

INITIAL DEVELOPMENT AND VALIDATION OF THE BULLYHARM: THE BULLYING, HARASSMENT, AND AGGRESSION RECEIPT MEASURE

WILLIAM J. HALL

University of North Carolina at Chapel Hill

This article describes the development and preliminary validation of the Bullying, Harassment, and Aggression Receipt Measure (BullyHARM). The development of the BullyHARM involved a number of steps and methods, including a literature review, expert review, cognitive testing, readability testing, data collection from a large sample, reliability testing, and confirmatory factor analysis. A sample of 275 middle school students was used to examine the psychometric properties and factor structure of the BullyHARM, which consists of 22 items and six subscales: physical bullying, verbal bullying, social/relational bullying, cyber-bullying, property bullying, and sexual bullying. First-order and second-order factor models were evaluated. Results demonstrate that the first-order factor model had superior fit. Results of reliability testing indicate that the BullyHARM scale and subscales have very good internal consistency reliability. Findings indicate that the BullyHARM has good properties regarding content validation and respondent-related validation, and is a promising instrument for measuring bullying victimization in school. © 2016 Wiley Periodicals, Inc.

Bullying: A Pressing School Problem

Bullying in schools is a significant threat to the educational and psychological well-being of students. Recent nationally representative studies show that 20–22% of middle and high school students are victims of bullying (Kann et al., 2015; U.S. Department of Education, 2015). Victims of bullying may experience educational problems (i.e., absenteeism and low academic performance), as well as psychological problems (i.e., low self-esteem, anxiety, psychosomatic problems, conduct problems, depression, and suicidal ideation and behavior; Arseneault, Bowes, & Shakoor, 2010; Arseneault et al., 2006; Dake, Price, & Telljohann, 2003; Gini & Pozzoli, 2009; Klomek, Sourander, & Gould, 2010; Nakamoto & Schwartz, 2010; Reijntjes et al., 2011; Reijntjes, Kamphuis, Prinzie, & Telch, 2010; Ttofi, Farrington, Lösel, & Loeber, 2011).

Bullying refers to unwanted aggressive behaviors enacted intentionally over time by an individual or group using some form of power to cause physical and/or psychological harm to another individual or group in a shared social context (Gladden, Vivolo-Kantor, Hamburger, & Lumpkin, 2014; Olweus, 2013). Thus, bullying is not playful or benign, and bullying behaviors are not random or accidental. Bullying is an intentional and interpersonal act that occurs in a social context where there is a relationship between the bully and victim. Typically, there is an asymmetric power relationship between the bully and victim (Olweus, 2013). For example, bullies may use their physical size or strength, interpersonal skills, popularity, strength in numbers, or dominant social status to target those who are small in stature or physically weak, seem passive and withdrawn, have few friends, or are members of socially disadvantaged groups. Bullying occurs in many social settings (e.g., schools, neighborhoods, and workplaces) and among various age groups. However, bullying

The author would like to thank Natasha Bowen for her invaluable assistance and support throughout the development of this measure and paper. The author also thanks the principal, computer teacher, and students who participated in the study—this would not have been possible without them. The author's research was supported by the National Research Service Award Postdoctoral Traineeship from the National Institute of Mental Health sponsored by Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, and the Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Grant Number: T32 MH019117.

Correspondence to: William J. Hall, University of North Carolina, Chapel Hill, NC 27599-7590. E-mail: wjhall@unc.edu

between students in schools is particularly pressing given that bullying arises during childhood and peaks during adolescence, which are vulnerable and formative periods of development (Goldblum, Craig, Pepler, & Connolly, 2007; Hall & Rounds, 2013). And, bullying in schools can threaten the well-being of students.

Bullying behaviors take various forms. Some bullying behaviors are direct (e.g., punching) and others are indirect (e.g., starting a rumor). More precisely, bullying can be physical (e.g., hitting), verbal (e.g., name calling), social/relational (e.g., spreading rumors), electronic (e.g., posting slurs online), possessional/property (e.g., defacing someone's property), and sexual (e.g., making offensive sexual gestures at someone). Bullying behaviors also vary in their frequency; however, definitions of bullying usually require patterns or repeated instances of aggressive behavior (Guerin & Hennessy, 2002). An exception may be made for a single yet severe incident of harassment with lasting effects (Olweus, 1993). Others assert that a single instance of harassment could represent bullying because there is a high likelihood that the harassment would be repeated (Gladden et al., 2014).

Existing Bullying Victimization Measures

It is critical to accurately and reliably measure experiences of bullying to inform the planning and to evaluate the effectiveness of antibullying interventions. However, two systematic reviews of bullying victimization measures (Hamburger, Basile, & Vivolo, 2011; Vivolo-Kantor, Martell, Holland, & Westby, 2014) revealed a number of measurement problems among existing instruments, including failure to assess all dimensions of bullying, questionable psychometric properties, and scale construction issues related to instructions and items with confusing wording as well as lack of appropriate recall time frames. These measurement issues call into question the validity of existing measures.

The two systematic reviews included 34 instruments assessing experiences of being bullied; however, only four (12%) of these instruments included the five components that characterize bullying (i.e., aggression, intentionality, harm, frequency, and power imbalance; Hamburger et al., 2011; Vivolo-Kantor et al., 2014). In addition, no instrument measured all of the different dimensions of bullying behaviors (i.e., physical, verbal, social/relational, electronic, property, and sexual bullying). Established bullying measures typically did not include items measuring cyber-bullying or sexual bullying (Hamburger et al., 2011; Vivolo-Kantor et al., 2014). Electronic bullying or cyber-bullying items were not incorporated into bullying scales developed before the current pervasion of cell phones and personal computers. However, cyber-bullying is a growing problem, especially among youth (Kowalski, Limber, & Agatston, 2012; Wang, Iannotti, & Nansel, 2009). In addition, sexual bullying items are virtually absent in existing scales (Hamburger et al., 2011; Vivolo-Kantor et al., 2014) despite evidence that girls and young women (Shute, Owens, & Slee, 2008) as well as youth who are or who are perceived to be lesbian, gay, bisexual, transgender, or queer (LGBTQ) are often subjected to sexual bullying (Dupper, 2013; Kimmel & Mahler, 2003; Rivers & Duncan, 2013; Varjas et al., 2008). Researchers at the Centers for Disease Control and Prevention (CDC) asserted that future measures should assess bullying of a sexual nature and cyber-bullying along with other traditional forms of bullying (Hamburger et al., 2011). The failure of existing measures to capture all of the major types of bullying is also reflected in the number of items as 41% of the 34 instruments consisted of 5–16 items, which may be insufficient to capture the full range of bullying behaviors.

