CORPUS PUBLISHERS

Current Research in Psychology and Behavioral Science (CRPBS)

ISSN: 2833-0986

Volume 4, Issue 3, 2023

Article Information

Received date : 06 March, 2023

Published date: 20 March, 2023

*Corresponding author

Rondy Yu, School of Education, 900 University Avenue, Riverside, CA 92521, USA

Key Words

Behavior; Bullying; Measurement; School

DOI: 10.54026/CRPBS/1091

Research Article

Development and Validation of a Brief Bullying Behavior Scale: Examination of the Health Behavior in School-Aged Children (HBSC) Survey Bullying Items

Rondy Yu[1], Wesley A. Sims[1] and Shane R. Jimerson[2*]

[1]*University of California, Riverside, USA*

[2]*University of California, Santa Barbara, USA*

Abstract

Bullying in schools is a longstanding challenge that has been repeatedly linked to deleterious student outcomes. Using a nationally representative sample of students in the United States (N=12,642, grades 5 to 10), this study examined the factor structure of the Health Behavior in School-Aged Children (HBSC) self-report survey's bullying items, along with one additional item related to the occurrence of physical fighting. Results provide preliminary support for the use of 19 items from the HBSC survey as a standalone scale for measuring student involvement in bullying for both male and female students.

Introduction

The topic of bullying in schools continues to garner significant attention in both the media and research community as studies around the world continue to reveal the deleterious impacts on the development of children and adolescents [1,2]. Defined as the repeated exposure to negative behaviors by one or more students [3], bullying is understandably concerning, particularly given how pervasive the problem is and its associated consequences. For instance, in the United States, the most recent youth risk behavior surveillance report indicates nearly one-fifth of adolescents have experienced bullying at school within the last year [4]. For victims of bullying, the experience has been found to be associated with depressive symptoms, suicidal thoughts/behaviors [5], and harmful, if not violent, acts of revenge [6,7]. While negative outcomes have commonly been discussed within the context of impact on the victim, it is also clear that both victims, defenders, and perpetrators suffer from academic, social-emotional, and behavioral problems related to bullying [8,9]. Thus, it is imperative that schools pay particular attention to the early identification of those involved in bullying, regardless of their role in it [10].

Assessment and Identification

The advantages of data-driven processes for the identification of strengths and challenges in educational settings are abundant in the extant empirical studies [11]. Like universal screening [12] and progress monitoring for instructional decision-making [13], assessment of bullying in schools serves important functions at multiple levels. At the macro level, assessment of bullying in schools can be used to determine large-scale prevalence rates and trends or inform decisions that impact the development of regulations and policies to promote school safety [14]. At the micro level, assessment of bullying in schools can serve to identify those students who may be involved in bullying or victimization [15] and inform amelioration efforts (i.e., intervention selection, progress monitoring; [16]. Given the frequency at which students are involved in bullying, there is a need for the identification of these issues at the school level that includes all students, including those with and without special needs [17]. Much has been learned about what schools can do to address bullying, which has led to the development of prevention programs and intervention strategies that have benefited many ([2,18-20]. However, the lack of proper and frequent assessment of bullying behavior continues to plague many schools across the nation. This point is clearly illustrated in a large-scale study conducted by Kumpulainen, et al. [21] that found less than a quarter of students involved in bullying received supports, including any mental health services.

Despite the availability of multiple measures of bullying behavior such as the California bullying victimization scale [22] and the widely used Olweus bully/victim questionnaire [23], there appears to be a paucity of research conducted on the use of brief behavior rating scales to progress monitor bullying in schools. Moreover, the utility of previously developed measures for progress monitoring in schools appear limited by design. As clinical concerns and legal mandates continue to call for additional efforts to identify, prevent, and intervene, it has been made clear that there is a need for more cost-effective and research-based methods for assessment [2,24,25]. If we are to move towards a more effective behavioral response-to-intervention framework for addressing bullying in our schools, we will need to have tools available that can be feasibly used in school settings.

Health Behavior in School-Aged Children (HBSC)

The Health Behavior in School-Aged Children (HBSC) survey is a potentially valuable survey for use internationally, as is was developed as part of a large-scale study sponsored by the World Health Organization [26] to examine the health-related attitudes and behaviors of children and adolescents. Between 2009 and 2010, it was administered across 41 countries in Europe and North America. The sample was selected using a multi-stage sampling strategy where schools were selected at random from a sample of schools, and then within each school, a random sample of students were selected to participate in the survey [27]. The survey is made up of items addressing a variety of health-related topics, including social relationships, physical health, dietary habits, exercise, body image, drug/alcohol use, and bullying behaviors. Researchers have examined various subsets of items from the HBSC with findings indicating moderate to good validity and reliability. Items related to bullying behavior were based on the Olweus Bully/Victim Questionnaire, which has been found to have good validity and reliability [28]. Recently, Roberson and Renshaw [29] investigated the factor structure of the bullying items of the HBSC self-report and concluded that their findings supported its structural validity for measuring student bullying behaviors. While initial psychometric evidence supporting the use of the HBSC as a measure of bullying behavior appears promising, further validation efforts are necessary to establish the psychometric properties.

