**Structural Validity of the HBSC Bullying Measure:   
Self-Report Rating Scales of Youth Victimization and Perpetration Behavior**

The issue of youth bullying involvement has received a great deal of research and media attention over the past few decades (Olweus, 2010) and remains a salient topic today (Hymel & Swearer, 2015). Part of the draw to study this phenomenon is the well-established link between bullying involvement and the numerous deleterious effects on youths’ social, academic, and psychological health (Jimerson, Nickerson, Mayer, & Furlong, 2012). For example, students who bully others—referred to hereafter as *perpetrators*—show more aggressive and dominant behavior patterns, exhibit fewer prosocial behaviors, and have elevated rates of externalizing behavior problems (Cook, Williams, Guerra, Kim, & Sadek, 2010; Veenstra, Lindberg, Oldehinkel, De Winter, & Ormel, 2005). On the other hand, students who are the targets of bullying—referred to hereafter as *victims*—show higher levels of internalizing behavior problems and are markedly less likely to attend and perform well in school (Cook et al., 2010; Haynie et al., 2001; Kochenderfer & Ladd, 1996; Swearer, Espelage, Vaillancourt, & Hymel, 2010).

The many negative effects associated with bullying involvement have been found across demographics, including sex, race/ethnicity, school grade-level, and nationality (Nansel, Overpeck, Pilla, Ruan, Simons-Morton, & Scheidt, 2001; Nansel, Craig, Overpeck, Saluja, & Ruan, 2004; Renshaw, Roberson, & Hammons, 2016). At the same time, some longitudinal studies of prevalence rates have shown that bullying involvement has been decreasing in the U.S. over the last several years, perhaps due in part to increased awareness of this public health issue and rise in anti-bullying legislation (e.g., stopbullying.gov). Though this may suggest reason for optimism, some estimated rates of bullying involvement are nonetheless intolerably high, with certain studies suggesting as many as 60.4% of youth endorsing involvement as a perpetrator and/or victim to some degree (Renshaw, Hammons, & Roberson, 2016).

Due to the elevated risks connected with bullying involvement, it is an important precondition to prevention and intervention that researchers and school professionals use technically adequate instruments to measure youths’ perpetration (i.e., bullying behavior directed toward others) and victimization (i.e., bullying behavior received from others). While some recent meta-analytic reviews of instruments for assessing bullying involvement show that more than 40 measures are currently available (Vessey, Strout, DiFazio, & Walker, 2014), these studies indicate major discrepancies in conceptual approaches to bullying measurement as well as varied evidence quality in regards to measurement validity (Vivolo-Kantor, Martell, Holland, & Westby, 2014). This variability in validity evidence makes it difficult to reach consensus about the nature and scope of youths’ bullying involvement (e.g., Renshaw, Hammons, & Roberson, 2016). Furthermore, these issues pose a challenge for practitioners in selecting appropriate assessment tools for screening or progress monitoring intervention and prevention efforts (Bradshaw, 2015). To remedy these issues, additional studies aimed at developing conceptually coherent and psychometrically robust measures of youths’ bullying victimization and perpetration are essential.

Though no single instrument has been adequately researched, Vessey et al. (2014) have suggested that the revisedOlweus Bully/Victim Questionnaire (OBVQ; Olweus, 1996) is perhaps the most technically adequate and commonly used bullying measure to date. While some other self-report bullying scales used for general prevalence estimation utilize only a scarce few items––such as the two items concerning school victimization in the Youth Risk Behavior Surveillance Survey (YRBSS; Center for Disease Control and Prevention, 2014) or the single victimization item in the National Crime Victimization Survey (NCVS; Bureau of Justice Statistics, 2015)––the OBVQ assesses both perpetration and victimization using multiple items targeting many classes of bullying behavior (i.e., verbal, physical, relational, racial, cyber, and general). This breadth makes the OBVQ one of the most conceptually broad instruments available.

In addition, given the evidence that both bullying perpetration and victimization are associated with overlapping yet distinct negative outcomes (Nansel et al., 2004; Renshaw, Roberson, & Hammons, 2016), the ability to derive distinct scores for both perpetration and victimization is an important feature of an instrument that intends to enhance the precision of risk estimates. Assessing bullying perpetration in conjunction with victimization also allows for the identification of students who experience both types of bullying (i.e., perpetrator–victims), whom research has identified as having even poorer outcomes compared to students involved in bullying as perpetrators-only or victims-only (Cook et al., 2010; Farmer et al., 2010; Veenstra et al., 2005). Although several existing measures use single, domain-general items to assess overall victimization and perpetration behavior—like those on the YRBSS or NCVS––research suggests that items targeting specific bullying behaviors yield higher prevalence rates, suggesting improved specificity (Renshaw, Hammons, & Roberson, 2016). Furthermore, it seems reasonable to suggest that measures composed of items targeting several specific bullying behaviors would have both greater usability (see Glover & Albers, 2007) and treatment utility (see Hayes, Nelson, & Jarrett, 1987) for practitioners using data obtained from these instruments to inform bullying prevention and intervention in schools.

Given the context sketched above, the purpose of the present study was to investigate the structural validity of an under-researched measure that targets both specific bullying victimization and perpetration behaviors: the self-report bullying items within the Health Behavior in School-aged Children survey (HBSC), sponsored by the World Health Organization (2014). These HBSC self-report items were modeled after the OBVQ and expanded in content to target parallel classes of bullying victimization and perpetration behaviors (e.g., physical victimization and physical perpetration). Although several studies investigating bullying involvement have been conducted using isolated items from HBSC self-report surveys (e.g., Renshaw, Hammons, & Roberson, 2016; Nansel et al., 2004), no study has yet to investigate the structural validity of the entire suite of bullying-related items. Thus, the purpose of the present study was to test the latent structure and invariance of responses to the HBSC self-report bullying items. Given that an item-content evaluation suggested the HBSC self-report survey targets at least two overarching classes of bullying behavior­—victimization and perpetration—the following hypotheses were proposed to guide the study:

1. Exploratory factor analysis (EFA) of responses to the HBSC bullying items will yield two distinct latent factors: *victimization* and *perpetration*.
2. Confirmatory factor analysis (CFA) of responses to this same set of items will corroborate the two-factor measurement model yielded by EFA.
3. Multigroup CFA of responses to the two-factor measurement model will demonstrate invariance across several key student demographic factors: grade-level, sex, and race/ethnicity.

