

A Psychometric Comparison of Two Self-Report Measures of Bullying and Victimization: Differences by Sex and Grade

Cynthia M. Hartung · Carolyn S. Little ·
Erica K. Allen · Melanie C. Page

Published online: 11 January 2011
© Springer Science + Business Media, LLC 2010

Abstract Numerous rating scales have been designed to assess bullying and victimization; however, the psychometric properties of these measures have been understudied. As a result, the accuracy of evaluations that determine the effectiveness of bullying policies or prevention programs is uncertain. The current study was designed to (a) compare the psychometric properties of two commonly-used student self-report measures of bullying and victimization in 3rd through 5th graders, (b) examine the psychometric properties of these two measures by sex and grade, and (c) examine sex and grade differences in levels of bullying and victimization. Participants were 532 elementary school students in Grades 3 through 5. Students completed the Revised Olweus Bully/Victim Questionnaire (OBVQ) and the Reynolds Bully-Victimization Scales (BVS). Measures were completed during guidance classes. In general, students admitted to more victimization than bullying. Further, internal consistency analyses indicated that the BVS had better reliability than the OBVQ which is likely a result of more items on BVS subscales than on OBVQ subscales. A number of interesting sex and grade differences were also identified. Implications for the assessment of school bullying and victimization are discussed.

Keywords Bullying · Victimization · Student self-report · Measurement · Elementary school

C. M. Hartung (✉) · C. S. Little · E. K. Allen
Department of Psychology, University of Wyoming,
1000 E. University Avenue, Dept. #3415, Laramie,
WY 82071, USA
e-mail: chartung@uwyo.edu

M. C. Page
Department of Psychology, Oklahoma State University,
Stillwater, OK, USA

Introduction

Bullying is a pervasive problem in the United States and internationally (Nansel et al., 2001). Bullying is defined as repeated exposure to a negative action, such as physical or verbal aggression, by one or more individuals, where an imbalance of power exists between those involved (Limber & Small, 2003; Olweus, 1993). This imbalance of power can be physical or social such as differences in physical stature or popularity (Olweus, 1993). Direct bullying includes physical aggression (e.g., hitting or shoving) and verbal aggression (e.g., name-calling, shouting, or accusing). Indirect bullying, or relational aggression, includes the infliction of emotional pain through social isolation, group exclusion, or manipulation of relationships (Crick & Grotpeter, 1996; Olweus, 1993).

Although assessment tools exist for measuring bullying and victimization, the psychometric properties of these measures have been understudied (Hartung & Scambler, 2007; Kyriakides, Kaloyirou, & Lindsay, 2006). For some measures, the psychometric properties are available but have not been published in peer-reviewed journals. For other measures, the psychometric properties have been published but the findings have not been replicated across laboratories (Crick & Grotpeter, 1996; Hartung & Scambler, 2007). Therefore, additional studies of the psychometric properties of these measures are needed. Further, some researchers have found sex and grade differences in levels of bullying and victimization (e.g., Casey-Cannon, Hayward, & Gowen, 2001; Nansel et al., 2001); however, few researchers have examined sex and grade differences in psychometric properties. Sex and grade differences in levels of bullying and victimization should be interpreted cautiously until the psychometric properties of measures across grade and sex have been confirmed.

The current study was designed to begin addressing the need for more research on the psychometric properties of these measures by comparing the reliability and validity of two commonly-used student self-report rating scales. In addition, sex and grade differences in psychometric properties are examined. Finally, sex and grade differences in levels of self-reported bullying and victimization are explored.

Sex and Grade Differences in Levels of Bullying and Victimization

Sex and grade differences in bullying and victimization appear to vary depending on the type of bullying involved. Boys typically report more victimization and more bullying than girls (Kyriakides et al., 2006; Nansel et al., 2001). Pepler et al. (2006) indicated that boys reported more bullying and sexual harassment than girls. Researchers have consistently shown that boys displayed higher rates of direct physical aggression than girls (Côté, Vaillancourt, Barker, Nagin, & Tremblay, 2007; Lindeman, Harakka, & Keltikangas-Jarvinen, 1997; Olweus, 1993) whereas girls and boys have shown similar rates of direct verbal aggression (Nansel et al., 2001). In contrast, girls have displayed higher rates of relational or indirect bullying than boys (Casey-Cannon et al., 2001; Murray-Close, Ostrov, & Crick, 2007; Nansel et al., 2001; Perry, Kusel, & Perry, 1988; van der Wal, de Wit, & Hirasing, 2003). In a longitudinal study, Côté et al. (2007) followed children from ages 2 to 8. They reported that boys showed higher levels of physical aggression than girls; whereas girls were more likely to show decreasing physical aggression and increasing indirect aggression during this period of development.

For grade differences in bullying, the results are mixed. Olweus (1993) reported that levels of bullying by boys increased slightly from second through ninth grades; whereas levels of bullying by girls decreased. Nansel et al. (2001) found that middle school youth (6th through 8th graders) reported more bullying than high school youth (9th through 10th graders). In contrast, Pepler et al. (2006) found that middle school youth (6th through 8th graders) reported less bullying than high school youth (9th through 12th graders). Some researchers have suggested that the relation between bullying and grade may be qualified by the type of bullying. Specifically, relational aggression has been shown to increase in girls during elementary school (Murray-Close et al., 2007). In addition, Perry et al. (1988) found that direct physical and verbal bullying peaks in middle school but that indirect or relational bullying does not peak until high school. Thus, grade differences in bullying may be qualified by sex and type of bullying.

