Journal of Interpersonal Violence
1–16

DOI: 10.1177/0886260515587665
jiv.sagepub.com
SAGE

Measuring Adolescent Violent Behavior Across Groups: Assessing Measurement Invariance of the Violent Behavior Checklist–Modified

Katie L. Cotter, PhD, MSW,[1]
Caroline B. R. Evans, PhD, MSW,[2]
and Paul R. Smokowski, PhD, MSW[3]

Abstract

Measures of violent behavior are often assumed to function identically across different groups (e.g., gender, race/ethnicity). However, failure to verify measurement invariance can lead to biased cross-group comparisons. The current study examines the measurement invariance of the Violent Behavior Checklist–Modified across genders and race/ethnicities. Using multiple group confirmatory factor analysis, configural and metric invariance are assessed in a sample of racially/ethnically diverse middle and high school students (N = 4,128) in two rural counties. Results indicate that the Violent Behavior Checklist–Modified has partial measurement invariance across genders and race/ethnicities. Specifically, four out of six items were non-invariant across genders, while one out of six items was non-invariant across race/ethnicities. Findings suggest that the latent factor of violence may be qualitatively different across males and females. Implications are discussed.

[1]Arizona State University, Phoenix, USA
[2]University of North Carolina at Chapel Hill, USA
[3]University of Kansas, Lawrence, USA

Corresponding Author:
Katie L. Cotter, Arizona State University, School of Social Work, 411 N. Central Avenue, Suite 800, Phoenix, AZ 85004-0689, USA.
Email: katie.cotter@asu.edu

Keywords
youth violence, rural, measurement invariance, adolescents

Adolescence is a tumultuous developmental period. Perhaps it is best summarized by Erikson (1962), who wrote, "In no other stage of the lifecycle [. . .] are the promise of finding oneself and the threat of losing oneself so closely allied" (p. 13). The burgeoning autonomy associated with adolescence results in a deluge of new experiences, including the exposure to risk factors that may increase vulnerability. Youth violence perpetration is one such risk factor that has major developmental consequences. While a single concerted definition of violence does not currently exist, the following definition provided by the World Health Organization is commonly cited:

> The intentional use of physical force or power, threatened or actual, against oneself, another person, or against a group or community, that either results in or has a high likelihood of resulting in injury, death, psychological harm, maldevelopment or deprivation. (Krug, Dahlberg, Mercy, Zwi, & Lozano, 2002, p. 5)

The consequences associated with youth violence have spawned a significant amount of intervention research in recent years. It is imperative that these interventions are accurately evaluated, which depends on the accuracy of the measurement tools used to assess them. Despite its importance, there is a particular dearth of research on whether different groups (e.g., genders and race/ethnicities) interpret violent behavior constructs and/or violent behavior survey items in the same way. According to social role theory, attitudes and behaviors are influenced in part by social norms (Eagly, 1987), suggesting that gender or cultural norms may influence individual's responses to survey questions. Measurement non-invariance refers to differences in observed scores that are due to something other than differences in the latent construct. Measurement non-invariance is problematic because to directly compare multiple groups, it is necessary to assume that measures perform identically across the groups (DeVellis, 2003).

Multiple group confirmatory factor analysis with latent variables allows researchers to test for measurement invariance across groups prior to testing substantive hypotheses. However, few researchers implement this strategy. Gregorich (2006) suggests that this may be due to a lack of awareness that measurement non-invariance threatens meaningful quantitative comparisons. Establishing measurement invariance for violence measures is particularly important given that many researchers compare rates of violence across

genders and race/ethnicities (Frisell, Pawitan, Langstrom, & Lichtenstein, 2012; Peterson, Esbensen, Taylor, & Freng, 2007; Topitzes, Mersky, & Reynolds, 2012; Zheng & Cleveland, 2013).

The current study tests for measurement invariance of the Violent Behavior Checklist–Modified across gender and racial/ethnic groups using multiple group confirmatory factor analysis. A detailed description of this process is provided.

Measurement Invariance Across Groups

A measure is considered to be invariant if respondents from different groups, with the same true score, also have the same observed score (i.e., an individual's probability of an observed score does not depend on group membership; Wu, Li, & Zumbo, 2007). When measures lack invariance, the term *non-invariance* is commonly used. In a factor analysis framework, a factor is used as a proxy for a person's true score and the items are the observed variables (Wu et al., 2007). Evidence of measurement invariance exists if the relationships among observed variables and factors are equivalent across groups. This indicates that a given measure functions the same way across groups. Partial measurement invariance exists when these relationships are equivalent across groups for some (but not all) items in a measure.