The Current Study

Initial examination of the structural validity of the HBSC survey bullying items appears promising. However, additional analysis beyond the initial examination [29] is required to establish the utility of the HBSC as a bullying progress monitoring instrument. A contemporary approach to assessment validation emphasizes that "validation" occurs over a series of studies designed to accumulate evidence that supports its proposed uses and interpretations (i.e., the Interpretation and Use Argument [IUA]) [30]. Rather than considering psychometric evidence in isolation (i.e., a single study), validation occurs through the repeated collection and dissemination of familiar reliability and validity evidence (e.g., content validity, interrater reliability), with samples representative of all intended assessment subjects (i.e., multiple studies; e.g., [30,31]. To this end, the current study aims were three-fold. First, this study examined the factor structure of the HBSC bullying items, including one additional item related to the occurrence of physical fighting, as this has been found to be positively associated with involvement in bullying victimization for male and female students [32]. In addition, a multiple groups factor analysis was used to determine whether the factor structure is applicable across both male and female students. Lastly, this study evaluated potential differential item performance for male and female ratings of involvement in bullying and victimization. Measurement invariance (i.e., equivalent item functioning) is anticipated across these groups.

Method

Participants

Participants included two independent samples compiled for students in the United States (U.S.) in grades five through 10 (N=12,642). Using SPSS software version 25, the dataset was split into two samples consisting of approximately 50% of observations each; assignment of observations to either sample was random. Sample 1 was used for the Exploratory Factor Analysis (EFA), and Sample 2 was used for both the Confirmatory Factor Analysis (CFA) and Multiple Groups CFA (MGCFA). Demographic characteristics (i.e., gender, grade level, and ethnicity) between the full and split samples were comparable (see Table 1).

Table 1: Demographic characteristics for the full and split samples.

	Sample 1		Sample 2		Full Sample	
	n	%	n	%	n	%
Total Sample	6,366	100	6,276	100	12,642	100
Gender						
Male	3,282	51.6	3,220	51.3	6,502	51.4
Female	3,082	48.4	3,054	48.7	6,136	48.6
Grade Level						
5th	875	13.7	842	13.4	1,717	7.6
6th	1,051	16.5	999	15.9	2,050	9.1
7th	11,164	18.3	1,257	20	12,421	54.9
8th	1,277	20.1	1,198	19.1	2,475	10.9
9th	1,043	16.4	1,029	16.4	2,072	9.2
10th	956	15	951	15.2	1,907	8.4
Ethnicity						
American Indian or Alaskan Native	106	1.7	116	1.8	222	1.8
Asian	218	3.4	251	4	469	3.9
Black	1,073	16.9	1,091	17.4	2,164	17.9
Hispanic	1,172	18.4	1,220	19.4	2,392	19.8
Mixed Race	429	6.7	399	6.4	828	6.8
Native Hawaiian or Pacific Islander	58	0.9	53	0.8	111	0.9
White	3,032	47.6	2,871	45.7	5,903	48.8

Procedure

Data was gathered through anonymous self-report surveys administered to students in grade five through 10 in 314 schools across the U.S. during the 2009-2010 school year. Participating schools included public and private schools from all 50 states and the District of Columbia. Student assent and parental consent were obtained by the schools, and the study protocol was approved by the Institutional Review Board of the Eunice Kennedy Shriver National Institute of Child Health and Human Development. Scripted survey administration was conducted by school personnel (e.g., teacher, counselor, nurse).

Measure

The complete HBSC survey includes 76, 86, or 88 items (for grades five to six, grades seven to nine, or grade 10, respectively) and takes approximately 45 minutes to complete [33]. Although there are three versions of the survey to account for varied developmental levels of respondents, all versions contained the same 24 bullying items, which asked how often students bullied others or were bullied by others at school in various forms (i.e., physical, verbal, relational, or cyber-bullying) and were scored using a 5-point scale (i.e., [not in this way in the past couple months], only once or twice, 2 or 3 times a month, about once a week, several times a week). All versions also included an item that asked how often the student engaged in physical fighting, which was also scored used a 5-point scale (i.e., I have not been in a physical fight, 1 time, 2 times, 3 times, 4 times or more).

Statistical Analyses

Data screening

Data screening procedures were performed using SPSS software version 25 prior to conducting statistical analyses to ensure the overall accuracy of the data and determine whether assumptions have been met for proceeding with factor analysis. A total of 271 students (2%) from the full sample had missing values and were therefore excluded from the analyses [34]. Descriptive statistics including the range, mean, and standard deviation were calculated for each variable; all values were found to fall within plausible ranges. All items of interest were categorical variables that showed violations of normality as indicated by skewness and kurtosis values, histograms, and boxplots. Bivariate correlations were also examined; multicollinearity was not indicated (i.e., no variable pairs correlated above .90) and all items were significantly correlated with each other ($p<0.01$). Given the appearance of an underlying relationship between the items of interest, proceeding with factor analysis was appropriate.