Many of the 34 measures reviewed had psychometric issues. For example, 15% of the measures did not report internal consistency reliabilities (Hamburger et al., 2011; Vivolo-Kantor et al., 2014). This is problematic because these measures may have poor internal consistency reliabilities (i.e., Cronbach's alpha values below .70). In addition, reports on the initial development and validation of the bullying victimization measures often did not report or did not collect sources of evidence

regarding content validation, including a literature review, expert review, and factor analysis (Bowen, 2008). Over two-thirds (68%) of the measures were presented without evidence of content validation, which is problematic (Hamburger et al., 2011; Vivolo-Kantor et al., 2014).

Items in many bullying scales have wording or construction issues. Over one-third (35%) of bullying victimization measures did not include a time frame for recalling and reporting being bullied (e.g., in the past week, month, or year; Hamburger et al., 2011; Vivolo-Kantor et al., 2014). Measures that lack recall time frames are less precise and leave room for interpretation by the respondent (Griffin & Gross, 2004). In addition, nine scales (27%) initially present participants with a definition of bullying so that respondents understand the construct being assessed (Furlong, Sharkey, Felix, Tanigawa, & Green, 2010). These definitions are often broad, lengthy, and complex, such as the 192 word definition in the Olweus Bully/Victim Questionnaire (Olweus, 1996). Criticisms and questions have been raised about instruments that use this definition-based self-report strategy as they may bias responses and inadequately capture specific subtypes of bullying (Cornell, Sheras, & Cole, 2006; Green, Felix, Sharkey, Furlong, & Kras, 2013; Greif & Furlong, 2006; Hamby & Finkelhor, 2000; Kert, Codding, Tryon, & Shiyko, 2010; Vaillancourt et al., 2008). Additionally, long and complex definitions may be cumbersome for child and adolescent participants to read and hold in their minds while they complete a questionnaire. In sum, there is a need for bullying victimization measures that overcome shortcomings of existing scales, and for which clear evidence of adequate psychometric qualities is presented. Better measures of bullying are essential for progress in school-based bullying research and intervention initiatives.

Current Study

This article describes the preliminary development and validation of the BullyHARM (Bullying, Harassment, and Aggression Receipt Measure), which was designed to assess students' experiences of bullying by peers in middle or high school settings. This study focused on two types of measurement validation: (a) content validation, which is the extent to which a scale includes all the main dimensions of a construct and that statistics support the factor structure, and (b) respondent-related validation, which is the extent to which a scale is appropriate for its intended population and measures what it is intended to measure with targeted respondents (Bowen, 2008). This study was driven by the following questions: (a) Do the items capture all of the main dimensions of bullying? (b) Is the scale developmentally appropriate for respondents in the target population? (c) Do participants interpret and understand the scale instructions and items as intended? (d) Do the data support the proposed dimensions of bullying and the underlying factor structure?

METHODS

The creation of the BullyHARM involved a mixed-methods scale development approach: literature review, expert review, cognitive measurement testing, readability testing, data collection from a large sample, reliability testing, and confirmatory factor analysis (CFA). These scale development steps were based on recommendations from Bowen and Guo (2012) and DeVellis (2012). This study was approved by the author's institutional review board.

Literature Review and Initial Scale Construction

The first step undertaken in the creation of the BullyHARM was a review of existing measures, conceptualizations of bullying, and types or dimensions of bullying. Based on a review of conceptual frameworks of bullying, research summaries of bullying behaviors, and the CDC compendium of bullying measures (Kaiser & Rasminsky, 2009; Hamburger et al., 2011; Sampson, 2009; Sveinsson & Morris, 2007), six main types of bullying were identified and defined: physical bullying,

verbal bullying, social or relational bullying, cyber-bullying, property bullying, and sexual bullying. Physical bullying involves bodily contact between the bully and victim. Verbal bullying involves the use of hurtful or fear-provoking spoken words. Social/relational bullying involves the manipulation of interpersonal relationships to hurt someone's reputation, friendships, and/or social acceptance. Cyber-bullying involves the use of technology, such as cell phones, computers, and the Internet, to hurt someone. Property bullying involves stealing or damaging someone's belongings. Finally, sexual bullying involves offensive actions of a sexual nature. Within these behavioral categories, lists of specific bullying behaviors were developed based on the literature and existing measures. From these lists, an initial item pool of 29 statements was created with three or more items developed for each of the six dimensions of bullying. All of the items were developed by the author.

Instructions for respondents were also drafted: "Below is a list of mean, hurtful, and offensive things other students might have done to you on purpose in the past month and you couldn't stop them. Read each statement and then pick one answer that best describes how often it happened to you." The terms "bully," "bullied," "bullying," "victim," "victimized," and "victimization" were deliberately not included in the instructions or items. Research suggests such terms may bias participant responses because of the emotions and stigma associated with being bullied, which could lead to underreporting of bullying and imprecise measurement (Cornell & Brockenbrough, 2004; Espelage & Swearer, 2003; Furlong et al., 2010; Greif & Furlong, 2006; Hamby & Finkelhor, 2000; Kert et al., 2010; Vaillancourt et al., 2008). In the BullyHARM, a series of specific bullying behaviors are presented to respondents who then indicate the frequency with which they have experienced each behavior. Focusing on bullying behaviors cues participants to reflect on specific, concrete incidents (Furlong et al., 2010) and may decrease the likelihood of "faking good" response bias (Chan, Myron, & Crawshaw, 2005). The instructions used the words "mean, hurtful, and offensive" to signal that the behaviors being assessed were aggressive, unwanted, and caused harm. The phrase, "you couldn't stop them," indicates that a perpetrating student exercised some form of power over the targeted student, who was unable to defend or stop the bullying.

The scale stem also included a recall time frame and specified the interpersonal context: "During the past month, other students at school . . ." Short time frames (e.g., the past week or past 24 hours) can minimize recall bias because they are remembered more easily and precisely; however, short time frames may not be representative of respondents' typical life experiences. On the other hand, long recall time frames (e.g., the past year) may be more representative but less accurately remembered. Thus, we chose a moderate recall time frame of 1 month. A 4-point rating scale was selected with the following response options: *not in the past month*, *1 or 2 times in the past month*, *about 1 time a week*, and *about 2 or more times a week*. We expected that middle and high school students could understand and apply these response options because there is a logical progression in terms of frequency.

Expert Review

The preliminary draft of the scale was reviewed by four experts: one school principal, one expert on adolescent interpersonal communication, and two school-based researchers with practice experience in schools. The experts included three women and one man, and three of them were White and one person was multiracial. The experts were asked to comment on the clarity of the instructions, content of the items, response format, and their appropriateness for the target population. Based on feedback, wording changes were made to the instructions and items for clarity and simplicity, and two items were removed due to interpretation issues.