For grade differences in victimization, a negative correlation has been shown with students in lower grades

reporting higher rates of victimization than students in higher grades (Dennis & Satcher, 1999; Embry & Luzzo, 1996; Olweus, 1993). Specifically, Olweus (1993) reported dramatic decreases in victimization for boys and girls from second through ninth grades. Similarly, Dennis and Satcher (1999) found that third graders reported being victims of name-calling more often than fifth graders. In addition, Embry and Luzzo (1996) found that second graders reported being victims of name-calling more often than sixth graders. Grade differences for victimization are more conclusive than for bullying with researchers consistently reporting that victimization is negatively correlated with grade.

Tools for Measuring Bullying and Victimization

Multiple techniques including (a) structured behavioral observations, (b) structured interviews, (c) peer and teacher nominations, and (d) student, parent and teacher rating scales are available for measuring bullying and victimization. Each of these procedures has advantages and disadvantages. Specifically, structured observations can be highly informative because verbal and relational bullying frequently occur when adults cannot overhear the exchanges. Although overt behavioral observations are prone to participant reactivity, covert observations have also been conducted. For example, Pepler and Craig (1995) collected unique data on bullying and victimization by using hidden video-recording devices overlooking playgrounds. Although this technique provided very useful information, informed consent may be required and school-wide parental consent is difficult to obtain (Crothers & Levinson, 2004; Pepler & Craig, 1995). Furthermore, behavioral observations require intensive time for coding behaviors, either live or videotaped (Espelage & Swearer, 2003; Hartung & Scambler, 2007).

Peer and teacher nominations, when used as a measure of bullying and victimization, typically involve asking students or teachers to match descriptors with students in the class (Espelage & Swearer, 2003; Ortega et al., 2001). For example, each student would be presented with a roster of all the students in the class and then asked to identify students who (a) bully, hit or tease others and/or (b) are frequently teased or harassed by others. Students are then categorized as bullies, victims, or bully-victims based on the total number of nominations. Peer and teacher nomination methods are a time-efficient procedure for identifying bullies and victims in schools; however, these procedures also require informed consent from all parents which makes them less practical (Espelage & Swearer, 2003). Furthermore, peer and teacher nominations typically result in categorical identification of students as bullies or victims but do not provide dimensional bullying and

victimization scores (Solberg & Olweus, 2003). It is more difficult to measure change over time, and in response to interventions, using categorical measures rather than dimensional measures. Finally, peer nominations are most practical in elementary school when children do not typically change classes and remain with the same group for most of the day (Espelage & Swearer, 2003).

Structured interviews for measuring bullying and victimization include items designed to obtain details about bullying and victimization encounters (Crothers & Levinson, 2004). Thus, the information obtained may include quantitative and qualitative data. Structured interviews are time-consuming, however, given that students are typically interviewed individually. Further, students may be less likely to respond honestly in a face-to-face interview; they may be more forthright when completing an anonymous rating scale. Given the amount of time it would take to individually interview all children in a school, this method is rarely-used on a school-wide basis.

Rating scales are a low-cost and efficient procedure for measuring the frequency and qualitative aspects of victimization and bullying (Hartung & Scambler, 2007). Although parent and teacher rating scales are available they are not commonly-used given that parents and teachers have limited knowledge of the amount of bullying that is taking place; this is especially true for bullying that may be covert such as verbal and relational aggression (Smith & Ananiadou, 2003). Student self-report measures can be efficiently-used on a school-wide level (Silverman & Rabian, 1999). In addition, self-report measures can be used at multiple time points to assess change (Espelage & Swearer, 2003). Further, the objective scoring of rating scales minimizes the need for highly trained clinicians to be involved in administration, scoring, and interpretation (Silverman & Rabian, 1999). Given the advantages of student self-report rating scales, they will be used in the current study. The existing knowledge about the psychometric properties of two commonly-used self-report measures, which will be used in the current study, are described next.

Two Commonly-Used Student Self-Report Rating Scales

The Revised Olweus Bully/Victim Questionnaire (OBVQ; Olweus, 1993) is one of the most frequently-used measures of bullying and victimization for students in 3rd through 12th grades (Greif & Furlong, 2006; Kyriakides et al., 2006). One advantage of the OBVQ is the inclusion of items that assess multiple aspects of bullying such as location, frequency, sex of perpetrator (Ross, 1996) and a couple of items that assess relational aggression (Olweus, 2004). According to an unpublished manuscript (Olweus, 2004), the OBVQ has shown good internal consistency

reliability ($\alpha = .80$) and has evidenced construct validity. In addition, bullying and victimization subscales significantly correlate ($r = .40-.60$) with other bullying and victimization measures, respectively (Olweus, 2004). The psychometric properties of the OBVQ were examined in a Greek Cypriot population and construct validity and reliability were found to be adequate (Kyriakides et al., 2006). Although the results of the Kyriakides et al. (2006) study are important, generalizability to other countries and ethnicities may be limited. Although the OBVQ is a widely-used measure, there is limited published data on the psychometric properties.