**Method**

**Participants**

The current study utilized the publicly-available 2009–2010 HBSC self-report survey dataset, which is composed of a stratified random sample of U.S. youth. A detailed description of the student sampling process and other information concerning the design and administration of the survey can be found in the publicly-available codebook (Iannotti, 2013). A summary of participant demographic characteristics for the original full sample (*N* = 11,449), the full sample with missing data removed (*N* = 9,979), and two random split-half subsamples derived for data analytic purposes is provided in Table 1. Participants in the original full sample identified approximately equally as male and female, were enrolled in grades 5–10 (grade-level representation range = 13.7–20.3%), and approximated racial/ethnic proportions of U.S. youth, with the most prevalent identities being White (49.6%), Hispanic/Latino/a (17.7%), and Black/African-American (15.3%). Most respondents indicated living in a suburban area (36.2%), followed by urban (30.4%), rural (25.7%), and unclassified areas (7.8%) least frequently. Respondent scores on a composite measure of family affluence arranged from 0 = *Low* to 9 = *High* were mostly in the upper-middle range of the scale (*M* = 5.96, *Median* = 6, *Q1* = 5, *Q3* = 7). Compared to the original full sample, relative proportions of each demographic were largely maintained across all other iterations of the sample (see Table 1).

**Measure**

The HBSC self-report survey is administered to a nationally-representative sample of U.S. students in grades 5–10 every four years. The survey assesses a broad range of health-related behaviors among youth, including drug/alcohol use, body image, attitudes about school, peer/family relationships, physical health, bullying involvement, and more. To operationalize “bullying” for respondents, a prompt with a formal definition––adapted from the OBVQ—was provided at the beginning of the item set:

We say a student is BEING BULLIED when another student, or a group of students, say or do nasty or unpleasant things to him or her. It is also bullying when a student is teased repeatedly in a way her or she does not like or when they are deliberately left out of things. But it is NOT BULLYING when students of about the same strength or power argue or fight. It is also not bullying when a student is teased in a friendly and playful way.

The HBSC self-report survey contains 22 items targeting specific classes of bullying behaviors––11 items for victimization (see Table 2) and 11 parallel items for perpetration (see Table 3)––covering teasing, social exclusion, physical aggression, spreading lies, various kinds of harassment, and multiple kinds of cyber bullying. The victimization and perpetration items were arranged in two item subsets and were prefaced with a similar item stem (i.e., *How often have you [been bullied/bullied another students(s)] at this school in the past couple of months in the ways listed below?*). Response options for all bullying involvement items were arranged along a five-point, relative frequency-based scale (1 = *I haven’t been bullied at school the past couple of months*,2 = *It has only happened once or twice*,3 = *2 or 3 times a month*, 4 = *About once a week*, 5 = *Several times a week*).

**Data Analyses**

The structural validity of responses to the HBSC self-report bullying involvement items was analyzed in multiple stages using the R statistical environment (R Core Team, 2016). Prior to the primary analyses, data inspection and cleaning procedures were undertaken, beginning with handling missing data. First, 647 cases (5.7%) from the original full sample were deleted list-wise as they had missing data across all bullying items of interest, suggesting respondents failed to complete this section of the survey. Of the remaining cases, 823 (7.2%) had partially missing data for the relevant variables and were also deleted list-wise. Because the number of cases with partially missing responses accounted for less than 10% of the sample, list-wise deletion methods were deemed acceptable and unlikely to significantly bias statistical estimates (Langkamp, Lehman, & Lemeshow, 2010). The remaining cases were then divided into two random split-halves for analytic purposes. Subsample one (S1; *N* = 4,989) was used to conduct EFA, whereas subsample two (S2; *N* = 4,990) was used to conduct CFA and measurement and structural invariance analyses. Descriptive statistics were calculated for both subsamples to investigate the distributionality of the target items and are provided in Table 4. Skewness and kurtosis estimates consistently were > |2.0| for all items, indicating significant non-normality of the distribution. Due to the relative-frequency nature of the response options for the bullying items, responses were treated as ordered categorical data for all primary analyses.

EFA were conducted using the psych (Revelle, 2016) and nFactors (Raiche, 2010) R packages. Rather than analyzing the more traditional Pearson correlation matrix, the polychoric correlation matrix was analyzed, as simulation studies have demonstrated the superiority of this approach in estimating the population Pearson values when data are ordinal (Holgado-Tello, Chacón–Moscoso, Barbero–García, & Vila–Abad, 2010). Given the non-normality of the data, a principal axis factoring extraction method was selected, as it is more robust to violations of normality than maximum likelihood. To enhance factor interpretation and account for association among extracted factors, a promax factor rotation method was employed. The number of factors to retain for the preferred measurement model was decided based on an array of analyses and decision rules, starting with scree plot visualization of eigenvalues and parallel analysis (Garrido, Abad, & Ponsoda, 2013). Both analyses indicated that a two-factor solution was the best description of these data. In line with this result, a two-factor solution was constrained for the first model but additional three- and four-factor solutions were also tested to explore the potential of alternative structures. Statistical assumptions of the EFA were evaluated in terms of sampling adequacy (Kaiser-Meyer-Olkin [KMO] > .90), sphericity (χ2 *p* < .05), and extracted item communalities (*h*2 ≥ .50). Additionally, rotated pattern matrix output for each EFA was inspected for strong factor loadings ≥ 0.3, small cross-factor loadings < 0.3, and theoretical interpretability.

CFA and measurement and structural invariance analyses were conducted using the lavaan (Rosseel, 2012), semTools (semTools Contributors, 2016), and psych (Revelle, 2016) R packages. Following all EFA, the preferred measurement model was analyzed with a CFA using a robust diagonal least squares estimator with mean- and variance-adjustment (i.e., WLSMV (Muthén & Kaplan, 1985; DiStefano & Morgan, 2014). Standardized factor loadings (λ) ≥ .50 were considered adequately large as this suggests the latent factor accounts for at least 25% of the variance in the item response. Data–model fit was assessed using a variety of fit indices including scaled-shifted model chi-square (χ2; Asparouhov & Muthén, 2010), comparative fit index (CFI; Brosseau-Liard & Savalei, 2014), and root mean square error of approximation (RMSEA; Brousseau-Liard, Savalei, & Li, 2012), as well as the standardized root mean square residual (SRMR). The following criterion values were used to indicate at least adequate data–model fit: χ2 with an associated *p* > .05 (Kline, 2016), CFI ≥ .90, RMSEA with 90% confidence interval ≤ .08, and SRMR ≤ .08 (Hu & Bentler, 1999). Of these, the χ2 index is perhaps the least useful for determining fit quality, as it is highly susceptible to bias from non-normality when the sample size is large, as in the present study, but was nonetheless reported for completeness. Additionally, it is noteworthy that research supporting the use of these decision rules has mostly been conducted using maximum likelihood estimation methods with continuous data. However, some recent studies have suggested that these fit criteria hold up reasonably well when WLSMV estimation is used with non-normal ordinal data when sample sizes are large (Zhao, 2015). Finally, latent construct reliability was evaluated using the *H* statistic recommended by Mueller & Hancock (2008), where values ≥ .70 were considered adequate.