EFA methodology

An EFA was conducted to identify the latent variables that account for the variance among the 25 selected items using Mplus software version 8 [35]. All items had five response categories and were treated as continuous variables [36]. Although normality is not a necessary assumption for factor analysis [37], the EFA was performed using maximum likelihood parameter estimates with robust standard errors and chi-square (i.e., MLR estimation) which is appropriate for non-normal continuous data with missing values [36,38]. To allow for correlation among factors, an oblique geomin rotation was used for this analysis [39]. Kaiser's "eigenvalue-greater-than-one" rule was employed to facilitate identification of factors appropriate for interpretation and retention, and Cattell's scree test was used to explore a graphical representation of the eigenvalues and divide the major factors from the minor factors. To supplement factor identification and retention based on noted limitations of Kaiser's eigenvalues and Cattell's scree plots [40], Parallel Analysis (PA) was also employed as a preliminary step in determining the number of appropriate factors. Using this method, factors with eigenvalues greater than those of a randomly generated correlation matrix were retained.

Other indices considered when identifying the appropriate number of HBSC factors include the chi-square test of model fit, Root-Mean-Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Standardized Root-Mean-Square Residual (SRMR), and factor loadings. The chi-square test of model fit, which measures the difference between the sample and predicted covariance matrices, indicates good fit when the result is non-significant. According to Fabrigar et al. [39], the chi-square statistic is greatly affected by sample size in that greater sample sizes can inflate chi-square; however, models with significant chi-square results can still be retained. Given the number of observations in Sample 1 (N=6,366), it was expected that chi-square would be significant. While chi-square tests the null hypothesis of perfect model fit, RMSEA recognizes that models can only be an approximation of what is observed in the real world. Following the criteria established by Browne and Cudeck [41], acceptable model fit is achieved when RMSEA is less than 0.08, and a value of 0.05 or below would be required to conclude good model fit. Similarly, SRMR values below .08 demonstrate acceptable fit [42]. CFI and TFI values between 0.90 and 0.95 indicate reasonable model fit, and values above 0.95 indicate good model fit [39,43]. Finally, consistent with the widely adopted criteria for the retention of factor loadings between 0.30 and 0.40 [43], only loadings above 0.35 were retained.

CFA methodology

Following the EFA, a CFA was then conducted using Mplus to determine whether the factor structure identified in the EFA can reproduce the observed relationships among the variables when applied to Sample 2. The CFA was conducted using the MLR estimator and oblique geomin rotation for the same reasons described for the EFA. Unit Loading Identification (ULI) constraints were employed in order to scale the factors. This was accomplished by setting factor loadings of reference items (i.e., those with the highest loadings) to 1.0 and allowing all other loadings to be freely estimated [44].

MGCFA methodology

Following the CFA, steps were taken to conduct a MGCFA using Mplus for establishing measurement invariance between male and female students. This process involves a series of model comparisons that adopt increasingly stringent parameter constraints. Through this process, parameters are constrained to be equal between groups, allowing for the evaluation of item equivalency between groups [43,45]. If a significant decrease is observed in the model fit, measurement non-invariance is indicated. The first step was to fit a model to the overall sample, followed by an evaluation of the separate CFA solutions for each group. After fitting the model separately for each group, simultaneous analyses were conducted. More specifically, configural, metric, and scalar invariance testing was performed, which involves the simultaneous analysis of CFA in both groups. Configural invariance testing was conducted by allowing all parameters (e.g., loadings, intercepts, and variances) to be freely estimated. The aim of this testing is to determine whether the same number of factors exist and if the same indicators load on each of the factors across both groups [43,46]. This model served as the baseline model against which other models were compared. Following configural invariance testing, metric invariance testing was conducted by holding the values of the factor loadings equal to determine whether they are different across groups. Finally, scalar invariance testing was conducted by setting factor loadings and intercepts to be equal to determine whether factor means and indicator intercepts are different across male and female students.

Chi-square difference testing was used to determine whether the addition of parameter constraints led to an increase in model misfit. Because analyses were conducted with MLR estimation, the test scaling correction was calculated using the following equation described by Muthén and Muthén [47],

$$c_d = \frac{\left(d_0 * c_0 - d_1 * c_1\right)}{\left(d_0 - d_1\right)}$$

With the difference test scaling correction value, the Satorra-Bentler scaled chi-square difference test [48] was then calculated with the following equation,

$$TR_d = \frac{\left(T_0 * c_0 - T_1 * c_1\right)}{\left(c_d\right)}$$

Furthermore, change in the CFI (greater than 0.01) was used as an additional measure of model misfit as recommended by Cheung and Rensvold [46].