The Reynolds Bully-Victimization Scale (BVS; Reynolds, 2003) is a promising scale designed for students in 3rd through 12th grades. According to Reynolds (2003), the BVS evidenced excellent internal consistency reliability and validity through factor analysis and construct, discriminant, and criterion-related validity. Using the standardization sample, factor analyses were conducted and a two factor solution was found (Reynolds, 2003). The coefficients for internal consistency reliability, for both bullying and victimization subscales, were excellent ($\alpha = .93$; Reynolds, 2003). These coefficients were high across sex and grade. Test-retest reliability was assessed by administering the scale 1 to 2 weeks after the first administration ($n = 207$). The test-retest reliability was considered good for both the bullying ($r = .81$) and victimization ($r = .80$) subscales (Reynolds, 2003). Criterion-related validity was assessed by comparing student self-report BVS with teacher-report BVS scores (Reynolds, 2003). There was a moderate correlation for the total sample ($r = .46, p < .001$) and for students in grades 3–6 ($r = .54, p < .001$); however, the correlation for students in grades 7–8 was not significant ($r = .24, NS$) suggesting that teacher reports may be less accurate for middle school youth than for elementary school children. Although the BVS is a promising measure, the psychometric properties are provided in the manual and have not been published in a peer-reviewed journal.

The Current Study

The current study was designed to (a) compare the psychometric properties of two commonly-used student self-report measures of bullying and victimization in 3rd through 5th graders, (b) to examine the psychometric properties of these two measures by sex and grade, and (c) to explore sex and grade differences in levels of bullying and victimization. The BVS (Reynolds, 2003) and OBVQ (Olweus, 1993) were selected because they include both bullying and victimization subscales, have shown adequate psychometric properties, and are used commonly for school-wide assessments.

Method

Participants

Participants were 532 elementary school students at six elementary schools in rural Oklahoma. The 532 children who participated were those whose parents consented to child participation. Participants included 59% of students in Grades 3 through 5. The sample included 178 third graders (91 boys, 87 girls), 158 fourth graders (87 boys, 71 girls), and 193 fifth graders (104 boys, 89 girls). The ethnic breakdown of the sample was 75.8% European American, 3.9% American Indian, 3.4% Asian American, 3.2% African American, 2.8% Hispanic/Latino, 8.8% “other,” and 2.1% chose not report their ethnicity. It is estimated that 47% of children in this school district received free or reduced price lunch (Great Schools, 2009). Given that this study required consent from parents, however, participants were not randomly selected. Thus, the socioeconomic status of the participants in the sample may not be consistent with the estimate for the entire school district.

Measures

The Revised Olweus Bully/Victim Questionnaire (OBVQ)

The OBVQ is a 39-item student self-report measure designed for students in Grades 3 to 12 (Olweus, 1993, 2004). Of the 39 items, 10 comprise the victimization subscale and 10 comprise the bullying subscale. These 20 items are rated on a 5-point scale (0 = *It hasn't happened to me in the past couple of months*, 1 = *only once or twice*, 2 = *2 or 3 times a month*, 3 = *about once a week*, and 4 = *about several times a week*). The remaining 19 items address details of bullying and victimization (e.g., Where does it take place? Who is the perpetrator? What efforts are made to stop or prevent bullying?) Although these items are designed to assess important qualitative aspects of bullying, they are beyond the scope of the present study which was designed to compare the bullying and victimization subscales across two measures. These items were administered, however, as part of a needs-assessment for the school district.

The first item on the OBVQ Victimization Subscale is a general question about how often the student has been the victim of bullying at school in the past couple of months. The remaining items on the OBVQ Victimization Subscale are questions about the frequency of specific forms of victimization (i.e., being called mean names, being left out, excluded or ignored; being hit, kicked, pushed, shoved, or locked indoors; having other students tell lies or spread false rumors about me; having money or other things taken

away from me or damaged; being threatened or forced to do things; being teased about my race or ethnicity, and being teased in a sexual manner). The sexual victimization item was not included for the current study at the request of the local school board; thus the Victimization Subscale for the current study consisted of 9 of 10 original OBVQ Victimization items.

The first item on the OBVQ Bullying Subscale is a general question about how often the student has been the perpetrator of bullying at school in the past couple of months. The remaining items on the OBVQ Bullying Subscale are questions about the frequency of specific forms of bullying (i.e., calling someone mean names, leaving someone out, excluding or ignoring; hitting, kicking, pushing, shoving, or locking someone indoors; telling lies or spreading false rumors about someone; taking money or other things away from someone or damaging someone's things; threatening or forcing someone to do things; teasing someone about his/her race or ethnicity, and teasing someone in a sexual manner). Again, the sexual item was not included; thus the Bullying Subscale for the current study consisted of 9 of 10 original OBVQ Bullying items.