Once a good-fitting CFA model was identified, multigroup CFA was then conducted with the same subsample (S2) to investigate measurement and structural invariance across three demographic factors: grade-level, sex, and race/ethnicity. Tests of measurement and structural invariance involved adding increasingly restrictive parameter constraints to the model. The order of invariance constraint level proceeded as follows: (a) *configural*/*baseline* (equal latent structure), (b) *weak* (adds equal factor loadings), (c) *strong* (adds equal intercepts), (d) *homogeneity of latent variance* (adds equal residuals and latent variances), and (e) *homogeneity of latent means* (adds equal latent means). Levels (a) through (c) refer properly to measurement invariance while levels (d) and (e) concern structural invariance (Vandenberg & Lance, 2000). Relative invariance from less to more constrained models was determined based on change in CFI ≤ .01 (Vandenberg & Lance, 2000). However, it is noteworthy that little research is available offering guidance on invariance evaluation criteria when using a WLSMV estimator for non-normal ordinal data. Another common test for invariance is the change in model χ2 from less to more restrictive models. Although, because there is disagreement in the literature regarding the appropriateness of this test when using WLSMV (cf. French & Finch, 2008; Asparouhov & Muthén, 2006) we have decided to forego this approach in favor of the change in CFI for the current investigation. Invariance results should therefore be interpreted with caution. Finally, once an appropriate factor structure was established with S2, subscale scores were calculated at the observed level to investigate their descriptive properties.

**Results**

**Latent Factor Structure**

**EFA (S1).** Results of the first analysis, which constrained a two-factor solution, showed strong KMO sampling adequacy (.94), lack of singularity—χ2(231) = 79308, *p* < .001—and adequate extracted item communalities (*h*2 range = .56–.94). Table 5 summarizes the factor loading output and psychometric characteristics of this measurement model. Eigenvalues for the two factors collectively accounted for 75.8% of the variance in these data. Pattern matrix factor loadings for all bullying victimization (λ range = .74­­–.89) and bullying perpetration (λ range = .77–.95) items were greater than the minimum threshold of .30, with no significant cross-loadings. Additionally, observed internal consistency estimates were strong for both factors (α > .70) while also showing a strong latent factor correlation (ϕ = .53).

Next, a second EFA constraining a three-factor solution was tested. Results from this model showed strong communalities among all items (*h*2 range = .56–.94) and the three factors collectively accounted for 80.5% of the variance. Inspection of the pattern matrix revealed the three factors consisted of (a) 11 items related to perpetration (λ range = .82–.93), (b) seven items related to traditional (i.e., non-cyber) victimization (λ range = .55–1.00­­), and (c) four items pertaining to cyber victimization (λ range = .69–.72). Although these loadings were strong, interpretation of the factors was not as clear as with the two-factor model, given that five items cross-loaded on other factors: pVerbal, pCompOut, pCellOut, vReligious, and vComp (see Table 2 and Table 3 for item content).

Finally, a four-factor EFA solution was tested. As with both previous models, the four-factor solution showed uniformly strong communalities (*h*2 range = .61–.96) and collectively accounted for 82.4% of the variance. Pattern matrix inspection revealed that the first three extracted factors represented the same factors as in the three-factor model—traditional victimization (λ range = .55–.96), perpetration (λ range = .82–.95), and cyber victimization (λ range = .72–.77)—plus a fourth factor where only the items vRacial (λ = .50) and vReligious (λ = .47) had factor loadings above .30. This model also showed three items with cross-loadings > .30: pCellOut, vRacial, and vReligious (see Table 2 and Table 3 for item content).

Considering these EFA results together, the preponderance of evidence converged in support of retaining the two-factor structure as the preferred measurement model for subsequent CFA and invariance analyses.

**CFA (S2).** To corroborate the two-factor measurement model indicated by EFA, CFA was conducted with S2 by regressing each of the 22 observed bullying items onto their respective latent factors of *perpetration* and *victimization* and allowing both latent factors to freely covary. Findings showed the model was characterized by strong latent factor loadings for perpetration (λ range = .72–.97) and victimization (λ range = .74–.97), as well as a large covariance between the factors (ϕ= .69). Model fit indices also indicated strong data-model fit—χ2 (208) = 3174.054, *p* < .001; CFI = .971; RMSEA = .053 [90% CI = .052–.055]; SRMR = .064 (see Figure 1). Additionally, both latent factors had strong internal reliability and average variance extracted (AVE): victimization *H* = .98 and AVE = .72, perpetration *H* = .99 and AVE = .82.

**Measurement and structural invariance analysis (S2).** Results from the multigroup measurement and structural invariance analyses showed good data–model fit at all levels of constraint across student race/ethnicity, grade-level, and sex (see Table 6 and Table 7). ∆CFI was uniformly < .01 for every level of invariance across all demographics. Additionally, the strictest invariance level, homogeneity of latent means, was the best fitting model for all three demographics, providing further evidence for measurement and structural invariance across groups.

**Observed subscale descriptive statistics (S2).** In addition to the latent-level analyses described above, basic descriptive statistics were also calculated at the observed level for total scores from the victimization and perpetration subscales (see Table 7). Results indicated that these scores were characterized by strong internal consistency (α > .90), and moderate (victimization *r* = .47) and large (perpetration *r* = .60) average inter-item correlations. Additionally, the scores showed significant non-normality (> |2.0|) and relatively-low mean bullying involvement frequency for both perpetration (*M* = 12.18) and victimization (*M* = 14.15).

