epidemiological studies, where participants are asked about their use or experience of specific behaviors (e.g., being slapped) without using inherently subjective labels of “violence” or “abuse.” There is also general agreement that proxy reporting vastly underestimates the prevalence of violence (Kuo, Mohler, Raudenbush, & Earls, 2000; Stoltenborgh, van IJzendoorn, Euser, & Bakermans-Kranenburg, 2011), and that self-reports are potentially more accurate. This is because many acts of violence—particularly for sexual violence—remain hidden or stigmatized and are never reported to third parties (Stoltenborgh et

al., 2011), or leave no visible signs of injury. Since the development of the Conflict Tactics Scale in the USA, which was the first act-based measure to explore intimate partner violence, other instruments including the World Health Organization Violence against Women scale (WHO VAW) (WHO, 2005) and the International Society for the Prevention of Child Abuse and Neglect Child Abuse Screening tool (ICAST) (IPSCAN, 2006) have augmented the framing and content of act-based measures of intimate partner violence and violence against children internationally. In practice, victimization questions from the above instruments are often adapted to capture self-reported perpetration, since little work has been done on developing or evaluating measures of perpetration.

Moreover, there has been little critical investigation or systematic study of the factors that may compromise the reliability and validity of act-based self-reported questionnaire items on violence (Mohr & Tulman, 2000). Reliability is the extent to which an instrument (or questionnaire) is free from error whereas validity is the extent to which the instrument measures what it purports to measure. Both of these properties can be compromised if reporting biases are present (e.g. respondents begin to think about the construct they are reporting on in a different way or define it differently at later points compared to at the outset, or the motivation to give a particular responses changes) (Smith et al., 2005). It is important therefore to understand the stability of self-reports over time and what may be driving changes in reporting, in addition to actual behavior change attributable to intervention exposure. Using data from a project in Cote d'Ivoire, we sought to explore whether and to what extent reports of perpetration of physical violence against children from their primary school teachers would change over the course of a short violence prevention training, when there had been no opportunity to perform any additional acts of violence. Our specific aims were to:

1. Describe and quantify differences in recall of teachers' violence perpetration against students over lifetime, past school term, and past week among teachers, immediately before and after a one-day violence prevention training during which teachers had no ability to use more violence against students; and,
2. Examine factors associated with changes in teachers' reports of lifetime, past school term, and past-week violence perpetration against their students.

Methods

Design

As part of an ongoing project to explore workings of a brief school-based violence prevention intervention (K. Devries et al., 2019), we surveyed teachers engaged in an interactive two-day training. Surveys were administered at three timepoints: immediately before and after the first training day and 4 months post-training, once the teachers were implementing classroom strategies. The current analysis used data collected before and after the first training day (i.e. "pre" and "post"). In the absence of any bias or random error, pre and post reports should have remained the same, as the teachers did not have a chance to use any more or less violence against their students during the training period.

Sampling and recruitment of teachers

We sampled all teachers invited to attend a series of training sessions as part of routine intervention delivery by Graines de Paix (GdP), in the city of Man, Cote d'Ivoire. The intervention is delivered by specially-trained teacher counsellors from the Ministry of Education, who are responsible for ongoing professional education and maintenance of teaching standards within defined geographical areas. These 'teacher-trainers', whose regular classroom observations grant them in-depth knowledge of school dynamics, select one to three teachers

from each primary school within their area to attend a two-day training. Counsellors choose teachers based on their perceived ability to influence their peers, to maximize intervention impact once teachers return to their schools. Each training group contained ~20 teachers. We invited all teachers engaged in training groups over a two-week period to participate.

Training content and delivery

The GdP intervention aimed to prevent and reduce violence from teachers to students in Ivorian primary schools by strengthening teachers' ability to use non-violent classroom management techniques. Training activities included personal reflection, group exercises, and guided discussions, and were supported by informational materials (e.g. training manual, activities booklet). Training aims were to increase knowledge, motivation and skills to reduce use of physical violence and improve pedagogical techniques. Further information about the content and structure of the intervention is available (K. Devries et al., 2019; "Graines de Paix,").

Data collection procedures

At the start of the one-day training session, a research assistant introduced the study and gave each teacher a written consent form and information sheet about local support services available for violence and mental health. Consenting teachers were given tablet computers to individually self-complete surveys in French. Surveys were again administered at the training day's end. Data from tablets were immediately uploaded to a password protected database, held on the London School of Hygiene and Tropical Medicine (LSHTM)'s secure server.

Child protection and ethical considerations

The study received ethical approval from LSHTM (ref 14014 and 14537) and the Centre Nationale d'Ethique de la Recherche (CNER) based in Abidjan. Consent forms indicated that if teachers disclosed the use of serious violence against a student, the research team would need to

pass their information to local child protection authorities. A predetermined set of criteria, approved by both ethics committees, were used to determine such referrals. Survey data were examined twice daily after uploaded to identify teachers needing referral. Over the three rounds of surveys, two referrals were made.