Measurement invariance is a necessary prerequisite for comparisons across groups. In practice, this means that without evidence of measurement invariance, cross-group comparisons can result in inaccurate conclusions. For instance, in a study on attitudes toward teen-dating violence researchers found evidence of measurement non-invariance across genders (Edelen, McCaffrey, Marshall, & Jaycox, 2009). Specifically, adolescents indicated more acceptance of retaliatory hitting when the victim was of his or her own gender. After accounting for this non-invariance, results indicated that compared with females, males were more accepting of cross-gender violence, which was not initially evident. These findings highlight that the existence of measurement non-invariance can lead to inaccurate findings.

Measurement non-invariance may result from construct bias, item bias, or method bias (Byrne & Watkins, 2003). Construct bias indicates that the construct of interest has differential meaning across groups. Item bias refers to differential interpretations across groups at the item level. Finally, method bias refers to differential responses across groups due to aspects of the assessment (e.g., certain groups may be more familiar with Likert-type scales) or aspects of administration (e.g., certain groups were given more guidance than others). When measurement non-invariance occurs, researchers should consider the source of the non-invariance and correct for it in future studies, if possible.

Measurement Invariance in Violent Behavior Across Genders

Although measurement invariance is a prerequisite for meaningful cross-group comparisons, few studies have tested for measurement invariance of adolescent violence measures. In a longitudinal study, the "anti-social conduct" of 1,037 participants was assessed at eight time points using a six-item scale that included three violent behavior items (i.e., physical fighting, bullying others, destroying property; Odgers et al., 2008). Focusing on the five data collection points that occurred in childhood and adolescence (i.e., between the ages of 7-15), researchers tested the scale for measurement invariance. Findings indicated that the scale was invariant across genders and concluded that it measured the same antisocial construct for males and females.

Another group of researchers tested for invariance of the Crime and Violence Scale (CVS) using a sample of 7,435 youth, the majority of whom (i.e., 73%) were under the age of 18 (Conrad, Riley, Conrad, Chan, & Dennis, 2010). Invariance was tested using differential item functioning (DIF). In contrast to the findings of Odgers and colleagues (2008), results showed that the CVS was non-invariant across genders based on the seriousness of the violent and criminal behavior. Females' scores tended to be composed of less violent crimes (e.g., slapping a person), whereas males' scores tended to be composed of more violent crimes (e.g., hurt another person resulting in the need for medical attention). These discrepant findings indicate the need for additional research.

Measurement Invariance in Violent Behavior Across Racial/Ethnic Groups

Given the importance of establishing measurement invariance prior to making cross-group comparisons, researchers have tested for measurement invariance based on race/ethnicity across a broad range of topics, such as school engagement (Glanville & Wildhagen, 2007), childhood trauma (Thombs, Lewis, Bernstein, Medrano, & Hatch, 2007), and obsessive compulsive symptoms (Garnaat & Norton, 2010). Despite this body of research, minimal research examining invariance across racial/ethnic groups on violent behavior scales exists. One group of researchers attempted to test for measurement invariance of a violence scale across racial/ethnic groups, but they were unable to complete the analysis due to the low response rates of select groups (Conrad et al., 2010). Another group of researchers tested for and reported evidence of measurement invariance in the externalizing behavior

subscale of The Diagnostic Interview Schedule for Children across African American, White, and Hispanic adolescents (Feaster et al., 2010). Given the lack of studies in this area, additional research is needed to explore whether violent behavior measures function differently across racial/ethnic groups.

Based on social role theory and given the previous work of Conrad et al. (2010), it was hypothesized that results of the current study would reveal partial measurement invariance across genders on the Violent Behavior Checklist–Modified. Specifically, it was hypothesized that items assessing more severe forms of violence (i.e., "beaten somebody up," "used any weapon in a fight," "gotten involved in a fight with one group of kids fighting another group of kids") would be non-invariant across genders. Given the lack of studies assessing invariance across race/ethnicity, this research question was considered exploratory and no specific hypotheses were delineated.