**Discussion**

Although several instruments assessing youths’ bullying involvement are currently available, to date, none appear as conceptually broad or as balanced across victimization and perpetration domains as the measure included within the HBSC self-report survey. The present study was the first to investigate the structural validity of responses to the 22 items that make up this measure, using the publicly-available, representative sample of U.S. youth in grades 5–10 from the 2009–2010 administration of the survey. We hypothesized that results from factor analyses would support a two-factor measurement model that represented bullying victimization and perpetration, and that this model would be invariant across key demographics. After testing the tenability of two-, three-, and four-factor EFA models, the results converged on the conclusion that the hypothesized two-factor model of bullying involvement was in fact the best structuredescribed by the data. CFA results provided additional evidence in support of this two-factor measurement model, showing strong data-model fit evidenced by a variety of absolute and incremental fit indices. Furthermore, findings from the follow-up multigroup CFA supported the measurement and structural invariance of responses across student sex, race/ethnicity, and grade-level. However, due to insufficient subgroup sample sizes for several racial/ethnic groups, invariance for race/ethnicity was only established for the three largest populations in the U.S. (i.e., White, Hispanic/Latino/a, and Black/African American students) and should therefore not be generalized beyond these groups based on this study alone.

Taken together, findings from the present study largely supported our hypotheses and suggest that the HBSC self-report bullying items can be considered structurally valid for measuring students’ bullying perpetration and victimization. Given that strong measurement and structural invariance evidence was derived from a representative sample of U.S. students, no adjustments to scoring or changes to administration procedures appear necessary when using the measure to assess bullying involvement among 5th–10th graders identifying as White, Hispanic/Latino/a, or Black/African American. Furthermore, because this study utilized such a large and demographically diverse dataset, the descriptive statistics derived from the observed subscales scores (see Table 8) could also be used as normative data for informing future research or practice, such as for identifying students at-risk for bullying involvement via universal screening or for progress monitoring the effectiveness of prevention or intervention efforts at the level of classrooms, grades, schools, or districts.

While the victimization and perpetration scales could be used in isolation, we recommend the use of both scales in conjunction when possible to identify the at least these three involvement groups with distinct outcome trajectories (i.e., victims, perpetrators, perpetrator-victims). Additionally, it is noteworthy that the covariance between the perpetration and victimization latent variables was large (see Figure 1), suggesting that there is a considerable rate of perpetrator-victims in our sample who are at-risk for the most detrimental outcomes. Although a thorough analysis of prevalence rates is beyond the scope of this measure development study, interested readers are directed to Renshaw, Roberson, & Hammons (2016) and Renshaw, Hammons, & Roberson (2016) for more detailed investigation of prevalence rates among different involvement group classifications using the 2009–2010 HBSC sample.

Results from this study should be considered in light of a few notable limitations. As mentioned earlier, the research supporting the use of the WLSMV estimator for non-normal ordinal data under various model and sample size conditions is still nascent. Although the considerably large sample size might allay some concern about the functionality of the estimates, substantive conclusions drawn from these results should be considered tentative. It is also not known to what extent the descriptive results presented herein (see Table 7) reflect contemporary bullying involvement among U.S. youth, as we are currently six years removed from the responses of participants from the 2009–2010 sample used in this study. Additionally, as an artifact of splitting the full sample into random halves, invariance analyses could only be conducted for the three most prevalent racial/ethnic groups, which limits the generalizability of findings for most ethnic minority groups within the U.S. Future research is therefore warranted to replicate these invariance analyses with the full sample from the next available HBSC administration, which should allow for the inclusion of more ethnic minority groups.

Future research investigating the concurrent and predictive validity of responses to the HBSC bullying measure with other physical, psychological, and academic wellbeing indicators is also encouraged. This would allow for a greater understanding of which classes of bullying behaviors represented by this measurement model are salient indicators of risk. Relatedly, this measure could also be co-validated with pre-existing bullying instruments to test the comparative validity of responses to competing measures for predicting valued student outcomes. Future studies within this line of research might also benefit from employing other data collection methodologies besides self-report (e.g., school discipline records, informant report ratings), as it will alleviate the potential bias in relying on a single data collection approach (see Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Similarly, randomizing the order in which items are presented may also help curb biased responding, as some evidence suggests respondents may be primed to respond differently depending on how previous questions were phrased (Huang & Cornell, 2015). Finally, in addition to basic psychometric validation studies, research investigating the treatment utility (see Hayes et al., 1987) of scores derived from the HBSC bullying measure is needed to inform use of the instrument for prevention and intervention purposes.

References

Asparouhov, T., & Muthén, B. (2006). Robust chi square difference testing with mean and variance adjusted test statistics. Retrieved from www.statmodel.com/download/webnotes/webnote10.pdf

Asparouhov, T., & Muthén, B. (2010). Simple second order chi-square correction. Retrieved from www. statjnodel. com/dowrJoad/WLSMV\_new\_chi2, 1

Bureau of Justice Statistics. (2015). *Indicators of school crime and safety*. Washington, DC: Bureau of Justice Statistics. Retrieved from www.bjs.gov/index.cfm?ty=pbse&sid=8

Bradshaw, C. P. (2015). Translating research to practice in bullying prevention. *American Psychologist*, *70*, 322–332. doi:10.1037/a0039114

Brosseau-Liard, P. E., & Savalei, V. (2014). Adjusting incremental fit indices for nonnormality. *Multivariate Behavioral Research*, *49*, 460–470. doi:10.1080/00273171.2014.933697

Brosseau-Liard, P. E., Savalei, V., & Li, L. (2012). An investigation of the sample performance of two nonnormality corrections for RMSEA. *Multivariate Behavioral Research*, *47*, 904–930. doi:10.1080/00273171.2012.715252

Centers for Disease Control and Prevention. (2014). *Youth Risk Behavior Surveillance System* *(YRBSS).* Retrieved from http://www.cdc.gov/mmwr/pdf/ss/ss6304.pdf

Cook, C. R., Williams, K. R., Guerra, N. G., Kim, T. E., & Sadek, S. (2010). Predictors of bullying and victimization in childhood and adolescence: A meta-analytic investigation. *School Psychology Quarterly*, *25*, 65–83. doi:10.1037/a0020149.

DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 425–438. doi:10.1080/10705511.2014.915373

Farmer, T. W., Petrin, R. A., Robertson, D. L., Fraser, M. W., Hall, C. M., Day, S. H., & Dadisman, K. (2010). Peer relations of bullies, bully-victims, and victims: The two social worlds of bullying in second-grade classrooms. *Elementary School Journal*, *110*, 364–392. doi:10.1086/648983

French, B. F., & Finch, W. H. (2008). Multigroup confirmatory factor analysis: Locating the invariant referent sets. *Structural Equation Modeling*, *15*, 96–113. doi: 10.1080/10705510701758349

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn’s parallel analysis with ordinal variables. *Psychological Methods*, *18*, 454–474. doi:10.1037/a0030005

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*, 117–135. doi:10.1016/j.jsp.2006.05.005

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist*, *42*, 963–974. doi:10.1037/0003-066X.42.11.963

Haynie, D. L., Nansel, T., Eitel, P., Crump, A. D., Saylor, K., Yu, K., & Simons-Morton, B. (2001). Bullies, victims, and bully/victims: Distinct groups of at-risk youth. *The Journal of Early Adolescence*, *21*, 29–49. doi:10.1177/0272431601021001002

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–García, I., & Vila–Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, *44*, 153–166. doi:10.1007/s11135-008-9190-y

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

Huang, F. L., & Cornell, D. G. (2015). The impact of definition and question order on the prevalence of bullying victimization using student self-reports. *Psychological Assessment*, *27*, 1484–1493. doi:10.1037/pas0000149

Hymel, S., & Swearer, S. M. (2015). Four decades of research on school bullying: An Introduction. *American Psychologist*, *70*, 293–299. doi:10.1037/a0038928

Iannotti, R. J. (2013). *Health behavior in school-aged children (HBSC), 2009–2010* *(ICPSR34792-v1).* Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Retrieved from http://doi.org/10.3886/ICPSR34792.v1

Jimerson, S. R., Nickerson, A. B., Mayer, M. J., & Furlong, M. J. (Eds.) (2012). *Handbook of school violence and school safety: International research and practice* (2nd ed.). New York, NY: Routledge.

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. New York, NY: Guilford.

Kochenderfer, B. J., & Ladd, G. W. (1996). Peer victimization: Cause of consequence of school maladjustment? *Child Development*, *67*, 1305–1317. doi:10.1111/j.1467-8624.1996.tb01797.x

Langkamp, D. L., Lehman, A., & Lemeshow, S. (2010). Techniques for handling missing data in secondary analyses of large surveys. *Academic Pediatrics*, *10*, 205–210. doi:10.1016/j.acap.2010.01.005.

Mueller, R. O., & Hancock, G. R. (2008). Best practices in structural equation modeling. In J. Osborne (Ed.), B*est practices in quantitative methods* (pp. 488–508). Thousand Oaks, CA: Sage. doi:10.4135/9781412995627.d38

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non‐normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171–189.

Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment. *The Journal of the American Medical Association*, *285*, 2094–2100. doi:10.1001/jama.285.16.2094.

Nansel, T. R., Craig, W., Overpeck, M. D., Saluja, G., & Ruan, W. J. (2004). Cross-national consistency in the relationship between bullying behaviors and psychosocial adjustment. *Archives of Pediatrics & Adolescent Medicine*, *158*, 730–736. doi:10.1001/archpedi.158.8.730

Olweus, D. (1996). *The Revised Olweus Bully/Victim Questionnaire*. Bergen, Norway: University of Bergen, Research Center for Health Promotion.

Olweus, D. (2010). Understanding and researching bullying: Critical issues. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 9–34)*.* New York, NY: Routledge.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879–903. doi:10.1037/0021-9010.88.5.879

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Raiche, G. (2010). nFactors: An R package for parallel analysis and non-graphical solutions to the Cattell scree test. R package version 2.3.3.

Renshaw, T. L., Hammons, K. N., Roberson, A. J. (2016). General vs. specific methods for classifying U.S. students’ bullying involvement: Investigating classification agreement, prevalence rates, and concurrent validity. *School Psychology Review*, *45*(4).

Renshaw, T. L., Roberson, A. J., & Hammons, K. N. (2016). The functionality of four bullying involvement classification schemas: Prevalence rates and associations with mental health and school outcomes. *School Mental Health, 8*, 332–343*.* doi:10.1007/s12310-015-9171-y

Revelle, W. (2016) psych: Procedures for personality and psychological research. Northwestern University, Evanston, Illinois, USA. https://CRAN.R-project.org/package=psych Version = 1.6.9.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. Retrieved from http://www.jstatsoft.org/v48/i02/.

semTools Contributors. (2016). semTools: Useful tools for structural equation modeling. R package version 0.4-14. Retrieved from https://CRAN.R-project.org/package=semTools

Swearer, S. M., Espelage, D. L., Vaillancourt, T., & Hymel, S. (2010). What can be done about school bullying? Linking research to educational practice. *Educational Researcher*, *39*, 38–47. doi:10.3102/0013189X09357622

Veenstra, R., Lindenberg, S., Oldehinkel, A. J., De Winter, A. F., & Ormel, J. (2005). Bullying and victimization in elementary schools: A comparison of bullies, victims, and bully/victims, and uninvolved preadolescents. *Developmental Psychology*, *41*, 672–682. doi:10.1037/0012-1649.41.4.672

Vessey, J., Strout, T. D., DiFazio, R. L., & Walker, A. (2014). Measuring the youth bullying experience: A systematic review of the psychometric properties of available instruments. *Journal of School Health*, *84*, 819­–843. doi:10.1111/josh.12210

Vivolo-Kantor, A. M., Martell, B. N., Holland, K. M., & Westby, R. (2014). A systematic review and content analysis of bullying and cyber-bullying measurement strategies. *Aggression and Violent Behavior*, *19*, 423–434. doi:10.1016/j.avb.2014.06.008

World Health Organization. (2014). Health behavior in school-aged children: World Health Organization collaborative cross-national survey. Retrieved from http://www.who.int/en

Zhao, Y. (2015). The performance of model fit measures by robust weighted least squares estimators in Confirmatory Factor Analysis (Doctoral dissertation, The Pennsylvania State University).