Measures

Teachers' use of violence against students was assessed using 32 items adapted from the International Society for the Prevention of Child Abuse and Neglect Child Abuse Screening Tool-Child Institutional (ICAST-CI) (e.g. "Have you ever hit a student with a stick?") (Supplement 1) (IPSCAN, 2006). The reliability and construct validity of the ICAST-CI were tested initially in four countries, and the instrument has since been translated into 20 languages and used widely (IPSCAN, 2006). Demographic characteristics of teachers (i.e. sex, marital status, number of children cared for, position at school, length of time at current job, and highest qualification) were modelled as binary/categorical; teacher age was modelled as continuous. Teachers' own experiences of intimate partner violence, non-partner sexual violence, and child sexual abuse were assessed using items adapted from the WHO Multi-country study on women's health and domestic violence against women (Garcia-Moreno, 2005) and treated as binary variables. The mental health and wellbeing of teachers was assessed via the Self-Report Questionnaire-20 (SRQ-20), which has been widely-used and validated in several low- and middle-income countries (Garcia-Moreno, 2005). Given a lack of validated cut-off in Cote d'Ivoire, the top 30% of the distribution was classified as having a "high" score indicative of probable depression/anxiety (K. M. Devries et al., 2014; K M Devries et al., 2011).

Composite measures were generated to assess intermediate intervention outcomes among teachers. Principal components analyses were carried out of data from the pre-training survey to

determine the number of factors to extract per measure. Exploratory factor analyses were then conducted using oblique (promax) rotation and the iterative principle factor estimation method. The internal consistency of the measures generated were calculated using Cronbach's alpha (Cronbach, 1951). Three measures were generated: an 8-item measure assessing teachers' awareness of the consequences of violence (0-low to 16-high, Cronbach's alpha: 0.81), a 10-item measure assessing teacher self-efficacy in the classroom (0-low to 30-high, Cronbach's alpha: 0.73), and a 6-item measure assessing teachers' acceptance of physical discipline practices in school (0-low to 18-high, Cronbach's alpha: 0.90). Scores were assigned to each response option, and mean scores and standard deviations were calculated.

Analyses

Teachers who did not complete a survey at pre and post (n=3) were removed. Individual items missing (n=6 at pre-survey) were coded as 'skipped.' To assess agreement in responses at each time point for violence as a binary measure, the number and percent of teachers endorsing violence perpetration by responding 'Yes' to any of 32 specific acts of violence was summarized. For each of the 32 items, the percent agreement and kappa statistic were calculated by comparing a categorical measure of violence (0-No, 1-Yes, 99-Skipped) at pre and post. To assess consistency in responses for violence as a continuous measure, a two-way mixed-effects intraclass correlation coefficient (ICC) based on a mean of multiple measurements (3,k) (Koo & Li, 2016) was calculated by comparing a continuous measure of violence at pre and post. Continuous measures were generated by coding 'Yes' as 1 and 'No' or 'Skipped' as 0 (score range: 0-low to 32-high).

The number of changes in teachers' reports of violence between surveys was calculated. A report was classified as having changed if any of the following six combinations was

identified: if the teacher reported 1) 'No' at pre and 'Yes' at post; 2) 'Skipped' at pre and 'Yes' at post; 3) 'Yes' at pre and 'No' at post; 4) 'Skipped' at pre and 'No' at post; 5) 'Yes' at pre and 'Skipped' at post; 6) 'No' at pre and 'Skipped' at post. The number of times a teacher changed his/her response out of 32 items was tabulated and the mean number of changes (with standard deviation and range) was derived.

Multiple linear regression was used to examine associations with the number of changes in teachers' reports of violence between surveys for each timeframe (i.e. lifetime, past school term, and past-week), modelled as continuous variables. We first conducted regression diagnostics of the dependent variables using studentized residuals, leverages, influence measures, q-q plots, and kernel density plots. Based on the results, we ran robust linear regression models using non-parametric bootstrapping with 1,000 replications to reduce the effects of extreme and non-normally distributed data for the number of changes in reports of violence (Hamilton, 2004). Stepwise hierarchical regression was performed at each timeframe. Teachers' background characteristics were added first (Model 1), followed by changes in teachers' intermediate outcomes (Model 2). For intermediate intervention outcomes, the change in the score was modeled as an independent variable by subtracting the beginning-of-training-day score from the end-of-training-day score. Beta coefficients and 95% confidence intervals (CIs) were derived. Collinearity between any pairs of variables was examined using variance inflation factors and none was identified. Analyses were conducted in STATA 14 (StataCorp, 2015).

Results

Surveys were completed by 160 teachers at pre and 157 teachers at post. About two-thirds of the sample was male (63.1%, n=99) and the average age was 37.18 years (SD=7.51).

Most were classroom teachers (75.2%, n=118) and the remainder were school directors, teaching assistants, or volunteers. Most had been in their current jobs for six or more years (59.2%, n=93). Roughly 20-30% reported lifetime non-partner sexual violence victimization, high mental health distress, or lifetime intimate partner violence victimization.

Across timeframes, large reductions in endorsements of any act of violence perpetration against students were observed when treating violence as a binary variable (i.e. 'any' vs. 'no' report of violence) (Table 1). For lifetime reports, the percentage of teachers answering 'yes' to any of 32 acts of violence dropped from 73.3% to 46.9% between surveys, representing a 34.8% decrease; percent decreases for past school-term and past-week reports were 41.6% and 47.8%, respectively. The kappa statistic for individual acts of violence perpetration ranged from 0.275 to 0.795 for lifetime reports, 0.274 to 0.855 for past school-term reports, and 0.189 to 0.798 for past-week reports, indicating a wide range of disagreement for individual items over time. When treating violence as a continuous measure, the 95% CIs for the ICC (3,k) ranged from 0.820 to 0.911 across timeframes (Table 2), suggesting a relatively high level of overall agreement.