Method

Participants

The United States Centers for Disease Control and Prevention funded the current study through a cooperative agreement with the North Carolina Academic Center for Excellence in Youth Violence Prevention (NC-ACE). The sample data came from the NC-ACE's Rural Adaptation Project (RAP), a 5-year longitudinal panel study of more than 5,000 middle school students from 28 public schools in two rural and economically disadvantaged counties in North Carolina. The data used in the current study were collected in spring 2013 (i.e., year 3 of the 5-year project). All middle school students in Grades 6 through 8 in County 1 were included in the sample. Because County 2 had a larger student population, a random sample of 40% of middle school students was included from County 2.

In both counties, data were collected using an online assessment tool that students completed in school computer labs that were closely monitored by research staff. Following school district policies, County 1 adopted the assessment as a part of normal school procedures and all students were included on the study roster. Parents from County 2 received a letter explaining the study; if they did not want their child(ren) to participate, they returned the letter requesting non-participation and their child was removed from the study roster. Students assented to participate by reading and electronically signing an assent screen prior to completing the online assessment and were informed that they were free to decline participation at any time. Each student received a US$5 gift card for his or her participation. To maintain confidentiality,

student assessments had an identification number attached and no identifying data were collected.

The initial sample consisted of 5,371 participants, but some participants were removed because their assessments were missing data for all variables of interest. In addition, participants self-identifying their racial/ethnic status as Hispanic/Latino, Asian, other, or mixed race were removed due to small sample sizes. The final analytic sample included 4,128 participants. A series of bivariate analyses (i.e., *t* tests, chi-square tests) were performed to identify demographic differences between students included in the analysis and students removed from the analysis. The results indicated no significant differences between the two groups in terms of age or gender. However, the unanalyzed sample was more likely to receive free/reduced price lunch than the analyzed sample (i.e., 11.62% more likely, $p < .001$). In addition, the unanalyzed sample was slightly less likely to live with a family with two adults than the analyzed sample (i.e., 4.78% less likely, $p < .01$). Finally, the unanalyzed sample was more likely to speak a language other than English at home than the analyzed sample (i.e., 13.73% more likely, $p < .001$).

The racial/ethnic composition of the final sample mirrored the diversity of the community: 37.06% ($n = 1,530$) of participants identified as White, 31.88% ($n = 1,316$) identified as American Indian (Lumbee), and 31.06% ($n = 1,282$) identified as African American. The sample was nearly evenly divided by gender, with 51.09% ($n = 2,109$) of participants identifying as female. The mean age of the sample was 14.03 years. More than half of the sample (63.69%; $n = 2,585$) received free/reduced price lunch, and 99.03% ($n = 4,077$) spoke English at home.

Measures

The School Success Profile (SSP; G. L. Bowen & Richman, 2008) is a youth self-report survey that measures adolescent's perceptions about their health and well-being, school experiences, friends, family, neighborhood, and self. A modified version of the SSP, the SSP+ was used for the RAP study. The SSP+ includes 25 of the subscales from the SSP, plus five additional subscales that measured constructs that were not captured by the SSP subscales. One such subscale (Violent Behavior Checklist; Dahlberg, Toal, Swahn, & Behrens, 2005; Nadel, Spellmann, Alvarez-Canino, Lausell-Bryant, & Landsberg, 1996) was used to measure violent behavior (perpetration) in the RAP study and is the focus of the current analysis.

Specifically, a modified version of the Violent Behavior Checklist (Dahlberg et al., 2005; Nadel et al., 1996) was used to measure adolescent

violent behaviors. Six out of 14 items were selected for this study due to the length of the SSP+ assessment. The items were selected based on relevance to physical violence (rather than verbal or emotional abuse) and the wording of some of the items was changed slightly for clarity. The six items asked adolescents how often on a Likert-type scale (i.e., never, once, sometimes, or often) they engaged in the following violent behaviors in the previous 12 months: "hit or kicked someone," "pushed or shoved someone," "beaten somebody up," "used any weapon in a fight," "gotten involved in a fight with one group of kids fighting another group of kids," and "used physical force to get others to do what you want." Cronbach's alpha reliabilities were .83 for males, .81 for females, .83 for African Americans, .83 for American Indians, and .81 for White adolescents.

Multiple Group Confirmatory Factor Analysis

Multiple group confirmatory factor analysis was carried out in several stages. First, configural invariance was tested. Configural invariance indicates that the same factor model exists across groups (Wu et al., 2007). To test configural invariance, the single-factor, six indicator model was constrained to be the same across groups. Following Cheung and Rensvold (2002), the overall root mean square error of approximation (RMSEA) was used to assess configural invariance. The Comparative Fit Index (CFI) and Tucker Lewis Index (TLI) were used as supplemental indices. These indices assess the relative improvement in fit of the specified model compared with a baseline model. Before running the analyses, cutoff values that indicated good model fit, were established for each fit index. In accordance with best practice, RMSEA values of .06 or lower and CFI and TLI values of .95 or higher were considered indicative of good model fit (Hu & Bentler, 1999).