Table 1

*2009–2010 HBSC Demographic Proportions for Sex, Grade-level, and Race/Ethnicity across all Samples*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Full Sample | | Full Sample after Removing Cases | | Subsample 1 | | Subsample 2 | |
|  | *N* | *%* | *N* | % | *N* | % | *N* | % |
| *Total Sample* | 11449 | 100 | 9979 | 100 | 4989 | 100 | 4990 | 100 |
| *Sex* |  |  |  |  |  |  |  |  |
| Male | 5883 | 51.4 | 5048 | 50.6 | 2518 | 50.5 | 2530 | 50.7 |
| Female | 5562 | 48.6 | 4930 | 49.4 | 2470 | 49.5 | 2460 | 49.3 |
| Missing Data | 4 | 0.0 | 1 | 0.0 | 1 | 0.0 | 0 | 0.0 |
| *Grade-level* |  |  |  |  |  |  |  |  |
| 5th | 1574 | 13.7 | 1207 | 12.1 | 617 | 12.4 | 590 | 11.8 |
| 6th | 1781 | 15.6 | 1524 | 15.3 | 750 | 15.0 | 774 | 15.5 |
| 7th | 2269 | 19.8 | 1963 | 19.7 | 966 | 19.4 | 997 | 20.0 |
| 8th | 2327 | 20.3 | 2061 | 20.7 | 1010 | 20.2 | 1051 | 21.1 |
| 9th | 1817 | 15.9 | 1662 | 16.7 | 855 | 17.1 | 807 | 16.2 |
| 10th | 1681 | 14.7 | 1562 | 15.7 | 791 | 15.9 | 771 | 15.5 |
| *Race/Ethnicity* |  |  |  |  |  |  |  |  |
| American Indian/ Alaskan Native | 206 | 1.8 | 176 | 1.8 | 83 | 1.7 | 93 | 1.9 |
| Asian | 448 | 3.9 | 406 | 4.1 | 194 | 3.9 | 212 | 4.2 |
| Black/African American | 1841 | 16.1 | 1522 | 15.3 | 771 | 15.5 | 751 | 15.1 |
| Hispanic/Latino/a | 2097 | 18.3 | 1768 | 17.7 | 861 | 17.3 | 907 | 18.2 |
| Native Hawaiian/Other Pacific Islander | 105 | 0.9 | 89 | 0.9 | 48 | 1.0 | 41 | 0.8 |
| White | 5485 | 47.9 | 4953 | 49.6 | 2512 | 50.5 | 2441 | 48.9 |
| Two or More | 757 | 6.6 | 679 | 6.8 | 334 | 6.7 | 345 | 6.9 |
| Missing Data | 510 | 4.5 | 386 | 3.9 | 186 | 3.7 | 200 | 4.0 |

Table 2

*2009–2010 HBSC Items Targeting Bullying Victimization*

|  |  |
| --- | --- |
| Code | Item |
| vVerbal | “I was called mean names, was made fun of, or teased in a hurtful way.” |
| vExclusion | “Other students left me out of things on purpose, excluded me from their group of friends, or completely ignored me.” |
| vPhysical | “I was hit, kicked, pushed, shoved around, or locked indoors.” |
| vRelational | “Other students told lies or spread false rumors about me and tried to make others dislike me.” |
| vRacial | “I was bullied with mean names and comments about my race or color.” |
| vReligious | “I was bullied with mean names and comments about my religion.” |
| vSexual | “Other students made sexual jokes, comments or gestures to me.” |
| vComp | “I was bullied using a computer or e-mail messages or pictures.” |
| vCell | “I was bullied using a cell phone.” |
| vCompOut | “I was bullied outside of school using a computer or e-mail messages or pictures.” |
| vCellOut | “I was bullied outside of school using a cell phone.” |

Table 3

*2009–2010 HBSC Items Targeting Bullying Perpetration*

|  |  |
| --- | --- |
| Code | Item |
| pVerbal | “I called another student(s) mean names, made fun of, or teased him or her in a hurtful way.” |
| pExclusion | “I kept another student(s) out things on purpose, excluded him or her from my group of friends or completely ignored him or her.” |
| pPhysical | “I hit, kicked, pushed, shoved around, or locked another student(s) indoors.” |
| pRelational | “I spread false rumors about another student(s) and tried to make others dislike him or her.” |
| pRacial | “I bullied another student(s) with mean names or comments about his or her race or color.” |
| pReligious | “I bullied another student(s) with mean names or comments about his or her religion.” |
| pSexual | “I made sexual jokes, comments or gestures to another student(s).” |
| pComp | “I bullied another student(s) using a computer, e-mail messages or pictures.” |
| pCell | “I bullied another student(s) using a cell phone.” |
| pCompOut | “I bullied others outside of school using a computer, e-mail messages or pictures.” |
| pCellOut | “I bullied others outside of school using a cell phone.” |

Table 4

*Descriptive Statistics for All Bullying Items in Subsample 1 and Subsample 2*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Subsample 1 | | | | | |  | Subsample 2 | | | | | |
|  | Mean | SD | Median | Q1, Q3 | Skew. | Kurt. |  | Mean | SD | Median | Q1, Q3 | Skew. | Kurt. |
| vVerbal | 1.56 | 1.09 | 1 | 1, 2 | 2.13 | 3.59 |  | 1.56 | 1.07 | 1 | 1, 2 | 2.14 | 3.70 |
| vExclusion | 1.44 | 0.97 | 1 | 1, 1 | 2.50 | 5.57 |  | 1.44 | 0.96 | 1 | 1, 1 | 2.51 | 5.70 |
| vPhysical | 1.26 | 0.76 | 1 | 1, 1 | 3.50 | 12.26 |  | 1.25 | 0.78 | 1 | 1, 1 | 3.57 | 12.66 |
| vRelational | 1.53 | 1.03 | 1 | 1, 2 | 2.22 | 4.18 |  | 1.51 | 1.02 | 1 | 1, 2 | 2.25 | 4.37 |
| vRacial | 1.25 | 0.81 | 1 | 1, 1 | 3.62 | 12.66 |  | 1.24 | 0.78 | 1 | 1, 1 | 3.71 | 13.52 |
| vReligious | 1.17 | 0.65 | 1 | 1, 1 | 4.52 | 20.94 |  | 1.17 | 0.67 | 1 | 1, 1 | 4.54 | 20.91 |
| vSexual | 1.42 | 0.98 | 1 | 1, 1 | 2.57 | 5.77 |  | 1.42 | 0.99 | 1 | 1, 1 | 2.60 | 5.85 |
| vComp | 1.13 | 0.56 | 1 | 1, 1 | 5.25 | 28.97 |  | 1.15 | 0.61 | 1 | 1, 1 | 4.87 | 24.53 |
| vCell | 1.13 | 0.56 | 1 | 1, 1 | 5.34 | 30.11 |  | 1.13 | 0.58 | 1 | 1, 1 | 5.09 | 26.95 |
| vCompOut | 1.13 | 0.56 | 1 | 1, 1 | 5.20 | 28.65 |  | 1.14 | 0.61 | 1 | 1, 1 | 4.99 | 25.68 |
| vCellOut | 1.12 | 0.55 | 1 | 1, 1 | 5.41 | 30.75 |  | 1.14 | 0.60 | 1 | 1, 1 | 5.04 | 26.10 |
| pVerbal | 1.35 | 0.80 | 1 | 1, 1 | 2.94 | 9.19 |  | 1.37 | 0.83 | 1 | 1, 1 | 2.90 | 8.72 |
| pExclusion | 1.24 | 0.69 | 1 | 1, 1 | 3.67 | 14.70 |  | 1.26 | 0.73 | 1 | 1, 1 | 3.51 | 13.02 |
| pPhysical | 1.19 | 0.66 | 1 | 1, 1 | 4.21 | 18.61 |  | 1.19 | 0.67 | 1 | 1, 1 | 4.15 | 17.91 |
| pRelational | 1.15 | 0.59 | 1 | 1, 1 | 4.71 | 23.66 |  | 1.15 | 0.59 | 1 | 1, 1 | 4.83 | 25.04 |
| pRacial | 1.13 | 0.57 | 1 | 1, 1 | 5.17 | 28.20 |  | 1.13 | 0.58 | 1 | 1, 1 | 5.22 | 28.65 |
| pReligious | 1.11 | 0.53 | 1 | 1, 1 | 5.74 | 34.64 |  | 1.11 | 0.55 | 1 | 1, 1 | 5.66 | 33.34 |
| pSexual | 1.17 | 0.64 | 1 | 1, 1 | 4.45 | 20.65 |  | 1.17 | 0.65 | 1 | 1, 1 | 4.51 | 20.96 |
| pComp | 1.10 | 0.51 | 1 | 1, 1 | 6.03 | 38.22 |  | 1.10 | 0.51 | 1 | 1, 1 | 6.20 | 40.22 |
| pCell | 1.10 | 0.51 | 1 | 1, 1 | 5.89 | 36.69 |  | 1.10 | 0.52 | 1 | 1, 1 | 5.98 | 37.82 |
| pCompOut | 1.10 | 0.51 | 1 | 1, 1 | 5.93 | 37.08 |  | 1.11 | 0.54 | 1 | 1, 1 | 5.84 | 35.44 |
| pCellOut | 1.10 | 0.51 | 1 | 1, 1 | 6.15 | 39.88 |  | 1.11 | 0.54 | 1 | 1, 1 | 5.85 | 35.82 |