Results

EFA

Preliminary data screening indicated non-normality across all variables as evidenced by skewness values (greater than |2.0|), kurtosis values (greater than |7.0|), histograms, and boxplots. A total of 149 (2%) students were excluded from the EFA due to missing data across all items. Initial examination of the eigenvalues revealed three factors with values greater than 1.0, supporting that three factors should be retained per Kaiser's rule. The retention of three factors is further supported by examination of the scree plot for the initial EFA, which showed three points above the point where the curve levels off. When considering the parallel analysis that compared the sample data's eigenvalues with those generated from a random dataset, results indicate that three factors should be retained; the eigenvalue for Factor 3 was 1.714 for the sample data and 1.09 for the random data. Model fit indices were also examined to identify the model that best fit the data. As seen in Table 2, fit statistics improved with every increase in the number of factors. Chi-square was significant ($p<0.001$) across all tested models, which was expected given the large sample size [39,43]. RMSEA indicated good fit for a three-factor model. Both CFI and TLI indicated adequate fit for a three-factor model. Upon further examination of the three-factor model, it was revealed that numerous items cross-loaded on multiple factors, and one item did not significantly load onto any factor. More specifically, multiple items loaded highly (greater than 0.30) onto more than one factor, and one item loaded weakly (less than 0.30) across all factors.

Table 2: Fit Indices for initial EFA models and EFA models with problematic items removed.

Models	χ^2	df	P	RMSEA	RMSEA 90%		TLI	SRMR
					CI	CFI		
Initial								
1-Factor	8017.975*	275	<0.001	0.067	[0.066, 0.069]	0.619	0.585	0.135
2-Factor	3555.278*	251	<0.001	0.046	[0.045, 0.047]	0.838	0.806	0.059
3-Factor	1703.743*	228	<0.001	0.032	[0.031, 0.034]	0.927	0.905	0.032
4-Factor	940.559*	206	<0.001	0.024	[0.022, 0.026]	0.964	0.947	0.019
5-Factor	820.970*	185	<0.001	0.024	[0.022, 0.025]	0.969	0.949	0.016
Problematic Items Removed								
1-Factor	4228.852*	152	<0.001	0.066	[0.064, 0.068]	0.715	0.679	0.139
2-Factor	1105.246*	134	<0.001	0.034	[0.032, 0.036]	0.932	0.913	0.036
3-Factor	594.990*	117	<0.001	0.026	[0.024, 0.028]	0.967	0.951	0.023
4-Factor	355.720*	101	<0.001	0.02	[0.018, 0.023]	0.982	0.97	0.016

Note: χ^2=chi-square test of model fit; RMSEA=Root-Mean Square Error of Approximation; CFI=Comparative Fit Index; TLI=Tucker-Lewis; SRMR=Standardized Root Mean Square *$p<0.001$.

Additional solutions were generated with problematic items deleted one at a time across iterations; fit statistics and factor loadings were reevaluated for each solution. After removal of six problematic items, a review of eigenvalues, the scree plot, and parallel analysis results supported a two-factor structure rather than a three-factor structure. Of the six items that were removed, five items (how often bullied others using computer outside school, bullied others using a cell phone outside school, got bullied using computer, got bullied using cell phone, called another student names) cross-loaded highly onto multiple factors, and one item (how often in a physical fight) did not load highly onto any factor. Examination of the factor loadings of the two-factor model showed that all items loaded highly onto a single factor. Fit statistics support a two-factor model; χ^2 (134)=1105.246, p<0.001, RMSEA=0.034, CFI=0.932, TLI=0.913, SRMR=0.036 (see Table 2). While the chi-square statistic is significant, all other fit indices calculated indicate adequate to good model fit. It is noteworthy that the three-factor model fit statistics are better when compared to the two-factor model; however, examination of the three-factor model revealed multiple problematic variables. When considering all of the available information, a three-factor model did not appear to be appropriate. Thus, a two-factor model was selected as the final EFA model with the identified latent factors of bullying and victimization. Factor loadings for the final EFA model are presented in Table 3.

Table 3: Geomin rotated factor loadings of final two-factor solution.

Indicator	Factors	
	Victim	Bully
How often are you bullied at school	0.7	-0.1
How often are you called names/teased	0.8	-0.09
How often are you left out of things	0.72	0
How often are you hit/kicked/pushed	0.63	0.16
How often others lied about me	0.74	0
How often bullied for my race/color	0.61	0.16
How often bullied for my religion	0.53	0.21
How often other students made sexual jokes about me	0.62	0.08
How often you bullied another student	0.11	0.36
How often left another student out of things	0.09	0.63
How often hit/kicked/pushed another student	0.08	0.69
How often told lies about another student	0.07	0.76
How often bullied others for their race/color	0.03	0.78
How often bullied others for their religion	0.01	0.84
How often made sexual jokes about another student	0.05	0.76
How often bullied others using a computer/e-mail	-0.03	0.92
How often bullied others using a cell phone	-0.02	0.92
How often bullied others using a computer/e-mail, outside of school	-0.02	0.9
How often bullied others using a cell phone, outside of school	-0.01	0.88

Note: Bolded loadings represent indicators loading on respective factor. Selection criteria: Loading>0.35.

CFA

A second independent and equivalent sample (Sample 2, n=6,276) was used to verify the factor structure identified in the EFA. Results indicate adequate to good fit with the exception of chi-square and TLI [42]; χ^2(151)=1286.019, p<0.001, RMSEA=0.035, CFI=0.909, TLI=0.897, SRMR=0.054. As seen in Table 4, all factor loadings are significant (p<0.001) with magnitudes ranging from moderate to large. The modification indices and Expected Parameter Change (EPC) values were also examined to determine whether respecification of the model would be appropriate in order to improve model fit. Several outliers were identified, including three pairs of variables with relatively high modification indices and EPC values. However, these EPC values were not substantial (range=0.37 to 0.38) which did not suggest modifications would markedly improve model fit. Given that the model generally has adequate to good fit, respecification was deemed unnecessary.