In the second stage of analysis, metric invariance (i.e., that all factor loadings or lambdas are equivalent across groups) was tested. Metric invariance was tested by constraining all lambdas to be equal across groups and using a chi-square difference test (i.e., the DIFFTEST option in MPLUS) to compare the constrained-lambda model with the unconstrained-lambda model. A statistically significant chi-square value would indicate that model fit got significantly worse after constraining the lambdas indicating the absence of metric invariance. Some researchers have argued that the metric invariance assumption is difficult to achieve and that cross-group comparisons can still be made if the non-invariant items make up only a small portion of the model (Byrne, Shavelson, & Muthén, 1989; Cheung & Rensvold, 2002). Therefore, the next step involved testing each lambda individually to determine which specific items were non-invariant.

In Stage 3, each lambda was constrained individually and chi-square difference tests were used to gauge changes in model fit. Again, a statistically significant chi-square value would indicate that the individual factor loading was non-invariant across groups. These stages were executed to test for invariance across genders (i.e., male and female) and then repeated to test for invariance across the three racial groups (i.e., African American, White, and American Indian). Analysis was conducted in Mplus version 7.0 (Muthén & Muthén 1998-2012) using the mean and variance-adjusted weighted least squares (WLSMV) estimator given ordinal variables.

Results

Upon specification of the measurement model in Stage 1, modification indices indicated significant improvement in the model χ^2 value if the errors of two observed variables "hit or kicked someone" and "pushed or shoved someone" were free to correlate. Minimal model modifications based on empirical evidence is considered defensible if they are theoretically sound and do not result in significant changes to the model parameters (Byrne et al., 1989). Each of these requirements was met. After constraining the single-factor model to be equal across genders, the overall RMSEA was .052 with a 90% confidence interval between .046 and .058. The CFI and TLI values were both .996. These results suggest that configural invariance existed across genders. After constraining the single-factor model to be equal across racial/ethnic groups, the overall RMSEA value was .043 with a 90% confidence interval between .037 and .050. In addition, the CFI and TLI values were both .997. This suggests that configural invariance can be assumed across racial/ethnic groups as well.

After testing for metric invariance across genders in Stage 2 (i.e., constraining all of the lambdas to be equal across genders), the chi-square difference test was statistically significant, $\chi^2(5) = 17.77$, $p < .001$. This indicates that the model fit got significantly worse when all of the lambdas were constrained to be equal and that all factor loadings were not equivalent between males and females. After testing for metric invariance across racial/ethnic groups (i.e., constraining all of the lambdas to be equal across racial/ethnic groups), the chi-square difference test was statistically significant, $\chi^2(5) = 23.15$, $p < .001$. This indicates that model fit got significantly worse when all of the lambdas were constrained to be equal and that all factor loadings were not equivalent across groups.

In Stage 3, to determine which lambda(s) was (were) non-invariant, each lambda was tested individually by comparing a model with the lambda freely estimated to a model with the lambda constrained to be equal across groups.