*Note.* Q1, Q3 = first and third quartile score; Skew. = Skewness; Kurt. = Kurtosis.

Table 5

*EFA Pattern Matrix Results for the Two-Factor Measurement Model*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Factor Loadings (λ) | | | |
| Item Code | | Perpetration (ξ1) | | Victimization (ξ2) | |
| pReligious | **0.946** | | 0.003 | |
| pComp | **0.944** | | 0.045 | |
| pCompOut | **0.935** | | 0.045 | |
| pRacial | **0.931** | | -0.029 | |
| pCell | **0.931** | | 0.062 | |
| pCellOut | **0.910** | | 0.065 | |
| pSexual | **0.869** | | 0.011 | |
| pPhysical | **0.862** | | -0.003 | |
| pRelational | **0.859** | | 0.084 | |
| pExclusion | **0.831** | | -0.028 | |
| pVerbal | **0.772** | | -0.042 | |
| vRelational | -0.162 | | **0.889** | |
| vVerbal | -0.234 | | **0.889** | |
| vExclusion | -0.148 | | **0.872** | |
| vPhysical | -0.001 | | **0.799** | |
| vRacial | 0.044 | | **0.782** | |
| vReligious | 0.117 | | **0.781** | |
| vSexual | -0.021 | | **0.775** | |
| vComp | 0.212 | | **0.758** | |
| vCell | 0.229 | | **0.739** | |
| vCompOut | 0.237 | | **0.738** | |
| vCellOut | 0.229 | | **0.736** | |
| Eigenvalues | | 9.36 | | 7.32 | |
| % variance | | 42.50 | | 33.30 | |
| α | | 0.94 | | 0.89 | |
| *r* | | 0.61 | | 0.47 | |

*Note.* α = Cronbach’s alpha; *r* = average inter-item correlation.

Table 6

*Model Fit of the Two-Factor CFA for Each Demographic Group Used for Multigroup Invariance Analyses*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Demographic | χ2 | *df* | RMSEA [90% CI] | CFI | SRMR | λVict. Range | λPerp. Range | St. Cov. |
| *Race/Ethnicity* |  |  |  |  |  |  |  |  |
| White | 1301.635 | 208 | .046 [.044, .049] | .974 | .067 | .74–.96 | .75–.98 | .60 |
| Hispanic/Latino/a | 782.303 | 208 | .055 [.051, .059] | .980 | .064 | .71–.97 | .71–.98 | .81 |
| Black/Af. American | 714.946 | 208 | .057 [.052, .062] | .973 | .064 | .68–.97 | .72–.97 | .79 |
| *Grade-level* |  |  |  |  |  |  |  |  |
| 5th | 430.592 | 208 | .043 [.037, .048] | .987 | .082 | .70–.99 | .78–.99 | .73 |
| 6th | 581.301 | 208 | .048 [.044, .053] | .986 | .069 | .73–.98 | .75–.99 | .77 |
| 7th | 834.013 | 208 | .055 [.051, .059] | .974 | .072 | .72–.97 | .69–.97 | .75 |
| 8th | 752.777 | 208 | .050 [.046, .054] | .974 | .071 | .70–.96 | .72–.98 | .59 |
| 9th | 549.378 | 208 | .045 [.041, .050] | .975 | .071 | .74–.97 | .74–.98 | .56 |
| 10th | 601.716 | 208 | .050 [.045, .054] | .988 | .059 | .70–.98 | .78–1.00 | .79 |
| *Sex* |  |  |  |  |  |  |  |  |
| Male | 1936.457 | 208 | .057 [.055, .060] | .975 | .071 | .74–.97 | .74–.98 | .66 |
| Female | 1469.310 | 208 | .050 [.047, .052] | .967 | .065 | .65–.97 | .73–.98 | .75 |

*Note.* χ2, RMSEA, and CFI estimates are presented with scaled-shifted corrections applied.All χ2 values were significant at the *p* < .001 level. St. Cov. = Standardized covariance between the victimization and perpetration latent variables.