Cognitive Measurement Testing

Next, cognitive measurement testing was employed. This form of cognitive testing involves collecting information from members of the target population while they respond to scale items (Willis, 2005). Cognitive measurement testing allows researchers to examine the question and response process as well as issues of clarity and understandability (de Leeuw, Borgers, & Smits, 2004). A combination of think aloud and verbal probing techniques (Willis, 2005) was used with five participants located using convenience sampling from the town around the authors' university. These participants were high school students who responded to a locally posted flyer about the development of the measure. All participants were 18 years of age; two were male and three were female; and two identified as White, one as Latina, one as multiracial, and one as Jewish. All of the participants reported being bullied in high school.

Participants read the instructions and each item aloud, commented in their own words on what they were being instructed to do and what each item brought to their mind, picked a response option, and explained why they picked the option. No participants indicated problems understanding the instructions. Participants understood the items 98.5% of the time (i.e., two participants each misunderstood an item but not the same item). Response congruency was examined by comparing participants' response choices and their explanations for choosing each response with researcher assumptions about response frequency: *Not in the past month* implies a frequency of 0, *1 or 2 times in the past month* implies a frequency of 1 or 2, *About 1 time a week* implies a frequency of 3–5, and *About 2 or more times a week* implies a frequency of 6–8 or more. Participants' responses and their explanations for choosing each response were congruent 100% of the time with the researcher assumptions about frequency. In addition, all participants felt that the response options were adequate and fitting, and all participants felt confident that they could accurately recall bullying incidents in the past month. At the conclusion of each interview, recommendations were made concerning wording changes, merging items, adding items, and removing items. Based on cognitive measurement testing findings, wording changes were made to four items, two pairs of similar items were merged, and three items were removed leaving the BullyHARM with 22 items and six subscales: physical bullying (five items), verbal bullying (three items), social bullying (three items), cyber-bullying (three items), property bullying (three items), and sexual bullying (five items; see Table 1).

Readability Testing

The Flesch–Kincaid grade level reading formula was applied to the instrument using Microsoft Word version 14.0 to evaluate the readability of the items and instructions. This formula assesses the readability of English text by considering the number of words per sentence and the number of syllables per word. Reading level scores correspond to U.S. grade school levels. The BullyHARM scored 6.3, which suggests that a sixth-grade student would be able to read and understand the instrument.

Data Collection for Psychometric Testing

Following readability testing, the 22-item BullyHARM was administered to a sample of 275 middle school students. Data from this sample were used for reliability testing and CFA. Given the preliminary nature of the study, a representative sample was not imperative at this stage. A purposive sampling strategy was used because we wanted to sample a diverse group of middle school students and the school principal was open to and supportive of the study. A middle school was targeted because bullying is prevalent during these grades (Nansel et al., 2001), and it was assumed that a scale that could be understood and used by middle school students could also be understood and used by high school students. Bullying also occurs during high school but less frequently (Nansel

Table 1
 Summary of Latent Factors, Scale Items, Descriptive Statistics, and Standardized Factor Loadings for the First-Order Factor Model ($N = 275$)

Factors and Items	Responses				Loading	SE
	<i>Not in the Past Month</i> n (%)	<i>1 or 2 Times in the Past Month</i> n (%)	<i>About 1 Time a Week</i> n (%)	<i>About 2 or More Times a Week</i> n (%)		
Physical bullying						
1. Pushed or pulled on me	201 (73.6)	50 (18.3)	15 (5.5)	7 (2.6)	.93	.02
2. Threw something at me	199 (73.2)	49 (18.0)	19 (7.0)	5 (1.8)	.82	.04
3. Kicked or tripped me	188 (69.6)	51 (18.9)	23 (8.5)	8 (3.0)	.88	.03
4. Hit, punched, or slapped me	214 (78.7)	43 (15.8)	10 (3.7)	5 (1.8)	.84	.03
5. Attacked me with a weapon or object	260 (95.9)	5 (1.8)	3 (1.1)	3 (1.1)	.95	.07
Verbal bullying						
1. Called me a bad name	143 (52.4)	61 (22.3)	47 (17.2)	22 (8.1)	.95	.02
2. Said something mean to me	138 (50.4)	63 (23.0)	51 (18.6)	22 (8.0)	.93	.02
3. Said something to scare or intimidate me	217 (79.8)	35 (12.9)	10 (3.7)	10 (3.7)	.90	.03
Social bullying						
1. Tried to turn people against me	189 (69.2)	47 (17.2)	22 (8.1)	15 (5.5)	.89	.03
2. Started or spread a false rumor about me	189 (69.0)	51 (18.6)	19 (6.9)	15 (5.5)	.88	.03
3. Excluded me from their group	222 (81.3)	27 (9.9)	14 (5.1)	10 (3.7)	.84	.04
Cyber-bullying						
1. Posted something bad about me on the Internet	248 (90.2)	18 (6.5)	5 (1.8)	4 (1.5)	.97	.02
2. Sent me a mean e-mail, instant message, or text message	244 (88.7)	19 (6.9)	7 (2.5)	5 (1.8)	.94	.03
3. Made a mean comment about me on the Internet	243 (88.4)	19 (6.9)	7 (2.5)	6 (2.2)	.96	.02
Property bullying						
1. Broke or messed up something that belonged to me	234 (85.1)	27 (9.8)	9 (3.3)	5 (1.8)	.92	.03
2. Wrote or drew something bad on my things	235 (85.5)	30 (10.9)	6 (2.2)	4 (1.5)	.93	.04
3. Took something of mine and wouldn't give it back	208 (75.9)	42 (15.3)	15 (5.5)	9 (3.3)	.83	.04
Sexual bullying						
1. Touched me in a sexual way that made me uncomfortable	234 (85.4)	23 (8.4)	11 (4.0)	6 (2.2)	.90	.03
2. Wrote, drew, or said bad sexual things about me	241 (87.6)	17 (6.2)	12 (4.4)	5 (1.8)	.87	.04
3. Made an offensive sexual gesture or body movement at me	237 (86.2)	18 (6.5)	13 (4.7)	7 (2.5)	.89	.04
4. Tried to pull down or pull off my clothes without my permission	260 (94.5)	5 (1.8)	4 (1.5)	6 (2.2)	.88	.07
5. Touched my private parts (breast, butt, or crotch) without my okay	237 (86.2)	25 (9.1)	8 (2.9)	5 (1.8)	.83	.04

Note. Response options were coded *Not in the past month* = 0, *1, or 2 times in the past month* = 1, *About 1 time a week* = 2, and *About 2 or more times a week* = 3. All factor loadings were significant at $p < .05$.

et al., 2001). Based on the principal's wishes, participants completed the survey in their regularly scheduled computer classes in April 2013. To minimize disruption of school activities, data collection was restricted to two days; therefore, 275 (59.3%) of the 464 students at the school were sampled because computer classes were held on a rotating basis on certain days of the week.