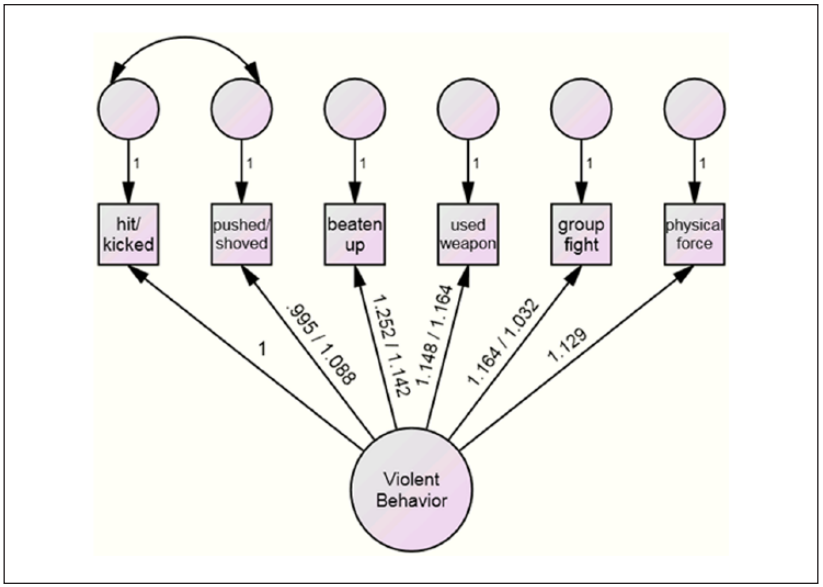

Figure 1. Multiple group confirmatory factor analysis for males and females.
Note. Path coefficients are unstandardized. A single parameter denotes a constrained path. For unconstrained paths, male parameters are listed first, followed by female parameters.

In terms of gender differences, chi-square difference tests indicated non-invariance in the following four items: "pushed or shoved someone," "beaten somebody up," "used any weapon in a fight," and "gotten involved in a fight with one group of kids fighting another group of kids." Specifically, the strength of the relationships between the following items and the latent variable were stronger for males compared with females: "beaten somebody up," "used any weapon in a fight," and "involved in a group fight." On the other hand, the relationship between the item "pushed or shoved someone" and the latent variable was stronger for females compared with males. The final model results are displayed in Figure 1. The final model had adequate model fit. The chi-square value was 222.266 (28), $p < .001$. The obtained RMSEA was .058 with a 90% confidence interval between .051 and .065. The model had a CFI and TLI of .996 and .995, respectively.

In terms of differences across racial/ethnic groups in Stage 3, chi-square difference tests indicated non-invariance in a single item: "used a weapon in a fight." This indicates that the strength of the relationship between this item and the latent factor is different for at least one racial/ethnic group. Additional chi-square difference tests were used to determine which racial/ethnic

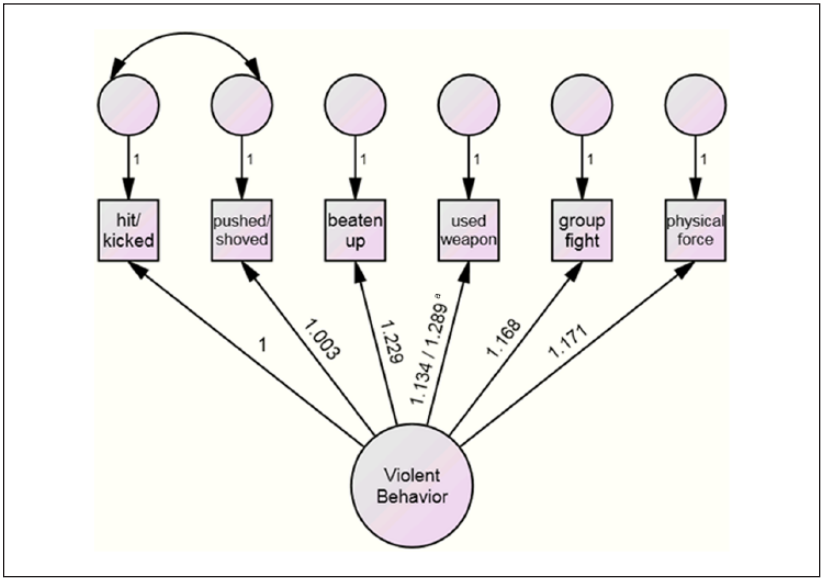

Figure 2. Multiple group confirmatory factor analysis for African American, American Indian, and White groups.
Note. Path coefficients are unstandardized. A single parameter denotes a constrained path.
[a]Unconstrained parameter value for White adolescents.

group(s) differed on this item. Results indicated that non-invariance existed for White students on this item; the relationship between the item and the latent variable was weaker for White adolescents compared with African American and American Indian adolescents. The final model had good model fit. The chi-square value was 199.348 (55), $p < .001$. Although a non-significant chi-square value is desirable, non-significant results are not uncommon, especially with large sample sizes. The obtained RMSEA was 0.044 with a 90% confidence interval between .037 and .050. The model had a CFI and TLI of 0.997. The final model results are displayed in Figure 2.

Discussion

Despite the proliferation of studies assessing measurement invariance, few studies have tested the measurement invariance of violent behavior scales. The current study examined measurement invariance of the Violent Behavior Checklist–Modified across genders and three racial/ethnic groups. In line with our hypothesis, findings indicated that partial invariance existed across

genders. Partial invariance existed across race/ethnicities as well, although the partial invariance across genders was more "severe" than the partial invariance across race/ethnicities. Four out of six factor loadings (or lambdas) were non-invariant across genders, while only one of six factor loadings was non-invariant across race/ethnicities.