The percentage of teachers who changed at least 1 of 32 responses ranged from 58.0% for past-week reports to 73.3% for lifetime reports (Table 2). A transition from either 'Yes' or 'Skipped' at pre to 'No' at post accounted for the largest proportion of changes observed (86.0% for lifetime, 79.0% for past school term, 63.1% for past-week). Of those who changed their responses between surveys, a majority changed between 1 and 3 responses (from 65.1% for past school term to 71.5% for lifetime reports); the remainder changed between 4 and 12 reports (from 18.4% for past-week to 34.7% for past school term). The mean number of changes at each timepoint was highest at lifetime and lowest at past-week.

In multivariable analyses of associations with the number of changes in teachers' reports of violence between surveys, four variables in crude analyses (Table 3) remained significant in adjusted analyses after adjusting for teachers' background characteristics and intermediate intervention outcomes (Table 4, Model 2). Having a high SRQ-20 score was associated with a higher number of changes in lifetime reports of violence compared to having a low SRQ-20 score (beta = 1.061, 95% CI = 0.229 to 2.404). Teachers' increasing age was associated with an increasing number of changes of past school term reports (beta = 0.067, 95% CI = 0.018 to 0.113) and past-week reports (beta = 0.052, 95% CI = 0.017 to 0.092), and was borderline significant for lifetime reports of violence (beta = 0.057, 95% CI = -0.001 to 1.03). Among intermediate outcomes and across all timeframes, an increasing number of changes in reports of violence was associated with an increase in teachers' awareness of the consequences of violence (lifetime beta = 0.177, 95% CI = 0.035 to 0.361; past school term beta = 0.233, 95% CI = 0.020 to 0.406; past-week beta = 0.241, 95% CI = 0.045 to 0.435) and a decrease in teachers' acceptance of physical discipline practices in school (lifetime beta = -0.152, 95% CI = -0.334 to -0.042; past school term beta = -0.169, 95% CI = -0.338 to -0.045; past week beta = -0.126, 95% CI = -0.232 to -0.042). These results did not change in our sensitivity analyses of possible effects from the clustering of teachers within schools.

Discussion

We found substantial changes in teachers' reported use of violence against their students before and after a one-day violence prevention training, with teachers reporting less violence at our later measurement point—despite teachers having no opportunity to use more or less violence. We found that 58% of teachers changed at least one past-week report of violence usage, and 73% changed at least one lifetime report of violence usage. Not all teachers were

equally likely to change reports; having a high SRQ-20 score for mental health distress and increasing age were associated with a greater number of changes. Of note, two of three intermediate intervention outcomes were associated with changes in reports: increasing awareness of the consequences of violence and decreasing acceptance of physical discipline practices in schools. These findings have important implications for measuring violence perpetration, particularly in the context of a violence-reduction intervention study.

There are several types of error which could help to explain discrepancies in reporting. We would expect some random error to occur, and in our sample, teachers reported both more and less violence at post-survey, which is consistent with this possibility. However, it is unlikely to account for the large discrepancies we observed. Our results also suggest some recall bias, as teachers made fewer changes to past-week reports, compared to past school term and lifetime reports. However, we would not expect a large amount of recall bias to occur over a one-day measurement period, so this is again unlikely to account for a large proportion of the discrepancies observed.

More likely explanations include both response shift and social desirability bias, given the delivery of a violence prevention training between pre- and post-reports. Response shift bias occurs when respondents change their understanding of the concept being assessed between surveys (Sprangers & Schwartz, 1999). In this study, teachers' understanding of violence may have changed between pre and post because of the one-day training. Formal investigation into response shift bias has been rare in the violence field (though is more common in other fields (Sprangers & Schwartz, 1999)), but the phenomenon threatens both the comparison of pre- and post-results within an intervention group as well as the comparison of longitudinal results

between control and intervention groups, when the response shift is not equal across groups (Shaw, Cross, & Zubrick, 2015).

Social desirability bias is present when respondents are influenced by what they perceived to be the socially-acceptable response and change their own response accordingly (Smith et al., 2005). In our study, for the many teachers reporting less violence at post-survey, the training may have increased their reluctance to disclose having used these acts against their students. Researchers in other settings have suggested program engagement as a reason for decreased reports of undesirable behaviors, such as violence, at endline (Cornell & Bandyopadhyay, 2010). Conversely, for the few teachers reporting more violence usage at post-survey, it may be that participation in the training enhanced their recognition of the importance of reporting these acts for prevention purposes. Researchers have suggested that participation in a program might increase the salience and hence accurate reporting of certain behaviors of violence at post- compared to pre-survey (Nixon & Werner, 2010; Orpinas et al., 2000).

We found that in teachers who reported higher scores on intermediate outcomes associated with the training intervention (change in awareness of the consequences of violence and change in acceptance of physical discipline practices in schools) were more likely to change their reports. This is consistent with both a possible response shift bias, and with socially desirable responding, as the results suggest that higher levels of engagement with anti-violence intervention material influenced reporting. Interestingly, we also found that participant characteristics including high SRQ-20 score and older age were associated with more changes in reports of violence perpetration for at least one timepoint. We have identified no studies which have investigated the role of respondents' background characteristics on stability of reporting over time, and it is therefore difficult to predict how these characteristics would interact with

reporting. Those participants who had more mental health difficulties and who were older in age may be more prone to recall bias or could also be more likely to have more personal experience with violence and traumatic consequences, and might therefore be more prone to either response shift bias or socially desirable responding when confronted with anti-violence intervention materials. Further research is needed to investigate these response patterns, to explore which types of bias may be present, and to come up with systematic ways to reduce them.