Participants. The 275 participants (52.4% male, 47.6% female) were between the ages of 12 and 17 (mean age = 13.3, $SD = .86$). The sample was diverse: 65.5% Latino/Latina, 12.7% White, 11.3% Black, 8.0% multiracial, 1.1% Asian, .7% American Indian, and .7% did not report their race/ethnicity. The middle school was located in a small town in central North Carolina. The sample size of 275 for the CFA could be considered large (Kline, 2005), and exceeded the minimum number needed to achieve a power level of .80 based on the degrees of freedom in the proposed models and alpha set at .05 (MacCallum, Browne, & Sugawara, 1996).

Procedure. The author administered the BullyHARM in the form of a web survey. Items were grouped according to subscale in the form of a grid to assist with coherence. And, the web survey host program randomly ordered the subscales for each participant to counterbalance and thereby minimize sequence effects. Demographic questions followed the BullyHARM items. Consented participants were told that they were participating in a short survey about "inappropriate and harmful behaviors," and the scale instructions were read out loud and presented on the computer screens. Participants were assured that their responses would be kept confidential, told that there were no right or wrong answers, and instructed to not talk during the survey and not look at their neighbors' computers. The author and the computer teacher monitored the classroom, while students completed the survey on the school computers. Most participants completed the survey in about 5 minutes. Each participant was entered into a random drawing for two \$25 gift cards.

Data Analysis

CFA was used to evaluate the factor structure of the BullyHARM because of the strong hypotheses about the latent factors based on a substantial body of research and theory. As previously stated, the literature review revealed six distinct dimensions of bullying behaviors. However, it was unclear if these distinct yet related constructs could be accounted for by a common, underlying second-order construct. Therefore, two measurement models were examined: a first-order factor model and a second-order factor model (Figure 1). In the first-order factor model, all of the latent factors (i.e., dimensions of bullying) were measured directly with observed variables. In the second-order factor model, the latent factors were themselves indicators of an underlying construct of bullying. We predicted that there would be strong, statistically significant loadings of the observed variables on each respective factor.

Mplus version 6.0 was used to perform the CFA. Weighted least squares mean and variance-adjusted (WLSMV) estimation was used because the data were ordinal and non-normally distributed because most participants were not bullied. WLSMV estimation with ordinal data uses a polychoric correlation matrix and an associated diagonal weight matrix with standard errors and mean and variance adjusted chi-square test statistics (Bollen, 1989). WLSMV estimation produces unbiased parameter estimates and standard errors with ordinal data (Flora & Curran, 2004).

The quality of each model was evaluated using multiple fit criteria: comparative fit index (CFI) $\geq .97$, Tucker-Lewis index (TLI) $\geq .97$, root mean square error of approximation (RMSEA) $\leq .05$, weighted root mean square residual (WRMR) $\leq .90$. These model fit criteria were based on recommendations by Browne and Cudeck (1993), Muthén and Muthén (1998–2007), and Schermelleh-Engel, Moosbrugger, and Müller (2003).

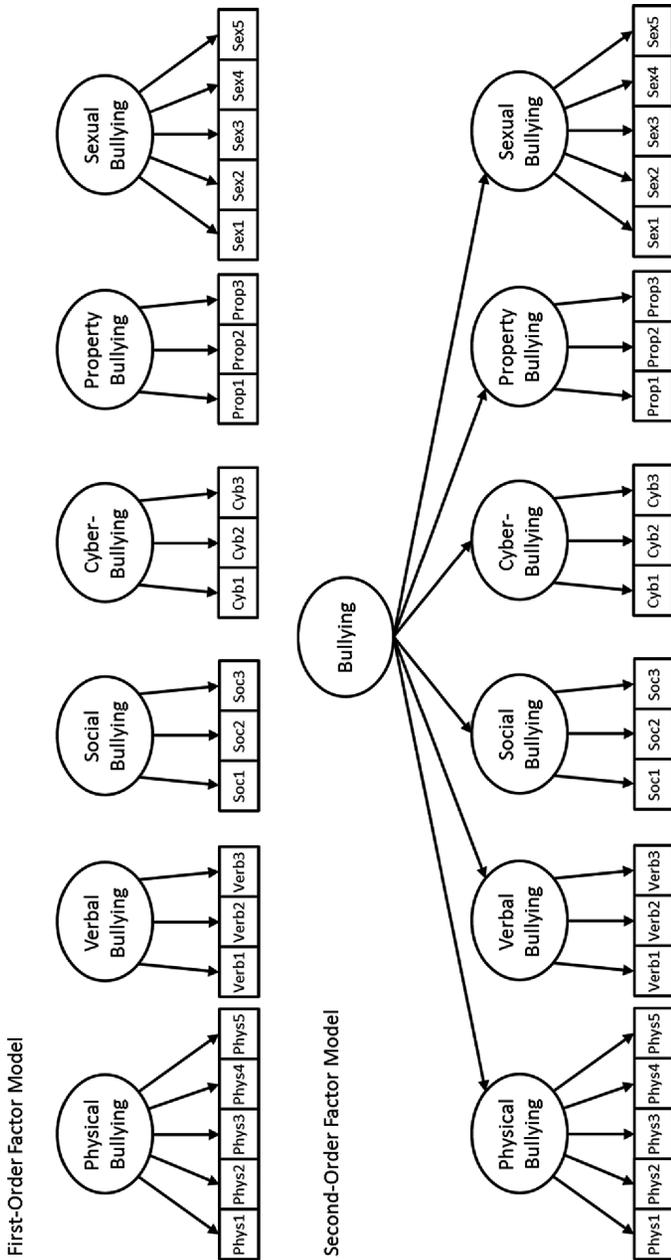


FIGURE 1. Confirmatory factor analysis models specified. Latent constructs are shown in ellipses and observed variables are shown in rectangles.

In addition, a derivatives difference test was employed to compare the models. The first-order factor model is nested in the second-order factor model. When using WLSMV in Mplus, a regular chi-square difference test is not possible because the distribution of the difference does not follow a normal chi-square distribution. The DIFFTEST procedure is used to compare nested models (Muthén & Muthén, 1998–2007). The less restrictive model or the model with fewer degrees of freedom is fit first, in this case the first-order factor model, then the more restrictive model is fit. A statistically significant difference test indicates the less restrictive model should be retained. If the difference test is not significant, the more parsimonious model can be retained because the additional constraints did not significantly worsen fit (Kline, 2005).