Table 4: Factor loadings of two-factor CFA solution.

Indicator	Factors	
	Victim	Bully
How often are you bullied at school	0.65	
How often are you called names/teased	0.74	
How often are you left out of things	0.71	
How often are you hit/kicked/pushed	0.7	
How often others lied about me	0.74	
How often bullied for my race/color	0.69	
How often bullied for my religion	0.65	
How often other students made sexual jokes about me	0.67	
How often you bullied another student		0.47
How often left another student out of things		0.66
How often hit/ kicked/pushed another student		0.71
How often told lies about another student		0.78
How often bullied others for their race/color		0.76
How often bullied others for their religion		0.81
How often made sexual jokes about another student		0.73
How often bullied others using a computer/e-mail		0.87
How often bullied others using a cell phone		0.87
How often bullied others using a computer/e-mail, outside of school		0.88
How often bullied others using a cell phone, outside of school		0.84

Note: All factor loadings are significant at p<0.001.

MGCFA

A MGCFA invariance evaluation was conducted to determine if the measurement properties of the bullying scale are equivalent between males and females, research aim 3. First, CFAs were conducted separately for males and females. In general, both models displayed adequate fit (see Table 5). Chi-square was significant across both male and female models as expected. While all other fit indices for the male model are adequate to good (RMSEA=0.036, SRMR=0.052, CFI=0.924, TLI=0.914), the fit indices appeared weaker for the female model. For the female model, all fit indices indicated good fit with the exception of CFI and TLI, approached adequate fit (RMSEA=0.036, SRMR=0.056, CFI=0.885, TLI=0.870). Examination of the factor loadings revealed all were statistically significant across both males and females (p<0.001). Overall, the model fit was judged to be of adequate fit for both males and females samples. After confirming adequate model fit separately for male and female students, a baseline model was fit in which the loading pattern was equivalent between both groups but all parameters (e.g., loadings, intercepts, and variances) were allowed to vary. In general, the baseline model presented adequate to good fit; χ^2(302)=2703.113, p<0.001, RMSEA=0.036, SRMR=0.058, CFI=0.907. Examination of the factor loadings indicate that they are all statistically significant for both groups. Thus, configural invariance was determined to exist.

Table 5: Fit statistics for multiple groups.

	χ^2	S-Bχ^2	df	Δdf	RMSEA (90% CI)	SRMR	CFI	ΔCFI
Single Group Solution Overall Sample (N=12,279)	2491.540*	-	151		0.036 [0.034, 0.037]	0.056	0.912	
Male (N=6,280)	1350.128*	-	151		0.036 [0.034, 0.037]	0.052	0.924	
Female (N=5,995)	1352.831*	-	151		0.036 [0.035, 0.038]	0.063	0.885	
Measurement Invariance Configural	2703.113*	-	302	-	0.036 [0.035, 0.037]	0.058	0.907	-
Metric (equal loadings)	2756.778*	73.235*	319	17	0.035 [0.034, 0.037]	0.06	0.906	0.01
Scalar (equal loadings and intercepts)	2991.464*	569.044*	336	17	0.036 [0.035, 0.037]	0.059	0.898	0.008

Note: S-Bχ^2: Satorra-Bentler scaled chi-square difference test; RMSEA=Root-Mean Square Error of Approximation; SRMR=Standardized Root Mean Square; CFI=Comparative Fit Index; *p<0.001.

Second, a weak factorial metric invariance model was fit to the data by constraining all factor loadings to be equal. The fit of this model appears to be adequate to good; $\chi^2(319)=2756.778$, $p<0.001$, RMSEA=0.035, SRMR =0.060, CFI=0.906. RMSEA and SRMR values fell within the range of good model fit, and the CFI value met criteria for acceptable model fit. To assess whether constraining factor loadings significantly worsened fit, both the Satorra-Bentler scaled chi-square difference test and CFI difference test were conducted. Difference testing did not indicate fit to significantly worsen; S-Bχ^2=73.235, $p<0.001$, ΔCFI=0.010. Given that the metric invariance model is not substantially worse when compared to the baseline model, it was concluded that metric invariance exists.

Third, a strong invariance scalar model was fit to the data by constraining factor loadings and item intercepts to be equal. Fit statistics for this model indicate good fit with the exception of CFI; $\chi^2(336)=2991.464$, $p<0.001$, RMSEA=0.036, SRMR=0.059, CFI=0.898. Results of the Satorra-Bentler scaled chi-square difference test and CFI difference test do not reveal a significant difference between the scalar invariance model and the metric invariance model; S-Bχ^2=569.044, $p<0.001$, ΔCFI=0.008. Thus, scalar invariance was determined to exist. Finally, factor means were compared between the two groups. The estimated factor mean for the female group on the Victimization factor was not found to be significantly different when compared to males (p=0.149). However, examination of the estimated factor mean for females on the Bullying factor was found to be significantly different when compared to males (p<0.001). The mean value for female students was 0.055 lower than the mean for male students (Figure 1).