Reynolds Bully-Victimization Scale

The BVS was designed for students in Grades 3 through 12 (Reynolds, 2003). The BVS contains 46 items including 23 bullying items and 23 victimization items. Responses are rated on a 4-point scale (0 = *never*, 1 = *sometimes*, 2 = *a lot of the time*, and 3 = *five or more times*). The BVS was standardized on a sample of 2,405 students; the sample was stratified based on sex, age, grade, race/ethnicity, and region. Therefore, normative data is available for interpreting the BVS. The BVS Victimization Subscale items were designed to measure the frequency with which the student is a victim of bullying. The BVS Bullying Subscale items were designed to measure the frequency with which the student is a perpetrator of bullying.

Procedure

The university IRB and the school board approved the procedures for the study. Teachers made consent forms available to parents during parent-teacher conferences or sent them home with students. Students were excluded from participation if their parents did not provide informed consent. Rating scale packets contained child assent forms and two rating scales. Students completed the rating scales during two regularly scheduled guidance class meetings approximately one week apart. The rating scales were administered in a counterbalanced order across grade and school. Students were identified by a participant number

and did not write their names on either rating scale. The assent forms and rating scales were read aloud by a research assistant because of students having variable reading skills. A school counselor and a clinical psychology graduate student or faculty member were also present in the event that a student became emotionally distressed while completing the rating scales. One student became distressed while completing the OBVQ and discontinued the study. No other participants discontinued the study. Completion of the BVS took approximately 15 minutes and completion of the OBVQ took approximately 30 minutes. These times are longer than would be typical if the rating scales had not been read aloud and if all students had adequate reading skills.

Results

Levels of Self-Reported Bullying and Victimization

Mean levels of self-reported bullying and victimization are reported for the total sample, and by sex, in Table 1. The possible range of scores for OBVQ subscales was 0–27. For BVS subscales the possible range of scores was 0–69. One-way ANOVAs were conducted to compare mean scores on bullying and victimization subscales within measure for the total sample. The assumptions of ANOVA were examined for the dependent variables. For the total sample and some subgroups (by sex or grade) the assumption of normality was violated because kurtosis was high. Although protection is provided for violating this assumption due to our large cell sizes (Keppel & Wickens, 2007), a more conservative alpha of .01 was adopted for all ANOVA analyses to account for this violation. ANOVAs were used rather than t-tests because partial-eta squared can be requested for ANOVAs in SPSS as a measure of effect size. Self-reported levels of BVS Victimization ($M = 13.03$, $SD = 13.13$) were significantly higher than self-reported levels of BVS Bullying ($M = 3.14$, $SD = 6.69$), $F(1, 531) = 328.11$, $p < .001$. Similarly, self-reported levels of OBVQ Victimization ($M = 5.65$, $SD = 6.27$) were significantly higher than self-reported levels of OBVQ Bullying ($M = 1.34$, $SD = 3.03$), $F(1, 531) = 265.63$, $p < .001$.

Next, because the two measures had different ranges of scores, standardized variables were created to allow comparisons of levels of bullying and victimization across measures. Again, one-way ANOVAs were conducted. Self-reported levels of BVS Bullying were not significantly different than self-reported levels of OBVQ Bullying. Similarly, self-reported levels of BVS Victimization were not significantly different than self-reported levels of OBVQ Victimization.

Sex Differences

One-way ANOVAs were conducted to examine possible sex differences in self-reported levels of bullying and victimization. There were no significant sex differences in levels of BVS Victimization, OBVQ Bullying or OBVQ Victimization (see Table 1). For levels of BVS Bullying, there was a marginally significant sex difference with boys admitting to more bullying than girls, $F(1, 528) = 4.77$, $p = .029$.

Grade Differences

Mean levels of self-reported bullying and victimization are detailed by grade in Table 2. Again, one-way ANOVAs were conducted to examine possible grade differences in levels of bullying and victimization. Tukey's post-hoc comparisons were conducted to examine significant differences among the three grades. There were no significant grade differences in levels of self-reported BVS Bullying or OBVQ Bullying. However, there was a significant effect of grade for BVS Victimization, $F(2, 531) = 27.50$, $p < .001$. Specifically, third graders reported significantly higher levels of victimization than fourth ($p = .001$) and fifth graders ($p < .001$). In addition, fourth graders reported significantly higher levels of victimization than fifth graders ($p = .002$). Similarly, there was a significant effect of grade for OBVQ Victimization, $F(2, 531) = 9.26$, $p < .001$. Again, third graders reported significantly higher levels of victimization than fifth graders ($p < .001$); however, differences between third and fourth graders and fourth and fifth graders were not statistically significant.

Factor Analysis

A principal components analysis with a varimax rotation was conducted for each measure to determine whether bullying and victimization items would cluster as separate subscales. As expected, a two-factor solution resulted for the OBVQ. Eigenvalues were 5.58 and 3.07 accounting for 30.99% and 17.04% of the variance, respectively. Evaluation of the items on each factor suggested separation of the variables into two distinct subscales: victimization and bullying (factor loadings ranged from .59 to .76 and from .55 and .75, respectively).