In line with our hypothesis regarding gender, items assessing more severe forms of violence (i.e., "beaten somebody up," "used any weapon in a fight," "gotten involved in a fight with one group of kids fighting another group of kids") were non-invariant across groups, such that stronger factor loadings were observed for males compared with females. However, contrary to our hypothesis, the item "pushed or shoved someone" was also non-invariant across genders. This less severe form of violence showed the opposite trend compared with the more severe items: Factor loadings were stronger for females than males. As a whole, these results coincide with existing research, which has documented partial invariance across genders on another crime/violence scale (Conrad et al., 2010).

The existence of non-invariance (or partial invariance) indicates that measure is not functioning equivalently for males and females. The source of this non-invariance can be at the construct, item, or method levels. Given that males and females were treated identically during the data collection procedures and that each gender is equally likely to have completed similar surveys in the past, method bias can be ruled out. After careful consideration, we also determined that it was unlikely that particular items were interpreted differently based on gender. The items describe behaviors using straightforward language that describe specific acts (e.g., hitting or kicking, pushing or shoving) that leave little room for interpretation. It is, however, possible that the construct of violent behavior has different qualitative meaning for males and females based on gender norms. According to social role theory, gender norms influence behavior and beliefs (Eagly, 1987). Gender norms may create unequal thresholds dictating what kind of behavior is acceptable for males versus females. For instance, males engaging in relatively less serious behaviors such as pushing/shoving may be considered normative "roughhousing" (the "boys will be boys" adage). At the same time, these same behaviors among females may be considered egregious. Thus, while this less serious form of violence is a strong indicator of "violent behavior" for females, pushing/shoving is a weaker indicator of "violent behavior" for males because it is a socially acceptable behavior for males. At the same time, the more serious behaviors such as beating others up and engaging in group fighting were stronger indicators of violent behavior for males compared with females.

Recommendations on dealing with non-invariance remain largely unresolved in the literature. In practice, four major options for dealing with partial

invariance are as follows: (1) allow cross-group comparisons on all items despite lack of measurement invariance, (2) restrict cross-group comparisons with those items that were invariant, (3) avoid cross-group comparisons on all items (Gregorich, 2006), or (4) constrain the invariant items to be equal and allow non-invariant items to vary (Byrne et al., 1989). Unfortunately, guidelines for choosing an appropriate option are lacking. Option 1 is probably least desirable as it ignores the evidence that a measure is functioning differently across groups. Although less problematic, Option 2 may lead to several different versions of a scale for multiple cross-group comparisons and the potential for incomplete coverage of the construct (Chen, 2008). Option 3 advocates that the presence of one or more non-invariant items suggests that the latent factor has different meaning across groups, and therefore the measure cannot be used for cross-group comparisons (Gregorich, 2006). Option 4 appears to introduce less bias than Option 1, but it creates questions as to why the non-invariant items are different and what the implications are for conceptualization of the construct. Chen (2008) suggests comparing the groups on the statistic of interest after constraining all items to be equal and after allowing non-invariant items to vary.

A lack of invariance in factor loadings can lead to artificial interaction effects in predictive relationships and inaccurate differences in means (Chen, 2008). Given that adolescent violence researchers consistently compare mean violence scores between groups and test for moderation by demographic groups (e.g., Frisell et al., 2012; Miller, Gorman-Smith, Sullivan, Orpinas, & Simon, 2009; Peterson et al., 2007), measurement invariance is a particularly important issue in this substantive area. That is, failing to test measurement invariance could lead to inaccurate conclusions and because research informs practice, these inaccurate conclusions could translate into ineffective intervention practices. For instance, if moderation analyses were completed to test for gender differences in risk factors for violent behavior without testing measurement invariance, it is possible that unidentified non-invariance could lead to artificial differences by gender. These inaccurate results could be used to inform gender-specific intervention programming, which in turn may not address the true risk factors for male and female violence.