No pattern emerged in our sample for the types of items showing low versus high agreement across timeframes, perhaps due in part to the complete lack of endorsement of some forms of violence. There was a wide variety in kappa statistics for individual measures of agreement, ranging from 'fair' (e.g. 0.271) to 'substantial' (e.g. 0.798) across items (Viera & Garrett, 2005). While the quantification of the consistency in reports revealed "good" to "excellent" agreement between the continuous measures of violence perpetration at pre and post (Koo & Li, 2016), these results may be masking problems that are seen at the level of individual items and may also reflect the psychometric problems of simply adding a list of behaviors to produce a continuous variable (in our sample, the prevalence of violence based on individual items was dramatically different between pre- and post-survey, dropping as much as 48% for past-week estimates). While it is well established that a multi-item scale is more reliable than individual, single items, this assumes that the multi-item scale (or continuous variable) is appropriately derived. In many cases simply adding up individual items does not produce a score that meets basic psychometric requirements for good measurement. A more robust approach would be to use the modern psychometric methods (e.g. Rasch methods (Rasch, 1960)) to develop appropriate and meaningful scales to measure violence. Data gathered using these scales

could then be used to investigate reporting biases such as those identified here in a more robust way.

Results should be interpreted in light of several limitations. Our sample was not random, as teachers were selected for the training sessions based on their perceived ability to influence their peers. It is possible that these teachers who influence their peers may have differed from other teachers in ways that affected the consistency of their responses, which could have biased our results in a conservative or anticonservative direction. For instance, these teachers may have been more affected by what they learned during the training day, leading to greater response shift bias, or they may have been less likely to change their responses since they are viewed as leaders at their schools. There may have been recall bias, particularly in participants' responses to lifetime reports of violence; however, surveying teachers immediately before and after a one-day training is likely to have minimized the effects of this form of bias. Finally, while our measures of intermediate intervention outcomes were created using factor analysis and showed good internal consistency, they have yet to be subjected to a full assessment of psychometric properties, which may have limited their accuracy.

Taken together, our findings suggest the need for caution in interpreting an individual's self-report of violence perpetration, which is often used as an outcome in trials of violence prevention programs (Foshee et al., 2014; Jewkes et al., 2008; Wolfe et al., 2009). Findings highlight that violence prevention training can influence reports of violence, even when the training has not afforded participants opportunity to use any more or less violence. Given the instability in reports of violence perpetration identified in this study, further research is needed to understand factors influencing the measurement of violence perpetration and victimization (Bender, 2017).

References

Bender, A. (2017). Ethics, methods, and measures in intimate partner violence research: The current state of the field. *Violence Against Women, 23*(11), 1382-1413.

Boden, J. M., Horwood, L. J., & Fergusson, D. M. (2007). Exposure to childhood sexual and physical abuse and subsequent educational achievement outcomes. *Child Abuse & Neglect, 31*(10), 1101-1114.

Cornell, D., & Bandyopadhyay, S. (2010). The assessment of bullying. In S. Jimerson, S. Swearer, & D. Espelage (Eds.), *Handbook of bullying in schools: An international perspective*. New York, NY: Routledge, Taylor & Francis.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334. doi:10.1007/BF02310555

Devries, K., Balliet, M., Knight, L., Proceurer, F., N'Diore, A., N'Guessan, D., . . . Quintero, L. (2019). Learn in Peace, educate without violence: Preliminary evidence for the effects of a brief, school-based violence prevention intervention in Cote d'Ivoire *Submitted to Social Science & Medicine*.

Devries, K., Child, J., Allen, E., Walakira, E., Parkes, J., Elbourne, D., . . . Naker, D. (2013). School violence, mental health and educational performance in Ugandan primary school children: A cross-sectional survey. *Pediatrics, 133*(1), 1-9.

Devries, K., Mak, J., Bacchus, L., Child, J., Falder, G., Petzold, M., . . . Watts, C. (2013). Intimate partner violence and incident depressive symptoms and suicide attempts: A systematic review of longitudinal studies. *PLoS Med, 10*(5).

Devries, K., Mak, J., Child, J., Falder, G., Bacchus, L., Petzold, M., . . . Watts, C. (2013). The global prevalence of intimate partner violence. *Science, Express Online, June 20, 2013*. doi:10.1126/science.1240937

Devries, K. M., Child, J. C., Allen, E., Walakira, E., Parkes, J., & Naker, D. (2014). School violence, mental health, and educational performance in Uganda. *Pediatrics, 133*(1), e129-137. doi:10.1542/peds.2013-2007

Devries, K. M., Kiss, L., Watts, C., Yoshihama, M., Deyessa, N., Mbwambo, J., . . . Garcia-Moreno, C. (2011). Violence against women is strongly associated with suicide attempts: evidence from the WHO multi-country study on women's health and domestic violence against women. *Social Science and Medicine, 73*(1), 79-86.

Foshee, V., Reyes, L., Agnew-Brune, C., Simon, T., Vagi, K., Lee, R., & Suchindran, C. (2014). The Effects of the Evidence-Based Safe Dates Dating Abuse Prevention Program on Other Youth Violence Outcomes. *Prev Sci, 15*, 907-916.