Table 7

*Measurement and Structural Invariance Results for the Two-Factor Model Across Demographics*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Demographic / Invariance Level | χ2 | *df* | RMSEA [90% CI] | CFI | ΔCFI |
| *Race/Ethnicity* |  |  |  |  |  |
| Configural/Baseline | 2818.169 | 624 | .051 [.049, .053] | .974 | –– |
| Weak | 2275.960 | 664 | .042 [.040, .044] | .981 | .007 |
| Strong | 2780.884 | 792 | .043 [.041, .045] | .977 | .004 |
| Latent Variance | 2684.975 | 796 | .042 [.040, .043] | .978 | .001 |
| Latent Means | 2059.119 | 800 | .034 [.032, .036] | .985 | .007 |
| *Grade-level* |  |  |  |  |  |
| Configural/Baseline | 3646.476 | 1248 | .048 [.046, .050] | .983 | –– |
| Weak | 2996.939 | 1348 | .038 [.037, .040] | .988 | .005 |
| Strong | 3818.532 | 1668 | .039 [.038, .041] | .984 | .004 |
| Latent Variance | 3708.599 | 1678 | .038 [.037, .040] | .985 | .001 |
| Latent Means | 2736.109 | 1688 | .027 [.025, .029] | .992 | .007 |
| *Sex* |  |  |  |  |  |
| Configural/Baseline | 3401.684 | 416 | .054 [.052, .055] | .973 | –– |
| Weak | 3219.269 | 436 | .051 [.049, .052] | .974 | .001 |
| Strong | 3483.713 | 500 | .049 [.047, .050] | .974 | .000 |
| Latent Variance | 3367.200 | 502 | .048 [.046, .049] | .974 | .000 |
| Latent Means | 2369.732 | 504 | .039 [.037, .040] | .983 | .009 |

*Note.* χ2, RMSEA, and CFI estimates are presented with scaled-shifted corrections applied. All χ2 values were significant at the *p* < .001 level.

Table 8

*Observed Descriptive Statistics of the HBSC Bullying Measure Composite Scores*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subscale | Items | Min, Max | Mean | SD | Median | Q1, Q3 | Skew. | Kurt*.* | *r* | α |
| Victimization | 11 | 11, 55 | 14.15 | 6.23 | 12 | 11, 14 | 3.31 | 13.24 | .47 | .90 |
| Perpetration | 11 | 11, 55 | 12.81 | 5.30 | 11 | 11, 12 | 5.02 | 29.64 | .60 | .94 |

*Note.* Q1, Q3 = first and third quartile score; Skew. = skewness; Kurt.= kurtosis; *r* = average inter-item correlation; α = Cronbach’s alpha.

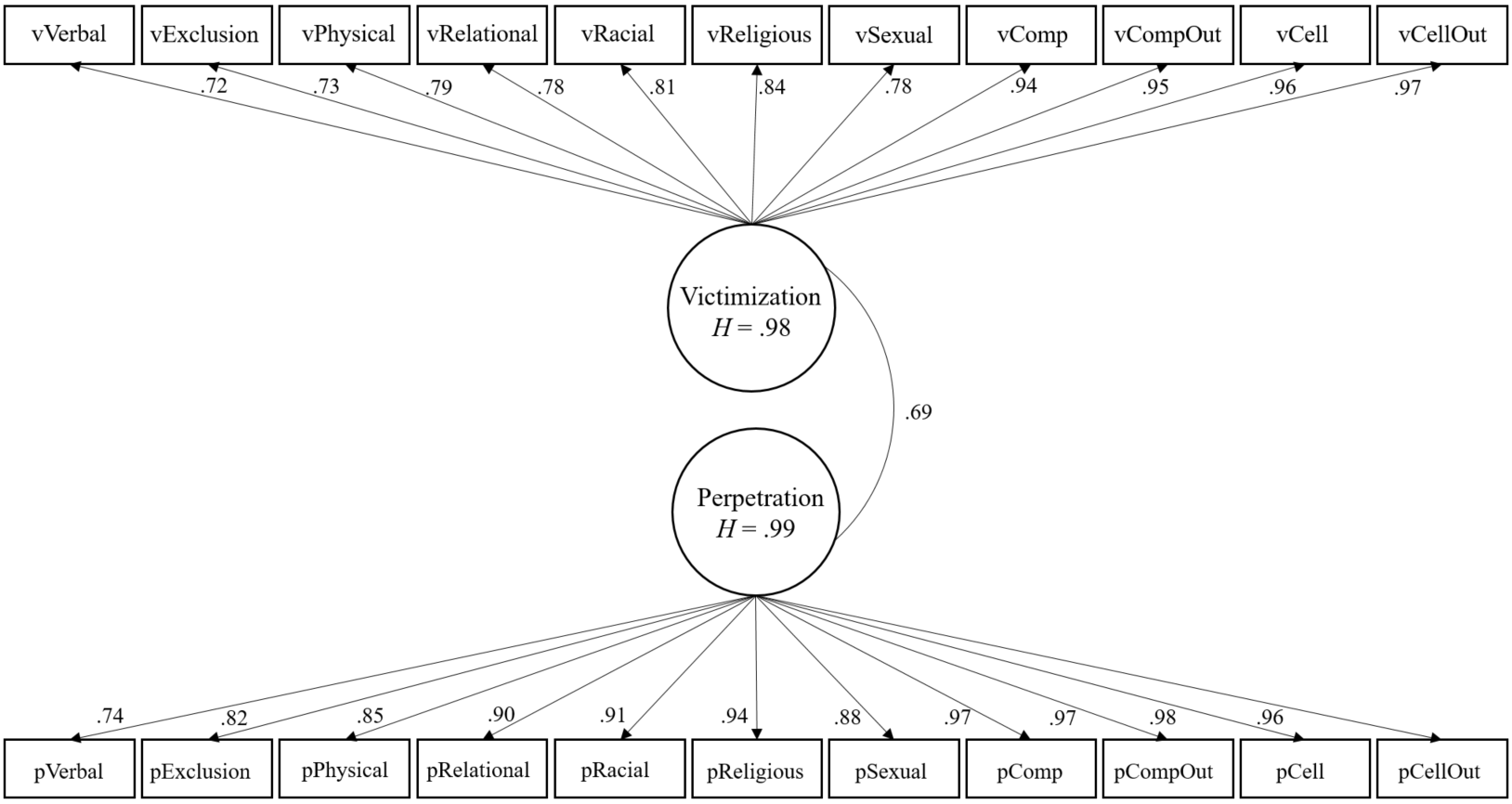
****

Figure 1. *Full structural model of the two-factor HBSC Bullying Measure.*

*Note*. The covariance and factor loading estimates are presented in standardized form.