Most often, biases in mean scores are such that the mean score of the group with the higher factor loadings is inflated and the mean score of the group with the lower factor loadings is deflated. This indicates that, at least in the current study, violent behavior among males is upwardly biased (i.e., a mean score higher than the "true" mean score), whereas violent behavior among females is downwardly biased (i.e., a mean score lower than the "true" mean score). In the current model, the mean violent behavior score was constrained to 0 for males for identification purposes. After constraining all

items to be equal, the mean score for females was −0.227 (*SE* = .045), which is significantly different than 0 ($p < .001$). After allowing non-invariant items to vary, the mean score for females was −0.211 (*SE* = .044), which is also significantly different than 0 ($p < .001$). In this case, the same conclusion (i.e., that male and female mean scores were significantly different) remained whether the items were constrained to be equal or the invariant items were allowed to vary. However, the extent of the difference between means was slightly exaggerated when all items were constrained to be equal. The measurement model (in which non-invariant items are allowed to vary and invariant items are constrained to be equal) can also be incorporated into substantive analyses through structural equation modeling (N. K. Bowen & Guo, 2012). Future studies using the Violent Behavior Checklist should use a similar approach before making cross-gender comparisons.

The second research question related to the invariance of the Violent Behavior Checklist across racial/ethnic groups. In contrast to the results based on gender, only a single item was non-invariant: "used any weapon in a fight." Specifically, the factor loading was weaker for White adolescents compared with African American and American Indian adolescents. Compared with the invariance testing across genders, the partial invariance across race/ethnicities is less problematic. It is reasonable to follow the second option and exclude this single item from cross-group comparisons and conduct additional tests to determine the validity of the remaining five items. The minimal non-invariance suggests that the construct of violent behavior functions similarly across racial/ethnic groups.

Overall, the current study has salient implications for adolescent violence research. Prior to testing substantive hypotheses regarding cross-group differences, it is imperative to test for measurement invariance. Failure to assess measurement invariance yields cross-group comparisons that may be biased. This is particularly important for researchers making cross-gender comparisons for violence perpetration. Additional research is needed to test for measurement invariance on victimization measures, aggression measures, and additional violent behavior measures. In addition to gender and racial/ethnic groups, invariance can be tested across age groups, socioeconomic statuses, and other demographics.

Limitations

The findings of the current study must be understood in light of its limitations. Generalizability of the results is limited as the current study took place in two rural, ethnically diverse, and socioeconomically disadvantaged communities in the Southeastern United States. Although there is high external

validity to the communities in which the study took place, caution is warranted when generalizing beyond the study context. In addition, the version of the Violent Behavior Checklist used in the current study is a shortened, slightly modified version of the original measure. While the findings of this study are certainly relevant for researchers using the Violent Behavior Checklist, these analyses should be replicated using the full version of the measure. Finally, the Violent Behavior Checklist is only one of many instruments used to measure violent behavior. Additional research is needed to assess the measurement invariance of other adolescent violence measures. It would also be ideal to test the measurement invariance of measures that capture the gender of the victim of violence as this information could introduce additional differences in the way in which the measure functions. Nevertheless, the findings of the current study emphasize the need for testing measurement invariance prior to comparing violent behavior across groups.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding for this research was provided through a cooperative agreement with the U.S. Centers for Disease Control and Prevention's National Center for Injury Prevention and Control (5 U01 CE001948-03).

References

Bowen, G. L., & Richman, J. M. (2008). *The school success profile*. Chapel Hill: University of North Carolina.

Bowen, N. K., & Guo, S. (2012). *Structural equation modeling*. New York, NY: Oxford University Press.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466. doi:10.1037//0033-2909.105.3.456

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, *34*, 155-175. doi:10.1177/0022022102250225

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*, 1005-1018. doi:10.1037/a0013193

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233-255. doi:10.1207/S15328007SEM0902_5

Conrad, K. J., Riley, B. B., Conrad, K. M., Chan, Y., & Dennis, M. L. (2010). Validation of the Crime and Violence Scale (CVS) against Rasch measurement model including differences by gender, race, and age. *Evaluation Review*, *34*, 83-115. doi:10.1177/0193841X10362162

Dahlberg, L. L., Toal, S. B., Swahn, M., & Behrens, C. B. (2005). *Measuring violence-related attitudes, behaviors, and influences among youths: A compendium of assessment tools* (2nd ed.). Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control.

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage.

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Lawrence Erlbaum.