Garcia-Moreno, C. J., H. Ellsberg, M. Heise, L. Watts, C. . (2005). *WHO multi-country study on womens health and domestic violence against women*. Retrieved from Geneva:

Gender and adolescence: global evidence (GAGE). Retrieved from https://www.gage.odi.org/

Graines de Paix. Retrieved from http://www.grainesdepaix.org/en

Hamilton, L. (2004). *Statistics with STATA*. Australia: Brooks/Cole CENGAGE Learning.

Hillis, S., Mercy, J. A., Amobi, A., & Kress, H. (2016). Global prevalence of past-year violence against children: A systematic review and minimum estimates. *Pediatrics, 137*(3), e20154079.

IPSCAN. (2006). *ICAST-C : The ISPCAN Child Abuse Screening Tool- Child Version. Manual and Proposed Guidelines for Pilot Administration*. Retrieved from http://www.canalproject.net/pdf/Questionnaire/7_Manual_Child_en.pdf

Jewkes, R., Nduna, M., Levin, J., Jama, N., Dunkle, K., Puren, A., & Duvvury, N. (2008). Impact of Stepping Stones on incidence of HIV and HSV-2 and sexual behaviour in rural South Africa: cluster randomised controlled trial. *Bmj, 337*, a506.

Koo, T., & Li, M. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*, 155-163.

Kuo, M., Mohler, B., Raudenbush, S. L., & Earls, F. J. (2000). Assessing exposure to violence using multiple informants: Application of hierarchical linear model. *Journal of Child Psychology and Psychiatry, 41*(8), 1049-1056.

Mohr, W., & Tulman, L. (2000). Children exposed to violence: Measurement considerations within an ecological framework. *Advanced Nursing Science, 23*(1), 59-68.

Nixon, C., & Werner, N. (2010). Reducing adolescents' involvement with relational aggression: Evaluating the effectiveness of the Creating A Safe School (CASS) intervention. *Psychology in the Schools, 47*, 606-620.

Norman, R. E., Byambaa, M., De, R., Butchart, A., Scott, J., & Vos, T. (2012). The long-term health consequences of child physical abuse, emotional abuse, and neglect: a systematic review and meta-analysis. *PLoS Med, 9*(11), e1001349. doi:10.1371/journal.pmed.1001349

Orpinas, P., Kelder, S., Frankowski, R., Murray, N., Zhang, Q., & McAlister, A. (2000). Outcome evaluation of a multi-component violence-prevention program for middle schools: The students for peace project. *Health Educ Res, 15*, 45-58.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

Shaw, T., Cross, D., & Zubrick, S. (2015). Testing for Response Shift Bias in Evaluations of School Antibullying Programs. *Evaluation Review, 39*(6), 527-554.

Smith, S., Lamping, D., Banerjee, S., Harwood, R., Foley, B., Smith, P., . . . Knapp, M. (2005). Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. *Health Technology Assessment, 9*(10).

Sprangers, M., & Schwartz, C. (1999). Integrating response shift into health-related quality of life research: a theoretical model *Social Science & Medicine, 48*(11), 1507-1515.

StataCorp. (2015). Stata (Version 14.0). Houston, Texas: Stata Corp.

Stoltenborgh, M., van IJzendoorn, M. H., Euser, E. M., & Bakermans-Kranenburg, M. J. (2011). A Global Perspective on Child Sexual Abuse: Meta-Analysis of Prevalence Around the World. *Child Maltreatment, 16*(2), 79-101. doi:10.1177/1077559511403920

UN. (2015). Sustainable Development Goals. Retrieved from https://sustainabledevelopment.un.org/topics/sustainabledevelopmentgoals

Viera, A., & Garrett, J. (2005). Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine, 37*(5), 360-363.

What works to prevent violence. Retrieved from https://www.whatworks.co.za/

WHO. (2005). *WHO multi-country study on women's health and domestic violence against women: summary report of initial results on prevalence, health outcomes and women's*

responses. Retrieved from Geneva: https://www.who.int/reproductivehealth/publications/violence/24159358X/en/

Wolfe, D., Crooks, C., Jaffe, P., Chiodo, D., HUghes, R., Ellis, W. E., . . . Donner, A. (2009). A School-Based Program to Prevent Adolescent Dating Violence: A Cluster Randomized Trial. *Arch Pediatr Adolesc Med, 163*(8), 692-699.

Tables

Table 1

Agreement between pre and post responses for acts of violence performed ever, in the past school term, and in the past week among teachers (n=157)