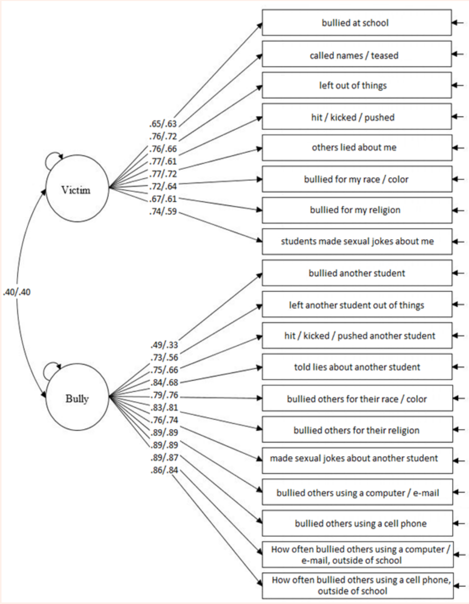

Figure 1: Path diagram for MGCFA with standardized loadings and factor correlations for males/females. All estimates are significant at p<0.001.

Discussion

Bullying and victimization behavior continues to be a persistent challenge for students and educators in schools around the world [2,4]. The negative outcomes for victims and perpetrators are now well-documented and significant (Espelage & Swearer 2003) [2,7,9]. As schools around the world increasingly turning to data and data-based decision-making to guide service-delivery policy and practice choices, collection of data to drive such decisions related to bullying behavior is critical. Unfortunately, the number of assessments for bullying behavior in schools appears limited. The HBSC survey appears as an emerging assessment option to address this noted assessment shortage [29]. Although promising, to date, the psychometric properties of the HBSC have yet to be thoroughly explored. The purpose of the study was to further examine the factor structure of a subset of items related to bullying behavior in the HBSC survey. Results of the EFA revealed a two-factor solution after the removal of six problematic variables, five of which cross-loaded onto multiple factors and one variable related to physical fighting that did not saliently load onto any factor. The final two-factor solution was retained with the following latent factors identified: Bullying and Victimization. Following the EFA, a CFA was conducted with a seperate sample and confirmed the acceptable fit of the initially identify two factor solution.

A MGCFA was conducted to determine whether the factor structure identified is applicable for both male and female students. Measurement invariance was established across the configural, metric, and scalar models. Results revealed equal form, factor loadings, and intercepts across gender, establishing adequate and equivalent fit for the two-factor, 19-item scale for both male and female students. Findings reveal that scores generated using the 19-item HBSC scale could be used for identification and progress monitoring of bullying behavior in schools, and comparisons of scores for male and female students appears appropriate. Finally, contrary to findings from the National Crime Victimization Survey suggesting female students are bullied more often than males (31% and 25%, respectively; [49,50], in the present study a comparison of the group means of the latent factors revealed that male and female students did not differ significantly in their reports of victimization. However, a significant difference was found in reports of involvement in bullying behavior, with female students reporting slightly lower involvement in bullying behaviors targeting other students.

Limitations and Future Directions

While findings from this study are promising and should be considered one of multiple additional steps in validating a modified, 19-item HBSC bullying item assessment, this study is not without limitations. First, the extraction of HBSC bullying items from the larger HBSC survey may influence student responses. Given the length of the HBSC, survey fatigue may have impacted responses for some participants, resulting in an inaccurate depiction of their experiences related to bullying in schools. Additionally, while the samples used in this study appear robust and generally representative of the demographic population of school children in the U.S., sampling is always imperfect, thus limiting generalizability of findings if only to a small degree. Further data collection in countries around the world is important to examine the psychometrics of the HBSC and appropriateness for use with diverse populations of children in countries around the world. Similarly, timing of data collection should be considered when evaluating generalizability of findings. For example, bullying behavior in schools is an ongoing phenomenon, with instances emerging throughout each academic year. If collected early or later in a school year, responses could change dramatically. It should also be noted that the data used in these analyses may be slightly dated. Since the data used in this study was collected, technology and the use of social media, has evolved significantly. Access to applications or websites that increase opportunities for students to interact, including engaging in relational aggression (i.e., online bullying) has increased significantly since 2010. Furthermore, advancements in hardware (e.g., smartphones, tablets, computers) as well as internet and networking accessibility have also contributed to advancements the capabilities of students to interact online. The findings reported likely do not reflect this evolution and therefore may not reflect contemporary prevalence of web-based bullying. Similarly, the socio-political climate has also changed significantly since this data has been collected. Political, racial, ethnic, and religious among other divisions have become extremely divisive and offer further opportunities to attack and harm others (i.e., bullying behavior).