Additionally, a two-factor solution resulted for the BVS. Eigenvalues were 12.84 and 6.94 accounting for 27.92% and 15.08% of the variance, respectively. Furthermore, evaluation of the items on each factor for this measure also resulted in two distinct subscales: victimization and bullying (factor loadings ranged from .36 and .79 and from .37 and .75, respectively).

Table 1 Mean levels of self-reported bullying and victimization for the total sample and by sex

	Total sample <i>n</i> = 532		Girls <i>n</i> = 247		Boys <i>n</i> = 282		ANOVA for sex <i>df</i> = 528		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>p</i>	η^2
BVS bullying (23 items)	3.14	6.69	2.47	4.69	3.74	8.03	4.77	.029	.00
BVS victimization (23 items)	13.03	13.13	13.41	13.52	12.60	12.47	0.51	.474	.00
OBVQ bullying (9 items)	1.34	3.03	1.27	2.84	1.41	3.20	0.28	.597	.00
OBVQ victimization (9 items)	5.65	6.27	6.02	6.60	5.29	5.95	1.01	.179	.00

Alpha of .01 was used to account for kurtosis violation

Table 2 Mean levels of self-reported bullying and victimization by grade

	3rd Grade <i>n</i> = 180		4th Grade <i>n</i> = 159		5th Grade <i>n</i> = 196		ANOVA for grade <i>df</i> = 531		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	<i>p</i>	η^2
BVS bullying (23 items)	3.23 ^a	5.59	3.20 ^a	7.15	3.01 ^a	7.25	0.06	.939	.00
BVS victimization (23 items)	18.02 ^a	14.92	13.02 ^b	12.15	8.40 ^c	10.12	27.50*	.000	.09
OBVQ bullying (9 items)	1.29 ^a	3.18	1.28 ^a	2.51	1.42 ^a	3.28	0.12	.884	.00
OBVQ victimization (9 items)	7.00 ^a	7.21	5.80 ^{ab}	5.87	4.26 ^b	5.30	9.26*	.000	.03

Means with no common superscripts are significantly different from one another. Alpha of .01 was used to account for kurtosis violation

* *p* < .01

For both the BVS and OBVQ, no items had loadings of .40 or higher on both bullying and victimization subscales. Therefore, across both measures, there was no evidence of cross-loadings. This finding suggests that bullying and victimization are separate constructs.

Internal Consistency Reliability

To establish internal consistency, Cronbach’s alpha was calculated for the bullying and victimization subscales. Cronbach’s alpha values of .69 or lower are referred to as “inadequate,” .70 to .79 are referred to as “adequate,” .80 to .89 are referred to as “good,” and .90 or higher as “excellent” (Charter, 2003; Henson, 2001). These values are shown on the diagonals in Table 3 for the total sample and separately for girls and boys. Similarly, alpha values are shown on the diagonals in Table 4 by grade. For the total sample, internal consistency was good for the OBVQ subscales and excellent for the BVS subscales. Across sex and grade, internal consistency was also good to excellent across measures and subscales with one exception. Specifically, internal consistency for OBVQ Bullying in fourth graders was only adequate ($\alpha = .79$).

For each of the four subscales (i.e., two subscales from each of two measures), comparisons of alpha values were conducted (a) between measures, (b) between sexes and (c) among grades. To test significant differences between alpha values, Fisher’s *z* transformations were used to compare independent correlations (Cohen, Cohen, West, & Aiken,

2003). For the total sample, and when examining differences between measures or sexes, one comparison was made for each construct or subscale. Therefore, an alpha value of .05 was used. When examining grade differences, three comparisons were made for each subscale (i.e., 3rd vs. 4th, 4th vs. 5th, and 3rd vs. 5th). A Bonferroni correction resulted in significance level of .017. Internal consistency comparisons by measure are shown in Table 5. Internal consistency comparisons by sex and grade but within measure are shown in Table 6.

Differences Across Measures

For the total sample, the internal consistency reliability for BVS Bullying was significantly higher than for OBVQ Bullying ($z = 6.96, p < .001$). This pattern held for boys, girls, fourth and fifth graders. However, for third graders, the internal consistency reliability for BVS Bullying was not significantly higher than for OBVQ Bullying. For the total sample, the internal consistency reliability for BVS Victimization was significantly higher than for OBVQ Victimization ($z = 7.06, p < .001$). This pattern held for boys, girls, third, fourth and fifth graders.

Sex Differences Within Measure

For OBVQ Bullying and Victimization, there was good reliability for both sexes and no significant sex differences.

Table 3 Multi-trait, Multi-method Matrix for Bullying and Victimization Subscales for Total Sample and by Sex