	Lifetime		Past school term		Past week	
Agreement for violence as a binary variable						
Teachers endorsing any usage of violence	**n(%)**	**Percent change**	**n(%)**	**Percent change**	**n(%)**	**Percent change**
Pre	115 (73.3)	34.8	89 (56.7)	41.6	46 (29.3)	47.8
Post	75 (46.9)		52 (33.1)		24 (15.3)	
Item-specific measures of agreement						
Item	**% agreement**	**Kappa**	**% agreement**	**Kappa**	**% agreement**	**Kappa**
1. Cursed at a student (to discipline them or make them listen)	72.0	0.463	71.3	0.419	75.8	0.323
2. Shouted or screamed at a student (to discipline them or make them listen)	63.7	0.393	65.6	0.417	74.5	0.415
3. Stopped a student from being with other children to make them feel bad, as a punishment	82.8	0.512	83.4	0.449	91.1	0.454
4. Threatened a student with bad marks that they didn't deserve	96.8	0.275	N/A	N/A	N/A	N/A
5. Embarrassed a student in front of other children to punish them	86.6	0.479	87.3	0.390	91.1	0.386
6. Kicked them out of the class as punishment	86.0	0.425	86.2	0.311	91.1	0.189
7. Slapped a student in the face or on the head	95.5	0.425	96.2	0.653	98.1	0.719
8. Slapped them with the hand on their arm or hand	82.2	0.418	86.0	0.455	85.4	0.301
9. Slapped a student in the palm of the hand with any type of 'flexible stick' (chicote) or ruler	87.3	0.642	87.3	0.532	89.8	0.464
10. Twisted their ear as punishment	87.9	0.309	91.1	0.274	93.0	0.244
11. Twisted their arm as punishment	99.4	0.664	99.4	0.664	99.4	0.664
12. Pinched a student on the body	89.8	0.458	91.7	0.348	96.8	0.604
13. Pulled their hair as punishment	98.7	0.745	99.4	0.665	99.4	0.665
14. Hit them by throwing an object at them	98.1	0.720	99.4	0.855	99.4	0.798
15. Hit them with a closed fist	N/A	N/A	N/A	N/A	N/A	N/A
16. Hit them with a stick	84.1	0.621	85.4	0.555	90.5	0.652

17. Caned them	98.7	0.795	98.7	0.795	98.7	0.494
18. Kicked them	N/A	N/A	N/A	N/A	N/A	N/A
19. Knocked them on the head as punishment	87.9	0.675	88.5	0.614	89.8	0.429
20. Forced a student to do hard chores, such as dig, slash a field, or do other hard labor as punishment	N/A	N/A	N/A	N/A	N/A	N/A
21. Hit the tip of their fingers or hands with a ruler or stick as punishment	99.4	0.664	N/A	N/A	N/A	N/A
22. Crushed their fingers or hands as punishment	N/A	N/A	N/A	N/A	N/A	N/A
23. Made them stand/kneel in a given position that hurts to punish them	89.1	0.596	89.1	0.541	91.7	0.356
24. Made them stay outside in the sun or rain to punish them	N/A	N/A	N/A	N/A	N/A	N/A
25. Burnt them as punishment	N/A	N/A	N/A	N/A	N/A	N/A
26. Taken their food away as punishment	97.4	0.490	97.4	0.325	N/A	N/A
27. Forced them to do something that was dangerous	N/A	N/A	N/A	N/A	N/A	N/A
28. Choked them	N/A	N/A	N/A	N/A	N/A	N/A
29. Tied them up (with a rope or belt) at school	N/A	N/A	N/A	N/A	N/A	N/A
30. Tried to cut them purposefully with a sharp object	N/A	N/A	N/A	N/A	N/A	N/A
31. Made them roll over on the ground until they were dizzy as punishment	N/A	N/A	N/A	N/A	N/A	N/A
32. Had sexual intercourse with any of your students	N/A	N/A	N/A	N/A	N/A	N/A
Agreement for violence as a continuous variable						
Intraclass correlation coefficient (3,k), (95%CI), p value	0.879 (0.833, 0.911)		0.874 (0.828, 0.908)		0.869 (0.820, 0.904)	

N/A: No kappa statistic could be calculated because no teachers endorsed this act at both pre and post survey.

Table 2

Summary statistics for changes in teacher reports of the use of violence against students at pre and post, for lifetime, past school term, and past-week across 32 items (n=157 teachers)

	Lifetime	**Past school term**	**Past week**
Teachers changing any response from pre to post- *n(%)*	115 (73.3)	109 (69.4)	91 (58.0)
'No' at pre, 'Yes' at post	22 (14.0)	18 (11.5)	10 (6.4)
'Skipped' at pre, 'Yes' at post	10 (6.4)	2 (1.3)	2 (1.3)
'Yes' at pre, 'No' at post	82 (52.2)	67 (42.7)	38 (24.2)
'Skipped' at pre, 'No' at post	53 (33.8)	57 (36.3)	61 (38.9)
'Yes' at pre, 'Skipped' at post	6 (3.8)	5 (3.2)	2 (1.3)
'No' at pre, 'Skipped' at post	11 (7.0)	12 (7.6)	14 (8.9)
Number of changes made by teachers from pre to post, of those reporting any change^- *n(%)*			
1 change	26 (22.6)	29 (26.6)	34 (37.4)
2 changes	27 (23.5)	27 (24.8)	19 (20.8)
3 changes	29 (25.2)	15 (13.7)	12 (13.2)
4 changes	11 (7.0)	15 (13.7)	11 (12.0)
5-7 changes	17 (10.8)	18 (16.5)	12 (13.2)
8-12 changes	5 (4.3)	5 (4.5)	3 (3.2)
Number of changes from pre to post- *mean(SD), range*	2.29 (2.32), 0-12	2.16 (2.31), 0-10	1.59 (2.07, 0-11)

Note: Data presented reflect changes in reports of individual acts of violence perpetration across 32 items adapted from the International Society for the Prevention of Child Abuse and Neglect Child Abuse Screening Tool-Child Institutional (ICAST-C).

^Among those changing any responses. Denominator for the number of teachers is 115 for lifetime, 109 for past school term, and 91 for past week.