Based on noted limitations, future research should continue to evaluate the psychometric properties of the 19-item assessment derived from the HBSC bullying items. Future research will be needed to evaluate the construct validity of this abbreviated scale. Such evaluations should include new samples and use of the 19-item assessment independently. Future works should also include new analytic techniques (e.g., Differential Item/Test Functioning [DIF/DTF]; longitudinal analyses [latent profile analysis]). Item invariance evaluation across additional identity characteristics including, but not limited to race, ethnicity, socioeconomic status, religion, and gender inclusive of fluid or nonconforming students. To enhance the scale's clinical utility, future research should address the ability of the 19-item measure to distinguish between those who are and those who are not experiencing significant involvement in bullying behavior. Lastly, future research should evaluate the applied use of data generated using the 19-item HBSC (i.e., translation to actual data-based decision-making).

Conclusion

Given the widespread and significant impact of bullying behaviors in many countries around the world [2], it is imperative that schools are able to appropriately screen and monitor school-oriented involvement in bullying behaviors. Use of valid, cost-effective, and brief assessments of bullying in schools promotes safer learning environments for all students by drawing attention to and monitoring bullying behavior. Results of this study build on previous preliminary work with a subset of bullying items from the HBSC survey as a standalone scale for measuring student involvement in bullying. As a brief behavior scale, there may be potential for its use in determining the frequency at which bullying behavior occurs (i.e., as a screener), selecting appropriate targets for prevention and response efforts, progress monitoring bullying behavior in schools, and gauging the effectiveness of prevention and intervention programs.

References

1. Hymel S, Swearer SM (2015) Four decades of research on school bullying: An introduction. American Psychologist 70(4): 293-299.
2. Jimerson SR, Swearer SM, Espelage DL, (Eds.), (2010) The handbook of bullying in schools: An international perspective. 1st (Edn.), Routledge Publishers, USA, pp. 1-624.
3. Olweus D (1993) Victimization by peers: Antecedents and long-term outcomes. In: Rubin KH, Asendorpf JB (Eds.), Social withdrawal, inhibition and shyness in childhood. Psychology Press, USA, pp. 315-341.
4. Underwood JM, Brener N, Thornton J, Harris WA, Bryan LN, et al. (2020) Overview and methods for the youth risk behavior surveillance system—United States, 2019. MMWR Supplements 69(1): 1-10.
5. Gladden RM, Vivolo-Kantor AM, Hamburger ME, Lumpkin CD (2014) Bullying surveillance among youths: Uniform definitions for public health and recommended data elements, Version 1.0. CDC and USDOE.
6. Garrett AG (2003) Bullying in American schools: Causes, preventions, interventions. McFarland & Company Inc. Publishers, USA, pp. 1-180.
7. Jan A, Husain S (2015) Bullying in elementary schools: Its causes and effects on students. Journal of Education and Practice 6(19): 43-56.
8. Franks J, Rawana E, Brownlee K (2013) The relationship between strengths in youth and bullying experiences at school. Educational & Child Psychology 30(4): 44-58.
9. Tennant JE, Klossing JJ, Demaray MK, Dorio N, Bixler T, et al. (2019) Internalizing problems of youth involved in bullying via different participant role combinations and gender. School Psychology Review 48(3): 222-236.
10. Sourander A, Jensen P, Rönning JA, Niemelä S, Helenius H, et al. (2007) What is the early adulthood outcome of boys who bully or are bullied in childhood? The finnish "From a Boy to a Man" study. Pediatrics 120(2): 397-404.
11. Riley-Tillman TC, Kalberer SM, Chafouleas SM (2005) Selecting the right tool for the job: A review of behavior monitoring tools used to assess student response to intervention. The California School Psychologist 10(1): 81-91.
12. Albers CA, Glover TA, Kratochwill TR (2007) Where are we, and where do we go now? Universal screening for enhanced educational and mental health outcomes. Journal of School Psychology 45(2): 257-263.
13. Stecker PM, Lembke ES, Foegen A (2008) Using progress-monitoring data to improve instructional decision making. Preventing School Failure: Alternative Education for Children and Youth 52(2): 48-58.
14. Greif JL, Furlong MJ (2006) The assessment of school bullying: Using theory to inform practice. Journal of School Violence 5(3): 33-50.
15. Cornell D, Cole JC (2012) Assessment of bullying. In: Jimerson S, Nickerson A, Mayer MJ, Furlong MJ, (Eds.), Handbook of School Violence and School Safety: International Research and Practice. Routledge Publishers, UK, pp. 298-312.
16. Baldry AC, Farrington DP (2004) Evaluation of an intervention program for the reduction of bullying and victimization in schools. Aggressive Behavior 30(1): 1-15.
17. Twyman KA, Saylor CF, Saia D, Macias MM, Taylor LA, et al. (2010) Bullying and ostracism experiences in children with special health care needs. Journal of Developmental & Behavioral Pediatrics 31(1): 1-8.
18. Knox J, Gibson S, Gönültaş S, Mulvey KL (2021) School connectedness and bystander intervention: The moderating role of perceived exclusion and privilege among African American students. School Psychology Review 50(2-3): 316-329.
19. Leff SS, Waasdorp TE, Paskewich BS, Winston FK (2021) Scaling and disseminating brief bullying prevention programming: Strengths, challenges, and considerations. School Psychology Review 50(2-3): 454-468.
20. Sharp S, Smith PK (2002) School bullying: Insights and perspectives. Routledge Publishers, UK.
21. Kumpulainen K, Räsänen E, Puura K (2001) Psychiatric disorders and the use of mental health services among children involved in bullying. Aggressive Behavior 27(2): 102-110.
22. Felix ED, Sharkey JD, Green JG, Furlong MJ, Tanigawa D (2011) Getting precise and pragmatic about the assessment of bullying: The development of the *California Bullying Victimization Scale*. Aggressive Behavior 37(3): 234-247.
23. Olweus D (1996) The revised olweus bully/victim questionnaire. Research Center for Health Promotion, University of Bergen, Norway.
24. Saylor CF, Nida SA, Williams KD, Taylor LA, Smyth W, et al. (2012) Bullying and Ostracism Screening Scales (BOSS): Development and applications. Children's Health Care 41(4): 322-343.
25. Swearer SM, Espelage DL, Napolitano SA (2009) Bullying prevention and intervention: Realistic strategies for schools. 1st (Edn.), Guilford Press, UK, pp. 1-170.
26. Currie C, Gabhainn SN, Godeau E, Roberts C, Smith R, et al. (2008) Inequalities in young people's health: HBSC international report from the 2005/2006 Survey. pp. 1-224.
27. Currie C, Inchley J, Molcho M, Lenzi M, Veselska Z, et al. (2014) Health Behaviour in School-Aged Children (HBSC) study protocol: Background, methodology and mandatory items for the 2013/14 survey. HBSC.
28. Wang J, Iannotti RJ, Nansel TR (2009) School bullying among adolescents in the United States: Physical, verbal, relational, and cyber. Journal of Adolescent health 45(4): 368-375.
29. Roberson AJ, Renshaw TL (2018) Structural validity of the HBSC bullying measure: Self-report rating scales of youth victimization and perpetration behavior. Journal of Psychoeducational Assessment 36(6): 628-643.
30. Kane M, Burns M (2013) The argument-based approach to validation. School Psychology Review 42(4): 448-457.