	Total sample				Girls				Boys			
	OBVQ bully	OBVQ victim	BVS bully	BVS victim	OBVQ bully	OBVQ victim	BVS bully	BVS victim	OBVQ bully	OBVQ victim	BVS bully	BVS victim
OBVQ Bully	(.84) ^a				(.85) ^a				(.84) ^a			
	<i>n</i> = 521				<i>n</i> = 242				<i>n</i> = 276			
OBVQ Victim	.30 ^c	(.86) ^a			.41 ^c	(.88) ^a			.22 ^c	(.85) ^a		
	<i>n</i> = 532	<i>n</i> = 515			<i>n</i> = 247	<i>n</i> = 238			<i>n</i> = 282	<i>n</i> = 274		
BVS Bully	.59 ^b	.23 ^c	(.93) ^a		.47 ^b	.27 ^c	(.89) ^a		.66 ^b	.23 ^c	(.94) ^a	
	<i>n</i> = 532	<i>n</i> = 532	<i>n</i> = 499		<i>n</i> = 247	<i>n</i> = 247	<i>n</i> = 234		<i>n</i> = 282	<i>n</i> = 282	<i>n</i> = 262	
BVS Victim	.26 ^c	.69 ^b	.33 ^c	(.94) ^a	.27 ^c	.72 ^b	.41 ^c	(.94) ^a	.26	.66 ^b	.33 ^c	(.93) ^a
	<i>n</i> = 532	<i>n</i> = 532	<i>n</i> = 532	<i>n</i> = 500	<i>n</i> = 247	<i>n</i> = 247	<i>n</i> = 247	<i>n</i> = 234	<i>n</i> = 282	<i>n</i> = 282	<i>n</i> = 282	<i>n</i> = 263

All of the within-trait correlations were significant with $p < .001$. *Bully* Bullying subscale, *Victim* Victimization subscale

^a Alpha value measures internal consistency reliability

^b Within-trait, cross-method correlation measures convergent validity of constructs (i.e., bullying and victimization)

^c Cross-trait correlations are compared to within-trait correlations as a measure of discriminant validity

Table 4 Multi-trait, multi-method matrix for bullying and victimization subscales by grade

	3rd Grade				4th Grade				5th Grade			
	OBVQ bully	OBVQ victim	BVS bully	BVS victim	OBVQ bully	OBVQ victim	BVS bully	BVS victim	OBVQ bully	OBVQ victim	BVS bully	BVS victim
OBVQ Bully	(.86) ^a				(.79) ^a				(.86) ^a			
	<i>n</i> = 177				<i>n</i> = 153				<i>n</i> = 191			
OBVQ Victim	.48 ^c	(.89) ^a			.33 ^c	(.83) ^a			.09 ^c	(.83) ^a		
	<i>n</i> = 180	<i>n</i> = 174			<i>n</i> = 159	<i>n</i> = 155			<i>n</i> = 193	<i>n</i> = 186		
BVS Bully	.47 ^b	.40 ^c	(.88) ^a		.63 ^b	.27 ^c	(.93) ^a		.67 ^b	.05 ^c	(.95) ^a	
	<i>n</i> = 180	<i>n</i> = 180	<i>n</i> = 159		<i>n</i> = 159	<i>n</i> = 159	<i>n</i> = 153		<i>n</i> = 193	<i>n</i> = 193	<i>n</i> = 187	
BVS Victim	.35 ^c	.68 ^b	.47 ^c	(.94) ^a	.24 ^c	.67 ^b	.37 ^c	(.93) ^a	.23 ^c	.68 ^b	.22 ^c	(.93) ^a
	<i>n</i> = 180	<i>n</i> = 180	<i>n</i> = 180	<i>n</i> = 162	<i>n</i> = 159	<i>n</i> = 159	<i>n</i> = 159	<i>n</i> = 154	<i>n</i> = 193	<i>n</i> = 193	<i>n</i> = 193	<i>n</i> = 184

All of the within-trait correlations were significant with $p < .001$. *Bully* Bullying subscale, *Victim* Victimization subscale

^a Alpha value measures internal consistency reliability

^b Within-trait, cross-method correlation measures convergent validity of constructs (i.e., bullying and victimization)

^c Cross-trait correlations are compared to within-trait correlations as a measure of discriminant validity

For BVS Bullying and Victimization, there was very good to excellent reliability for both sexes. There were no significant sex differences for BVS Victimization. However, for BVS Bullying, reliability was significantly higher for boys than for girls ($z = 3.61$, $p < .001$).

Grade Differences Within Measure

For BVS Victimization, there was excellent reliability and no significant differences among grades. For BVS Bullying, there was good to excellent reliability across grades, but there were some significant grade differences among reliabilities. Specifically, reports from fourth and fifth graders resulted in significantly higher reliabilities than reports from third graders ($z = 2.57$, $p < .001$ and

$z = 4.37$, $p < .001$, respectively); however, the difference between the reliabilities of fourth and fifth graders' reports did not reach significance.

For OBVQ Bullying, there was adequate to good reliability for all three grades and there were no significant grade differences. For OBVQ Victimization, third graders had significantly higher reliability than fifth graders ($z = 2.24$, $p = .013$). The difference between alpha values for third and fourth graders was only marginally significant ($p = .017$) and the difference for fourth and fifth graders was not significant.