Only .5% of values were missing, and no item had more than 1.8% missing responses. Full information maximum likelihood was used to handle missing data. This procedure allows for all cases to be included in analyses, even if they are missing values on some variables. We also examined variance inflation factor (VIF) scores to check for multicollinearity, and there were no VIF scores greater than 5.

RESULTS

Descriptive Statistics

Table 1 shows the frequency counts and response percentages for each item. Verbal bullying was the most common type of bullying reported followed by social bullying, physical bullying, property bullying, sexual bullying, and cyber-bullying. For each bullying behavior, a majority of participants reported never being bullied. However, about half (49.6%) of participants reported experiencing at least one form of bullying behavior to some degree in the past month.

Internal Consistency Reliability

SPSS version 19 was used to assess the reliability for each of the six subscales and for the overall scale. Cronbach's alphas $\geq .80$ indicate very good reliability (DeVellis, 2012). Internal consistency reliability was very good for the overall scale ($\alpha = .93$) and each subscale: physical bullying ($\alpha = .85$), verbal bullying ($\alpha = .85$), social bullying ($\alpha = .82$), cyber-bullying ($\alpha = .91$), property bullying ($\alpha = .83$), and sexual bullying ($\alpha = .85$). High internal consistency reliability values suggest that the items used to measure a construct are highly interrelated (Kline, 2005).

Model Estimation and Evaluation

Table 2 shows the WLSMV estimates of chi-square and degrees of freedom for each model and the pre-established fit indices. The first-order factor model met all of the prestated model fit criteria (i.e., CFI $\geq .97$, TLI $\geq .97$, RMSEA $\leq .05$, and WRMR $\leq .90$), whereas the second-order factor model only met two of the criteria. In addition, results of the DIFFTEST procedure comparing the

Table 2
Goodness-of-Fit Statistics for the BullyHARM Models (N = 275)

Model	χ^2	df	CFI	TLI	RMSEA (90% CI)	WRMR
First-order factor model	311.98*	194	.98	.98	.047 (.037–.056)	.82
Second-order factor model	348.43*	203	.97	.97	.051 (.042–.060)	.96

Note. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; WRMR = weighted root mean square residual.

* $p < .05$.

Table 3
Interfactor Correlations for the First-Order Factor Model (N = 275)

Factor	1	2	3	4	5	6
1. Physical bullying	–					
2. Verbal bullying	.68	–				
3. Social bullying	.57	.83	–			
4. Cyber-bullying	.63	.70	.82	–		
5. Property bullying	.74	.66	.72	.66	–	
6. Sexual bullying	.56	.71	.81	.76	.65	–

Note. All latent interfactor correlations were significant at $p < .05$.

models showed that the more restrictive model (i.e., the second-order factor model) had statistically significantly worse fit, $\chi^2(9) = 32.66$, $p < .05$. The findings indicate that the first-order model should be retained.

Table 1 shows the CFA results of the first-order factor model, including the factor loadings and standard errors. The standardized factor loadings ranged from .82 to .97 and all were statistically significant. Table 3 shows the interfactor correlations, all of which were statistically significant. Correlations ranged from .56 to .83, supporting the independence of the dimensions. The variances of the six factors were all statistically significantly different from 0, indicating the latent variables captured meaningful variation among sample members.

DISCUSSION

The findings will be discussed in the order of the research questions driving the study: (a) Do the items capture all of the major dimensions of bullying? (b) Is the scale developmentally appropriate for respondents in the intended target population? (c) Do participants interpret and understand the scale instructions and items as intended? (d) Do the data support the proposed dimensions of bullying and the underlying factor structure?

Dimensions of Bullying in the Scale

The generation and selection of items that comprise the BullyHARM involved a thorough literature review, expert review, and cognitive measurement testing to include indicators of all the major types of bullying: physical, verbal, relational/social, cyber, property, and sexual. Unlike existing bullying scales, the BullyHARM assesses all six of the most commonly reported types of bullying and contains three to five indicators for each bullying dimension. In the literature, there are some instances of specific bias-based or discriminatory bullying (e.g., racist bullying or homophobic bullying; Elamé, 2013; Peguero, 2012). Specific items related to these forms of bullying were not included in the BullyHARM primarily because adolescents may not be able to ascertain why they are being bullied. One study found that only 17% of youth who were victimized could generally ascertain the intention of the perpetrator (Everett & Price, 1995). Indeed, it would seem difficult for a student to always know if the bullying they are experiencing is motivated by prejudice or bias against some aspect of their identity. The BullyHARM uses a behavior-based measurement strategy that includes a range of specific behaviors, which are grouped based on the ways bullying behaviors are carried out. Specific instances of discriminatory bullying can be captured within verbal bullying, which can include name calling and slurs. Discriminatory bullying involves groups who are disproportionately targeted for bullying and the reasons they are targeted—perhaps

because they are vulnerable or members of marginalized social groups (Elamé, 2013; Peguero, 2012). Demographic questions regarding identity characteristics (i.e., race/ethnicity, national origin, socioeconomic status, sex, gender identity, sexual orientation, and ability/disability status) should accompany the administration of the BullyHARM to assess groups more likely to be targeted for bullying.

Developmental Appropriateness of the Scale

Readability findings indicate that the BullyHARM is developmentally appropriate for youth with at least a sixth-grade reading level. Unfortunately, some middle school students do not have a sixth-grade reading level. Nonetheless, the scale is moderately short with only 22 items, and the items are brief and grouped by category for respondent ease. Most participants completed the survey in about 5 minutes, which is appropriate given that the attention span of the typical adolescent is 10–20 minutes (Sousa, 2011). A few participants with developmental disabilities took longer to complete the survey. In addition, a few English language learners had difficulty completing the survey because of their limited mastery of English. Finally, three participants asked what the words “offensive,” “intimidate,” and “gesture” meant during the survey administration. Thus, these words may need to be substituted for simple language in future versions of the BullyHARM. Nonetheless, in terms of respondent burden and time for administration, the BullyHARM may be more appealing for young respondents when compared to some other bullying measures, which contain 50–100 items (e.g., Social Bullying Involvement Scales; Fitzpatrick & Bussey, 2011) or long and complex definitions of bullying that respondents must bear in mind while completing a measure (e.g., Olweus Bully/Victim Questionnaire; Olweus, 1996).

Interpretability and Understandability of the Scale

Cognitive measurement testing results suggest that students interpreted and understood the BullyHARM instructions and items as intended. The item pool was revised multiple times based on recommendations from experts and adolescent participants for clarity, conciseness, and coherence. During the web survey data collection, participants were instructed to raise their hands if they had questions or problems regarding the survey. The vast majority of participants had no problems completing the survey. As described above, a few participants had trouble reading and understanding a few select words in the survey. Nonetheless, revisions related to specific words and wording in the scale could simplify the measure so that it could be more easily understood by students with reading comprehension challenges.