31. Cook DA, Brydges R, Ginsburg S, Hatala R (2015) A contemporary approach to validity arguments: A practical guide to Kane's framework. Medical Education 49(6): 560-575.

32. Muula AS, Herring P, Siziya S, Rudatsikira E (2009) Bullying victimization and physical fighting among Venezuelan adolescents in Barinas: Results from the global school-based health survey 2003. Italian Journal of Pediatrics 35(1): 38.

33. Iannotti RJ (2013) Health Behavior in School-Aged Children (HBSC), 2009-2010. Inter-university Consortium for Political and Social Research.

34. Langkamp DL, Lehman A, Lemeshow S (2010) Techniques for handling missing data in secondary analyses of large surveys. Academic Pediatrics 10(3): 205-210.

35. Muthén LK, Muthén BO (2017) Mplus User's Guide. 7th (Edn.), pp. 1-950.

36. Newsom J (2018) Practical approaches to dealing with nonnormal and categorical variables. Newsom Psy 523/623 Structural Equation Modeling, Spring, pp. 1-4.

37. Tabachnick B, Fidell L (2001) Using multivariate statistics. 7th (Edn.), Pearson Education Publishers, UK, pp. 1-848.

38. Yuan KH, Bentler PM (2000) Three likelihood based methods for mean and covariance structure analysis with nonnormal missing data. Sociological Methodology 30(1): 165-200.

39. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ (1999) Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods 4(3): 272-299.

40. Kanyongo GY (2005) The influence of reliability on four rules for determining the number of components to retain. Journal of Modern Applied Statistical Methods 5(2): 332-343.

41. Browne MW, Cudeck R (1993) Alternative ways of assessing model fit. In: Bollen KA, Long JS (Eds.), Testing structural equation models. Sage Publishers, USA, pp. 136-162.

42. Hu LT, Bentler PM (1998) Fit indices in covariance structure modeling: Sensitivity to under parameterized model misspecification. Psychological Methods 3(4): 424-453.

43. Brown TA (2015) Confirmatory factor analysis for applied research. 2nd (Edn.), Guilford Publications, USA.

44. Steiger JH (2002) When constraints interact: A caution about reference variables, identification constraints, and scale dependencies in structural equation modeling. Psychological Methods 7(2): 210-227.

45. Cheung GW, Rensvold RB (1999) Testing factorial invariance across groups: A reconceptualization and proposed new method. Journal of Management 25(1): 1-27.

46. Cheung GW, Rensvold RB (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling 9(2): 233-255.

47. Muthén LK, Muthén BO (2005) Chi-square difference testing using the Satorra-Bentler scaled chi-square. Mplus.

48. Satorra A, Bentler PM (2001) A scaled difference chi-square test statistic for moment structure analysis. Psychometrika 66(4): 507-514.

49. DeVoe J, Murphy C (2011) Student reports of bullying and cyber-bullying: Results from the 2009 school crime supplement to the national crime victimization survey. National Center for Education Statistics, pp. 1-52.

50. Zhu Q, Cheong Y, Wang C (2022) Relation between peer victimization and mental health among Chinese students: Perceived school climate and covitality as potential moderators. School Psychology Review.