Table 3

Crude associations with changes in reports of violence from pre- to post-survey (lifetime, past school term, and past-week)^ (n=157 teachers)

	Total n(%)	Lifetime (Beta, 95%CI)	Past school term (Beta, 95%CI)	Past week (Beta, 95%CI)
Teacher Characteristics				
Sex				
Male	99 (63.1%)	Ref	Ref	Ref
Female	58 (36.9%)	-0.562 (-1.123, 0.047)	-0.558 (-1.128, 0.022)	-0.037 (-0.537, 0.555)
Marital status				
Single	41 (26.1%)	Ref	Ref	Ref
In a relationship	116 (73.9%)	0.468 (-0.221, 1.057)	0.612 (-0.129, 1.211)	0.338 (-0.198, 0.824)
Number of children cared for (n=156)				
0 children	12 (7.6%)	Ref	Ref	Ref
1-3 children	80 (51.0%)	0.381 (-1.327, 1.482)	0.435 (-1.133, 1.564)	0.078 (-1.797, 0.943)
4+ children	64 (40.8%)	0.438 (-1.128, 1.558)	0.396 (-1.248, 1.448)	0.116 (-2.101, 0.892)
Position at school				
Director	21 (13.4%)	Ref	Ref	Ref
Classroom teacher	118 (75.2%)	-0.297 (-1.302, 0.476)	0.403 (-0.649, 1.206)	0.214 (-0.622, 0.840)
Teacher assistant/volunteer	18 (11.5%)	0.057 (-1.272, 1.356)	0.932 (-0.453, 2.576)	0.715 (-0.461, 1.785)
Length of time at current job				
Less than 1 year	20 (12.7%)	Ref	Ref	Ref
1-5 years	44 (28.0%)	-0.246 (-1.160, 0.681)	-0.111 (-1.091, 0.691)	-0.108 (-1.038, 0.614)
6+ years	93 (59.2%)	0.440 (-0.351, 1.200)	0.531 (-0.428, 1.329)	0.130 (-0.724, 0.892)
Highest qualification				
Secondary school	76 (48.4%)	Ref	Ref	Ref
Higher than secondary	71 (45.2%)	-0.434 (-1.010, 0.231)	-0.422 (-1.163, 0.247)	-0.143 (-0.686, 0.387)
Other	10 (6.4%)	0.680 (-0.954, 2.131)	-0.323 (-1.399, 0.951)	-0.165 (-1.004, 1.339)
Victim of intimate partner violence				
No	123 (78.3%)	Ref	Ref	Ref
Yes	34 (21.7%)	0.823 (0.017, 1.691)	0.666 (-0.308, 1.661)	0.665 (-0.667, 1.589)
Victim of non-partner sexual violence				
No	110 (70.1%)	Ref	Ref	Ref
Yes	47 (29.9%)	0.825 (0.176, 1.440)	1.027 (0.267, 1.914)	0.664 (0.094, 1.368)
Victim of child sexual abuse				
No	141 (89.8%)	Ref	Ref	Ref
Yes	16 (10.2%)	0.225 (-0.786, 1.580)	0.443 (-0.760, 2.134)	0.297 (-0.672, 1.977)

Mental health distress Low SRQ High SRQ	110 (70.1%) 47 (29.9%)	Ref 1.498 (0.707, 2.548)	Ref 0.836 (-0.029, 1.778)	Ref 0.723 (0.102, 1.728)
	Total Mean(SD), range	**Lifetime (Beta, 95%CI)**	**Past school term (Beta, 95%CI)**	**Past week (Beta, 95%CI)**
Teacher Characteristics				
Age	37.18 (7.57), 21-60	0.046 (-0.006, 0.105)	0.043 (-0.011, 0.0960)	0.024 (-0.015, 0.067)
Intermediate Intervention Outcomes^^				
Change in awareness of the consequences of violence (0-low to 16-high)	0.257 (2.27), -9 to 9	0.173 (0.059, 0.288)	0.224 (0.091, 0.368)	0.205 (0.077, 0.357)
Change in self-efficacy (0-low to 30-high)	0.924 (3.2), -16 to 12	0.070 (-0.070, 0.179)	0.059 (-0.113, 0.170)	0.059 (-0.0381, 0.166)
Change in acceptance of physical discipline practices in school (0-low to 18-high)	-0.631 (2.55), -8 to 6	-0.224 (-0.357, -0.122)	-0.239 (-0.399, -0.120)	-0.120 (-0.216, -0.043)

^Assessed using robust linear regression with 1,000 non-parametric bootstrapped replications.

^^Changes are post- minus pre- survey.

Table 4

Adjusted associations with changes in reports of violence from pre- to post-survey (lifetime, past school term, and past-week)^ (n=157 teachers)