Convergent Validity

Next, multitrait-multimethod (MTMM) matrices were created to allow examination of convergent and

Table 5 Comparisons of internal consistency values within construct across measure

Alpha value 1	Alphas value 2	Fisher's z	One-tailed p -value
Total sample BVS bullying ($\alpha = .93$)	Total sample OBVQ bullying ($\alpha = .84$)	6.96*	<.001
Total sample BVS victimization ($\alpha = .94$)	Total sample OBVQ victimization ($\alpha = .86$)	7.06*	<.001
Boys BVS bullying ($\alpha = .94$)	Boys OBVQ bullying ($\alpha = .84$)	5.96*	<.001
Boys BVS victimization ($\alpha = .93$)	Boys OBVQ victimization ($\alpha = .85$)	4.63*	<.001
Girls BVS bullying ($\alpha = .89$)	Girls OBVQ bullying ($\alpha = .85$)	1.80	.036
Girls BVS victimization ($\alpha = .94$)	Girls OBVQ victimization ($\alpha = .88$)	3.91*	<.001
3rd BVS bullying ($\alpha = .88$)	3rd OBVQ bullying ($\alpha = .86$)	0.75	.227
3rd BVS victimization ($\alpha = .94$)	3rd OBVQ victimization ($\alpha = .89$)	2.87	.002
4th BVS bullying ($\alpha = .93$)	4th OBVQ bullying ($\alpha = .79$)	5.03*	<.001
4th BVS victimization ($\alpha = .93$)	4th OBVQ victimization ($\alpha = .83$)	4.09*	<.001
5th BVS bullying ($\alpha = .95$)	5th OBVQ bullying ($\alpha = .86$)	5.55*	<.001
5th BVS victimization ($\alpha = .93$)	5th OBVQ victimization ($\alpha = .83$)	4.49*	<.001

* Bonferroni correction applied of $p < .008$

Table 6 Comparisons of internal consistency values within construct across sex and grade

Alpha value 1	Alphas value 2	Fisher's z	One-tailed p -value
Girls OBVQ bullying ($\alpha = .85$)	Boys OBVQ bullying ($\alpha = .84$)	0.40	.345
Girls OBVQ victimization ($\alpha = .88$)	Boys OBVQ victimization ($\alpha = .85$)	1.36	.086
Girls BVS bullying ($\alpha = .89$)	Boys BVS bullying ($\alpha = .94$)	3.61*	<.001
Girls BVS victimization ($\alpha = .94$)	Boys BVS victimization ($\alpha = .93$)	0.91	.182
3rd grade OBVQ bullying ($\alpha = .86$)	4th grade OBVQ bullying ($\alpha = .79$)	2.02	.022
3rd grade OBVQ victimization ($\alpha = .86$)	5th grade OBVQ bullying ($\alpha = .86$)	0.00	.500
4th grade OBVQ bullying ($\alpha = .79$)	5th grade OBVQ victimization ($\alpha = .86$)	2.05	.020
3rd grade OBVQ victimization ($\alpha = .89$)	4th grade OBVQ victimization ($\alpha = .83$)	2.13	.017
3rd grade OBVQ victimization ($\alpha = .89$)	5th grade OBVQ victimization ($\alpha = .83$)	2.24*	.013
4th grade OBVQ victimization ($\alpha = .83$)	5th grade OBVQ victimization ($\alpha = .83$)	0.00	.500
3rd grade BVS bullying ($\alpha = .88$)	4th grade BVS bullying ($\alpha = .93$)	2.57*	.005
3rd grade BVS victimization ($\alpha = .88$)	5th grade BVS bullying ($\alpha = .95$)	4.37*	<.001
4th grade BVS bullying ($\alpha = .93$)	5th grade BVS victimization ($\alpha = .95$)	1.60	.054
3rd grade BVS victimization ($\alpha = .94$)	4th grade BVS victimization ($\alpha = .93$)	0.73	.234
3rd grade BVS victimization ($\alpha = .94$)	5th grade BVS victimization ($\alpha = .93$)	0.76	.223
4th grade BVS victimization ($\alpha = .93$)	5th grade BVS victimization ($\alpha = .93$)	0.00	.500

* Bonferroni correction applied of $p < .050$ for the total sample, girls and boys

* Bonferroni correction applied of $p < .017$ for 3rd, 4th, and 5th graders

discriminant validity (Campbell & Fiske, 1959). The two traits were bullying and victimization; the two methods were the OBVQ and BVS self-report measures. It was expected that the constructs of bullying and victimization would show good convergent validity. To examine convergent validity the within-trait, cross-method correlations (i.e., OBVQ Bullying with BVS Bullying and OBVQ Victimization with BVS Victimization) were examined. It was expected that these correlations would be statistically significant. For the total sample, including both sexes (see Table 3) and all grades (see Table 4), OBVQ and BVS Bullying were significantly correlated as were OBVQ and

BVS Victimization ($p < .001$). All correlations had medium to large effect sizes.

Comparison of Convergent Validity Within Sex and Grade

To evaluate convergent validity across traits, Fisher's z transformations were used to compare independent correlations (Cohen et al., 2003). A total of six comparisons were conducted and a Bonferroni correction resulted in an alpha value of .008. The results for the cross-trait comparisons (i.e., bullying vs. victimization) are shown in Table 7.

For the total sample, convergent validity for victimization was significantly stronger than for bullying ($z = 2.77$, $p = .003$). Similarly, for girls and third graders, convergent validity for victimization was significantly stronger than for bullying ($z = 4.39$, $p < .001$ and $z = 3.00$, $p = .001$, respectively). However, for boys, fourth and fifth graders convergent validity differences were not significant.