Dimensions and Factor Structure of the Scale

Findings support the factor structure of the BullyHARM and show good psychometric properties for the instrument. Results of reliability testing indicate that the scale and subscales have very good internal consistency reliability. In addition, all of the standardized factor loadings for the items on their respective factors are high (i.e., $\geq .82$) in the final model, which indicates that the data support the factor structure of the BullyHARM. The moderately high interfactor correlations suggest that there are close relationships between scores on the factors; however, model fit results indicate that the first-order factor model structure fits the data well and demonstrated better fit than the second-order factor model. This suggests that although the bullying factors are related, there are also meaningful differences between the types of bullying behaviors. Physical bullying, verbal bullying,

relational/social bullying, cyber-bullying, property bullying, and sexual bullying are all similar in that they are all aggressive behaviors that cause harm. However, they are also behaviors that are qualitatively different from one another and each type may be fueled by various causal factors and produce various negative effects.

Limitations

This study has several limitations. First, there may have been social desirability response bias where participants underreported experiences of bullying. Collecting data in the school's computer class was not the most private setting for collecting information about a sensitive topic like bullying. Second, the students who participated in cognitive measurement testing were high school students and not middle school students; thus, certain wording issues that came up during the data collection with middle school students (e.g., questions about the word "gesture") could have been prevented if cognitive testing had also included middle school students. Third, grouping the items based on presumed subscales during data collection may have led to some response bias. Finally, the sampling strategy is a limitation because the sample is not representative of the population of middle school students.

Future Research

The validation of a scale is an ongoing process where the measure is evaluated in different ways with different samples (Bowen, 2008). Future studies should investigate properties related to score performance validation and practice-related validation. Examining score performance validation would involve comparing scores on the BullyHARM with other established bullying instruments and measures of constructs that are highly related to bullying victimization (e.g., mental distress or social isolation). Test-retest reliability of the BullyHARM should also be examined.

Another area for future research is investigating elements of practice-related validation. School-based research demands practical measures that can be used efficiently under real-world conditions. Evidence regarding the practicality and utility of the BullyHARM should be collected from educators and school-based evaluation professionals to ensure that the instrument ultimately benefits students by making schools safer.

Finally, a perpetration version of the scale could also be developed in the future. This would involve keeping most of the primary elements of the measure intact and making changes to the wording of the instructions, item stem, and items to reflect bullying perpetration rather than victimization. Also, Spanish versions of the scale should also be developed and evaluated.

Implications for Practice

Bullying is a prevalent and persistent problem that threatens the mission of schools by contributing to educational and psychological problems among students. To inform intervention initiatives and to evaluate their effectiveness, bullying should be assessed on a regular basis. Measuring bullying in the first few months of the school year would provide a baseline, and subsequent measurements during the school year could be timed to evaluate specific interventions implemented (e.g., changes in the supervision of students or curricular lessons about bullying). Questions about where bullying occurs (e.g., hallways, classrooms, cafeteria, bathrooms, locker rooms, and buses) should also be

included in surveys to guide intervention planning. Questions about student demographics (i.e., race/ethnicity, national origin, socioeconomic status, sex, gender identity, sexual orientation, and ability/disability status) should also be included to assess groups who are more likely to be targeted for bullying.

Although bullying is often assessed anonymously, anonymity prevents school practitioners from identifying students who have experienced frequent bullying and need mental health services. The claim that anonymous surveys encourage more accurate reports of bullying (Solberg & Olweus, 2003) has been largely unexamined (Cornell & Bandyopadhyay, 2010). However, one study found no significant differences between rates of bullying perpetration or victimization between students randomly assigned to take a survey anonymously or not (Chan et al., 2005). In a study about illicit behavior (e.g., stealing), there were virtually no differences in responses between adolescents who completed an anonymous survey and adolescents who completed a survey where they were assured that their responses would be held in confidence and their names could be connected to their responses via identification numbers by only the assessment team (O'Malley, Johnson, Bachman, & Schulenberg, 2000). Based on this evidence, confidential but not completely anonymous surveys may be sufficient for protecting students' privacy and linking students who have been bullied to mental health services.

Decision making about which students have been victimized by bullying and need assistance is not easy. There is little consensus among scholars about who can be classified as a victim of bullying (Swearer, Siebecker, Johnsen-Frericichs, & Wang, 2010). For example, Solberg and Olweus (2003) used a cutoff point of being bullied at least two or three times per month, whereas Srabstein, McCarter, Shao, and Huang (2006) used a cutoff point of at least once a week to identify victims. Other researchers have classified students with bullying victimization scores falling in the 75th percentile as victims (Kokkinos & Panayiotou, 2004). Given that time and resources for mental and behavioral health services can be limited in school settings, school practitioners may decide to prioritize students with the highest bullying victimization scores and then work their way down. Bullying victimization can lead to psychological and educational problems.

Another administration issue relates to format: paper questionnaire or online survey. School-based researchers and evaluators should weigh the pros and cons of each method before choosing a format. Paper questionnaires may promote student privacy and students are accustomed to filling out bubble answer sheets often used in standardized testing; however, online surveys require fewer material resources and allow for efficient administration, data collection, and data analysis. Only one study was found in the literature that compared adolescent responses about bullying from web-based surveys and paper questionnaires. The researchers found, in general, significantly higher levels of cyber-bullying victimization in the online survey, but no significant differences between administration formats regarding physical, verbal, and property bullying perpetration; physical, verbal, and property bullying perpetration victimization; and cyber-bullying perpetration (Shpiegel, Klomek, & Apter, 2015).

A final practice consideration stems from the findings showing high correlations between bullying factors yet the first-order factor model fit the data better than the second-order factor model. Thus, although bullying behaviors are related, they also differ by dimension, and each type of bullying may have a similar yet distinct etiology. Therefore, preventive interventions for bullying behaviors may be similar yet different. For example, a preventive intervention for physical bullying would likely target some similar but also different causal factors than a preventive intervention for cyber-bullying. Examining scores by the different types of bullying may allow school practitioners to develop more targeted interventions. Just as different dimensions of bullying may have different etiologies, they may also produce different effects for students who are victimized, and therefore

require different interventions and responses from school practitioners. For example, a student who is physically bullied may have different needs than a student who is verbally and socially bullied. Examining scores on the different dimensions of bullying will allow school personnel to identify students who are affected by one or more types of bullying and provide tailored interventions to address the associated negative sequelae. The final version of the BullyHARM is in Table A1 and is available for public use. Measures with evidence of reliability and validity are essential to surveil bullying and also inform and evaluate school-based interventions.