	Lifetime		Past school term		Past week	
	Model 1 (Beta, 95%CI)	**Model 2 (Beta, 95%CI)**	**Model 1 (Beta, 95%CI)**	**Model 2 (Beta, 95%CI)**	**Model 1 (Beta, 95%CI)**	**Model 2 (Beta, 95% CI)**
Teacher Characteristics						
Sex						
Male	Ref	Ref	Ref	Ref	Ref	Ref
Female	-0.522 (-1.384, 0.123)	-0.316 (-1.051, 0.324)	-0.495 (-1.357, 0.335)	-0.274 (-0.959, 0.514)	0.058 (-0.702, 0.738)	0.321 (-0.344, 1.119)
Age	0.034 (-0.033, 0.088)	0.057 (-0.001, 1.03)	0.043 (-0.020, 0.093)	0.067 (0.018, 0.113)	0.028 (-0.027, 0.070)	0.052 (0.017, 0.092)
Marital status						
Single	Ref	Ref	Ref	Ref	Ref	Ref
In a relationship	0.513 (-0.431, 1.387)	0.592 (-0.314, 1.374)	0.728 (-0.137, 1.716)	0.724 (-0.135, 1.533)	0.390 (-0.262, 1.268)	0.421 (-0.321, 1.071)
Number of children cared for (n=156)						
0 children	Ref	Ref	Ref	Ref	Ref	Ref
1-3 children	0.664 (-1.319, 2.547)	0.548 (-1.356, 1.915)	0.626 (-1.114, 2.204)	0.539 (-1.237, 1.850)	0.245 (-1.713, 1.316)	0.257 (-1.263, 1.303)
4+ children	0.539 (-1.448, 2.665)	0.147 (-1.960, 1.663)	0.503 (-1.600, 2.341)	0.082 (-2.164, 1.560)	0.176 (-2.658, 1.300)	-0.021 (-1.905, 1.099)
Position at school						
Director	Ref	Ref	Ref	Ref	Ref	Ref
Teacher	-0.151 (-1.429, 1.034)	-0.047 (-1.334, 1.218)	0.407 (-1.422, 1.755)	0.550 (-1.667, 1.788)	0.290 (-1.390, 1.075)	0.607 (-1.170, 1.367)
Teacher assistant/ volunteer	0.859 (-0.695, 2.943)	0.516 (-0.868, 1.991)	1.616 (-0.296, 3.801)	1.309 (-0.576, 3.041)	1.00 (-1.249, 2.631)	1.189 (-0.211, 3.230)
Length of time at current job						
Less than 1 year	Ref	Ref	Ref	Ref	Ref	Ref
1-5 years	0.379 (-0.959, 2.231)	0.275 (-1.106, 1.754)	0.579 (-0.854, 2.364)	0.525 (-0.908, 2.107)	0.223 (-1.321, 1.490)	0.397 (-0.972, 2.022)
6+ years	0.507 (-0.852, 2.013)	0.311 (-1.235, 1.811)	0.706 (-0.77, 2.705)	0.603 (-0.826, 2.562)	0.299 (-0.900, 1.953)	0.349 (-0.947, 2.109)
Highest qualification						

Secondary school	Ref	Ref	Ref	Ref	Ref	Ref
Higher than secondary	-0.583 (-1.414, 0.216)	-0.142 (-0.962, 0.777)	-0.524 (-1.426, 0.307)	-0.110 (-0.911, 0.746)	-0.174 (-1.082, 0.406)	0.222 (-0.443, 0.881)
Other	0.058 (-1.658, 1.702)	-0.131 (-1.995, 1.374)	-0.848 (-2.339, 1.060)	-1.142 (-2.539, 0.206)	-0.376 (-1.815, 1.295)	-0.609 (-1.908, 0.633)
Victim of intimate partner violence						
No	Ref	Ref	Ref	Ref	Ref	Ref
Yes	0.447 (-0.666, 1.545)	0.262 (-0.910, 1.352)	0.346 (-1.195, 1.723)	0.149 (-1.498, 1.477)	0.382 (-0.813, 1.514)	0.088 (-1.173, 1.153)
Victim of non-partner sexual violence						
No	Ref	Ref	Ref	Ref	Ref	Ref
Yes	0.380 (-0.569, 1.207)	0.209 (-1.037, 0.987)	0.491 (-0.631, 1.507)	0.266 (-0.934, 1.184)	0.401 (-0.453, 1.261)	0.341 (-0.387, 1.173)
Victim of child sexual abuse						
No	Ref	Ref	Ref	Ref	Ref	Ref
Yes	-0.196 (-1.228, 0.930)	0.336 (-0.768, 1.515)	0.377 (-1.089, 1.821)	0.906 (-0.046, 2.044)	0.218 (-1.070, 2.060)	0.882 (-0.792, 2.100)
Mental health distress						
Low SRQ	Ref	Ref	Ref	Ref	Ref	Ref
High SRQ	1.231 (0.271, 2.420)	1.061 (0.229, 2.404)	0.725 (-0.418, 1.906)	0.551 (-0.473, 1.717)	0.504 (-0.295, 1.863)	0.489 (-0.293, 1.364)
Intermediate Intervention Outcomes^^						
Change in awareness of the consequences of violence (0-low to 16-high)		0.177 (0.035, 0.361)		0.233 (0.020, 0.406)		0.241 (0.046, 0.435)
Change in self-efficacy (0-low to 30-high)		0.006 (-0.129, 0.131)		0.012 (-0.133, 0.122)		0.036 (-0.074, 0.129)
Change in acceptance of physical discipline practices in school (0-low to 18-high)		-0.152 (-0.334, -0.042)		-0.169 (-0.338, -0.045)		-0.126 (-0.232, -0.042)

Notes: Results for Model 1 are adjusted for all other teacher characteristics; results for Model 2 are adjusted for all other teacher characteristics and intermediate intervention outcomes.

^Assessed using robust multiple linear regression with 1,000 non-parametric bootstrapped replications.

^^Changes are end-of-training-day score minus beginning-of-training-day score.