Edelen, M. O., McCaffrey, D. F., Marshall, G. N., & Jaycox, L. H. (2009). Measurement of teen dating violence attitudes: An item response theory evaluation of differential item functioning according to gender. *Journal of Interpersonal Violence*, *24*, 1243-1260. doi:10.1177/0886260508322187

Erikson, E. (1962). Youth: Fidelity and diversity. *Daedalus*, *91*, 5-27. Retrieved from http://www.jstor.org/stable/20025167

Feaster, D. J., Robbins, M. S., Henderson, C., Horigian, V., Puccinelli, M. J., Burlew, A. K., & Szapocznik, J. (2010). Equivalence of family functioning and externalizing in adolescent substance users of different race/ethnicity. *Journal of Substance Abuse Treatment*, *38S1*, S113-S124. doi:10.1016/j.jsat.2010.01.010

Frisell, T., Pawitan, Y., Langstrom, N., & Lichtenstein, P. (2012). Heritability, assertive mating, and gender differences in violent crime: Results from a total population sample using twin, adoption, and sibling models. *Behavioral Genetics*, *42*, 3-18. doi:10.1007/s10519-011-9483-0

Garnaat, S. L., & Norton, P. J. (2010). Factor structure and measurement invariance of the Yale-Brown Obsessive Compulsive Scale across four racial/ethnic groups. *Journal of Anxiety Disorders*, *24*, 723-728. doi:10.1016/j.janxdis.2010.05.004

Glanville, J. L., & Wildhagen, T. (2007). The measurement of school engagement: Assessing dimensionality and measurement invariance across race and ethnicity. *Educational and Psychological Measurement*, *67*, 1019-1041. doi:10.1177/0013164406299126

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11, Suppl. 3), S78-S94. doi:10.1097/01.mlr.0000245454.12228.8f

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55. doi:10.1080/10705519909540118

Krug, E. G., Dahlberg, L. L., Mercy, J. A., Zwi, A. B., & Lozano, R. (2002). *World report on violence and health*. Geneva, Switzerland: World Health Organization.

Miller, S., Gorman-Smith, D., Sullivan, T., Orpinas, P., & Simon, T. R. (2009). Parent and peer predictors of physical dating violence perpetration in early adolescence:

Tests of moderations and gender differences. *Journal of Clinical & Adolescent Psychology*, *38*, 538-550. doi:10.1080/15374410902976270

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Nadel, H., Spellmann, M., Alvarez-Canino, T., Lausell-Bryant, L., & Landsberg, G. (1996). The cycle of violence and victimization: A study of the school-based intervention of a multidisciplinary youth violence-prevention program. *American Journal of Preventive Medicine*, *12*, 109-119. Retrieved from http://www.sciencedirect.com/science/journal/07493797

Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology*, *20*, 673-716. doi:10.1017/S095457908000333

Peterson, D., Esbensen, F., Taylor, T. J., & Freng, A. (2007). Youth violence in context: The roles of sex, race, and community in offending. *Youth Violence and Juvenile Justice*, *5*, 385-410. doi:10.1177/1541204006297369

Thombs, B. D., Lewis, C., Bernstein, D. P., Medrano, M. A., & Hatch, J. P. (2007). An evaluation of the measurement equivalence of the Childhood Trauma Questionnaire—Short Form across gender and race in a sample of drug-abusing adults. *Journal of Psychosomatic Research*, *63*, 391-398. doi:10.1016/j.jpsychores.2007.04.010

Topitzes, J., Mersky, J. P., & Reynolds, A. J. (2012). From childhood maltreatment to violent offending: An examination of mixed-gender and gender-specific models. *Journal of Interpersonal Violence*, *27*, 2322-2347. doi:10.1177/0886260511433510

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment Research & Evaluation*, *12*(3), 1-26. Available from http://pareonline.net/

Zheng, Y., & Cleveland, H. H. (2013). Identifying gender-specific developmental trajectories of nonviolent and violent delinquency from adolescence to young adulthood. *Journal of Adolescence*, *36*, 371-381. doi:10.1016/j.adolescence.2012.12.007

Author Biographies

Katie L. Cotter, PhD, is an assitant professor at Arizona State University, School of Social Work. Her major research interests are youth violence and risk and protective factors associated with adolescent development.

Caroline B. R. Evans, PhD, is a research associate with the North Carolina Academic Center for Excellence in Youth Violence Prevention.. Her major research interests are the ecological risk and promotive factors associated with bullying and victimization.

Paul R. Smokowski, PhD, is Dean of the University of Kansas School of Social Welfare and director, North Carolina Academic Center for Excellence in Youth Violence Prevention. His major research interests are youth violence prevention and resilience in ethnic minority children and adolescents.