APPENDIX

Table A1

The Bullying, Harassment, and Aggression Receipt Measure (BullyHARM)

Below is a list of mean, hurtful, and offensive things other students might have done to you on purpose and you couldn't stop them. Read each statement and then pick one answer that best describes how often it happened to you.

During the past month, other students at school . . .	Not in the Past Month	1 or 2 Times in the Past Month	About 1 Time a Week	About 2 or More Times a Week
pushed or pulled on me	①	②	③	④
threw something at me	①	②	③	④
kicked or tripped me	①	②	③	④
hit, punched, or slapped me	①	②	③	④
attacked me with a weapon or object	①	②	③	④
called me a bad name	①	②	③	④
said something mean to me	①	②	③	④
said something to scare or intimidate me	①	②	③	④
tried to turn people against me	①	②	③	④
started or spread a false rumor about me	①	②	③	④
excluded me from their group	①	②	③	④
posted something bad about me on the Internet	①	②	③	④
sent me a mean e-mail, instant message, or text message	①	②	③	④
made a mean comment about me on the Internet	①	②	③	④
broke or messed up something that belonged to me	①	②	③	④
wrote or drew something bad on my things	①	②	③	④
took something of mine and wouldn't give it back	①	②	③	④
touched me in a sexual way that made me uncomfortable	①	②	③	④
wrote, drew, or said bad sexual things about me	①	②	③	④
made an offensive sexual gesture or body movement at me	①	②	③	④
tried to pull down or pull off my clothes without my permission	①	②	③	④
touched my private parts (breast, butt, or crotch) without my okay	①	②	③	④

REFERENCES

- Arseneault, L., Bowes, L., & Shakoor, S. (2010). Bullying victimization in youths and mental health problems: "Much ado about nothing"? *Psychological Medicine*, *40*, 717–729.
- Arseneault, L., Walsh, E., Trzesniewski, K., Newcombe, R., Caspi, A., & Moffitt, T. E. (2006). Bullying victimization uniquely contributes to adjustment problems in young children: A nationally representative cohort study. *Pediatrics*, *118*, 130–138.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Sons.
- Bowen, N. K. (2008). Validation. In W. A. Darity, Jr. (Ed.), *International encyclopedia of the social sciences* (2nd ed., Vol. 7, pp. 569–572). Detroit, MI: Macmillan Reference USA.
- Bowen, N. K., & Guo, S. (2012). *Structural equation modeling*. New York, NY: Oxford University Press.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Chan, J. H., Myron, R., & Crawshaw, M. (2005). The efficacy of non-anonymous measures of bullying. *School Psychology International*, *26*, 443–458.
- Cornell, D. G., & Bandyopadhyay, S. (2010). The assessment of bullying. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 265–270). New York, NY: Routledge.
- Cornell, D. G., & Brockenbrough, K. (2004). Identification of bullies and victims: A comparison of methods. *Journal of School Violence*, *3*, 63–87.
- Cornell, D. G., Sheras, P. L., & Cole, J. C. (2006). Assessment of bullying. In S. R. Jimerson & M. J. Furlong (Eds.), *Handbook of school violence and school safety* (pp. 587–602). Mahwah, NJ: Erlbaum.
- Dake, J. A., Price, J. H., & Telljohann, S. K. (2003). The nature and extent of bullying at school. *Journal of School Health*, *73*, 173–180.
- De Leeuw, E., Borgers, N., & Smits, A. (2004). Pretesting questionnaires for children and adolescents. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 409–429). Hoboken, NJ: John Wiley and Sons. doi:10.1002/0471654728.ch20
- DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks, CA: Sage.
- Dupper, D. R. (2013). LGBTQ bullying and sexual bullying in schools. In D. R. Dupper (Ed.), *School bullying: New perspectives on a growing problem* (pp. 40–55). New York, NY: Oxford University Press.
- Elamé, E. (2013). Discriminatory bullying: A new intercultural challenge. New York, NY: Springer.
- Espelage, D. L., & Swearer, S. M. (2003). Research on school bullying and victimization: What have we learned and where do we go from here? *School Psychology Review*, *32*, 365–384.
- Everett, S. A., & Price, J. H. (1995). Students' perceptions of violence in the public schools: the MetLife survey. *Journal of Adolescent Health*, *17*, 345–352.
- Fitzpatrick, S., & Bussey, K. (2011). The development of the Social Bullying Involvement Scales. *Aggressive Behavior*, *37*, 177–192.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466–491.
- Furlong, M. J., Sharkey, J. D., Felix, E. D., Tanigawa, D., & Green, J. G. (2010). Bullying assessment: A call for increased precision of self-reporting procedures. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 329–345). New York, NY: Routledge.
- Gini, G., & Pozzoli, T. (2009). Association between bullying and psychosomatic problems: A meta-analysis. *Pediatrics*, *123*, 1059–1065.
- Gladden, R. M., Vivolo-Kantor, A. M., Hamburger, M. E., & Lumpkin, C. D. (2014). *Bullying surveillance among youths: Uniform definitions for public health and recommended data elements, Version 1.0*. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, and U.S. Department of Education.
- Goldbaum, S., Craig, W. M., Pepler, D., & Connolly, J. (2007). Developmental trajectories of victimization: Identifying risk and protective factors. In J. E. Zins, M. J. Elias, & C. A. Maher (Eds.), *Bullying, victimization, and peer harassment: A handbook of prevention and intervention* (pp. 143–160). New York, NY: Haworth Press.
- Green, J. G., Felix, E. D., Sharkey, J. D., Furlong, M. J., & Kras, J. E. (2013). Identifying bully victims: Definitional versus behavioral approaches. *Psychological Assessment*, *25*, 651–657.
- Greif, J. L., & Furlong, M. J. (2006). The assessment of school bullying: Using theory to inform practice. *Journal of School Violence*, *5*, 33–50.
- Griffin, R. S., & Gross, A. M. (2004). Childhood bullying: Current empirical findings and future directions for research. *Aggression and Violent Behavior*, *9*, 379–400.
- Guerin, S., & Hennessy, E. (2002). Pupils' definitions of bullying. *European Journal of Psychology of Education*, *17*, 249–261.

- Hall, W. J., & Rounds, K. A. (2013). Adolescent health. In the Public Health Social Work Section of the American Public Health Association, R. H. Keefe, & E. T. Jurkowski (Eds.), *Handbook for public health social work* (pp. 59–80). New York, NY: Springer.
- Hamburger, M. E., Basile, K. C., & Vivolo, A. M. (2011). *Measuring bullying victimization, perpetration, and bystander experiences: A compendium of assessment tools*. Atlanta, GA: Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, Division of Violence Prevention.
- Hamby, S. L., & Finkelhor, D. (2000). The victimization of children: Recommendations for assessment and instrument development. *Journal of the American Academy of Child and Adolescent Psychiatry*, 39, 829–840.
- Kaiser, B., & Rasminsky, J. S. (2009). Bullying. In B. Kaiser & J. S. Rasminsky (Eds.), *Challenging behavior in elementary and middle school* (pp. 280–306). Upper Saddle River, NJ: Pearson.
- Kann, L., Kinchen, S., Shanklin, S. L., Flint, K. H., Kawkins, J., Harris, W. A., . . . Zaza, S. (2014). Youth Risk Behavior Surveillance—United States, 2013. *Morbidity and Mortality Weekly Report*, 63, 1–168.
- Kert, A. S., Coddling, R. S., Tryon, G. S., & Shiyko, M. (2010). Impact of the word "bully" on the reported rate of bullying behavior. *Psychology in the Schools*, 47, 193–204.
- Kimmel, M. S., & Mahler, M. (2003). Adolescent masculinity, homophobia, and violence: Random school shootings, 1982–2001. *American Behavioral Scientist*, 46, 1439–1458.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York, NY: Guilford.
- Klomek, A. B., Sourander, A., & Gould, M. (2010). The association of suicide and bullying in childhood to young adulthood: A review of cross-sectional and longitudinal research findings. *Canadian Journal of Psychiatry*, 55, 282–288.
- Kokkinos, C. M., & Panayiotou, G. (2004). Predicting bullying and victimization among early adolescents: Associations with disruptive behavior disorders. *Aggressive Behavior*, 30, 520–533.
- Kowalski, R. M., Limber, S. P., & Agaston, P. W. (2012). *Cyber bullying: Bullying in the digital age* (2nd ed.). Malden, MA: Wiley-Blackwell.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130–149.
- Muthén, L. K., & Muthén, B. O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.
- Nakamoto, J., & Schwartz, D. (2010). Is peer victimization associated with academic achievement? A meta-analytic review. *Social Development*, 19, 221–242.
- Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth. *Journal of the American Medical Association*, 285, 2094–2100.
- Olweus, D. (1993). *Bullying: What we know and what we can do*. Oxford, England: Blackwell Publishers.
- Olweus, D. (1996). *The revised Olweus Bully/Victim Questionnaire*. Mimeo. Bergen, Norway: Research Center for Health Promotion (HEMIL Center), University of Bergen.
- Olweus, D. (2013). School bullying: Development and some important challenges. *Annual Review of Clinical Psychology*, 9, 751–780.
- O'Malley, P. M., Johnston, L. D., Bachman, J. G., & Schulenberg, J. (2000). A comparison of confidential versus anonymous survey procedures: Effects on reporting of drug use and related attitudes and beliefs in a national study of students. *Journal of Drug Issues*, 30, 35–54.
- Peguro, A. A. (2012). Schools, bullying, and inequality: Intersecting factors and complexities with the stratification of youth victimization at school. *Sociology Compass*, 6, 402–412.
- Reijntjes, A., Kamphuis, J. H., Prinzie, P., Boelen, P. A., Van der Schoot, M., & Telch, M. J. (2011). Prospective linkages between peer victimization and externalizing problems in children: A meta-analysis. *Aggressive Behavior*, 37(3), 215–222.
- Reijntjes, A., Kamphuis, J. H., Prinzie, P., & Telch, M. J. (2010). Peer victimization and internalizing problems in children: A meta-analysis of longitudinal studies. *Child Abuse & Neglect*, 34, 244–252.
- Rivers, I., & Duncan, N. (2013). *Bullying: Experiences and discourses of sexuality and gender*. New York, NY: Routledge.
- Sampson, R. (2009). A guide to bullying in schools. In D. N. Rickler (Ed.), *Bullying in schools* (pp. 1–41). New York, NY: Nova Science Publishers.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23–74.
- Shpiegel, Y., Klomek, A. B., & Apter, A. (2015). Bullying, cyberbullying, depression and suicide ideation among youth: Comparing online to paper-and-pencil questionnaires. *International Journal of Child and Adolescent Health*, 8, 161–167.
- Shute, R., Owens, L., & Slee, P. (2008). Everyday victimization of adolescent girls by boys: Sexual harassment, bullying or aggression? *Sex Roles*, 58, 477–489.
- Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggressive Behavior*, 29, 239–268.
- Sousa, D. A. (2011). *How the brain learns* (4th ed.). Thousand Oaks, CA: Corwin.

- Srabstein, J. C., McCarter, R. J., Shao, C., & Huang, Z. J. (2006). Morbidities associated with bullying behaviors in adolescents. School based study of American adolescents. *International Journal of Adolescent Medicine and Health*, 18, 587–596.
- Sveinsson, A. V., & Morris, R. J. (2007). Conceptual and methodological issues in assessment and intervention with school bullies. In J. E. Zins, M. J. Elias, & C. A. Maher (Eds.), *Bullying, victimization, and peer harassment: A handbook of prevention and intervention* (pp. 9–26). New York, NY: Haworth Press.
- Swearer, S. M., Siebecker, A. B., Johnsen-Frerichs, L. A., & Wang, C. (2010). Assessment of bullying/victimization: The problem of comparability across studies and across methodologies. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 305–327). New York, NY: Routledge.
- Tofi, M. M., Farrington, D. P., Lösel, F., & Loeber, R. (2011). Do victims of school bullies tend to become depressed later in life? A systematic review and meta-analysis of longitudinal studies. *Journal of Aggression, Conflict, and Peace Research*, 3, 63–73.
- U.S. Department of Education. (2015). *Student reports of bullying and cyber-bullying: Results from the 2013 School Crime Supplement to the National Crime Victimization Survey*. Washington, DC: National Center for Education Statistics.
- Vaillancourt, T., McDougall, P., Hymel, S., Krygsman, A., Miller, J., Stiver, K., & Davis, C. (2008). Bullying: Are researchers and children/youth talking about the same thing? *International Journal of Behavioral Development*, 32, 486–495.
- Varjas, K., Dew, B., Marshall, M., Graybill, E., Singh, A., Meyers, J., & Birckbichler, L. (2008). Bullying in schools towards sexual minority youth. *Journal of School Violence*, 7, 59–86.
- 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.
- Wang, J., Iannotti, R. J., & Nansel, T. R. (2009). School bullying among adolescents in the United States: Physical, verbal, relational, and cyber. *Journal of Adolescent Health*, 45, 368–375.
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.