Comparison of Convergent Validity Across Sex and Grade

Next, to compare convergent validity across sex and grade, Fisher's z transformations were used. A total of six comparisons were conducted and a Bonferroni correction resulted in an alpha value of .006. The results for comparisons across sex and grade are shown in Table 8.

There was a significant sex difference for bullying with boys' self-reports resulting in stronger convergent validity than girls' self-reports ($z = 3.23$, $p < .001$). There was no significant sex difference in convergent validity for victimization.

A significant grade difference emerged for bullying with fifth graders' self-reports showing stronger convergent validity than third graders ($z = 2.88$, $p = .002$). There were no significant convergent validity differences between third and fourth graders or between fourth and fifth graders. For victimization, there were no significant grade differences in convergent validity.

Discriminant Validity

To evaluate discriminant validity for bullying and victimization, within-trait correlations (e.g., OBVQ Bullying with BVS Bullying) were expected to be significantly larger than cross-trait correlations (e.g., OBVQ Bullying with OBVQ Victimization and BVS Bullying with OBVQ Victimization). Because of the dependent nature of the correlations, pair-wise comparisons between correlations were conducted using Steiger's formula for dependent

correlations (Steiger, 1980). For the total sample, each within-trait correlation (i.e., one for bullying and one for victimization) was compared to four cross-trait correlations. Four comparisons were conducted for each convergent validity value; therefore, a Bonferroni correction resulted in an alpha value of .013. Results are shown in Table 9. For bullying in the total sample all four comparisons were statistically significant ($p < .001$). Similarly for victimization in the total sample, all four comparisons were statistically significant ($p < .001$). Thus, both bullying and victimization demonstrated excellent discriminant validity in the total sample.

These analyses were also conducted separately by grade and sex. For victimization, all comparisons were statistically significant ($p < .001$) for each subgroup (i.e., boys, girls, 3rd, 4th, and 5th graders). Thus, victimization demonstrated excellent discriminant validity across sex and grade.

For bullying, all comparisons were statistically significant for boys, fourth and fifth graders ($p < .001$). However, discriminant validity for bullying was not as strong in girls and third graders (see Table 9). For girls, only 2 of 4 comparisons for bullying were statistically significant. For third graders none of the four comparisons for bullying were statistically significant.

Discussion

The current study was designed to (a) compare the psychometric properties of two commonly-used student self-report measures of bullying and victimization in 3rd through 5th graders, (b) to examine the psychometric properties of these two measures by sex and grade, and (c) to explore sex and grade differences in levels of bullying and victimization. Specifically, self-reported levels of victimization and bullying were examined by sex and grade. In addition, internal consistency reliability and construct validity were compared across measures and

Table 7 Cross-measure, within-trait comparisons of convergent validity correlations within sex and grade

Correlation 1	Correlation 2	Fisher's z	One-tailed p -value
OBVQ victim with BVS victim ($r = .69$)	OBVQ bully with BVS bully ($r = .59$)	2.77*	.003
Girls OBVQ victim with BVS victim ($r = .72$)	Girls OBVQ bully with BVS bully ($r = .47$)	4.39*	<.001
Boys OBVQ victim with BVS victim ($r = .66$)	Boys OBVQ bully with BVS bully ($r = .66$)	0.00	.500
3rd grade OBVQ victim with BVS victim ($r = .68$)	3rd grade OBVQ bully with BVS bully ($r = .47$)	3.00*	.001
4th grade OBVQ victim with BVS victim ($r = .67$)	4th grade OBVQ bully with BVS bully ($r = .63$)	0.61	.270
5th grade OBVQ victim with BVS victim ($r = .68$)	5th grade OBVQ bully with BVS bully ($r = .67$)	0.18	.429

Bully Bullying subscale, Victim Victimization subscale

* Bonferroni correction applied of $p < .008$

Table 8 Cross-measure, within-trait comparisons of convergent validity correlations by sex and grade

Correlation 1	Correlation 2	Fisher's <i>z</i>	One-tailed <i>p</i> -value
Boys OBVQ bully with BVS bully ($r = .66$)	Girls OBVQ bully with BVS bully ($r = .47$)	3.23*	<.001
Boys OBVQ victim with BVS victim ($r = .66$)	Girls OBVQ victim with BVS victim ($r = .72$)	1.31	.095
3rd grade OBVQ bully with BVS bully ($r = .47$)	4th grade OBVQ bully with BVS bully ($r = .63$)	2.11	.018
3rd grade OBVQ bully with BVS bully ($r = .47$)	5th grade OBVQ bully with BVS bully ($r = .67$)	2.88*	.002
4th grade OBVQ bully with BVS bully ($r = .63$)	5th grade OBVQ bully with BVS bully ($r = .67$)	0.64	.261
3rd grade OBVQ victim with BVS victim ($r = .68$)	4th OBVQ victim w BVS victim ($r = .67$)	0.17	.434
3rd grade OBVQ victim with BVS victim ($r = .68$)	5th OBVQ victim w BVS victim ($r = .68$)	0.00	.500
4th grade OBVQ victim with BVS victim ($r = .67$)	5th OBVQ victim w BVS victim ($r = .68$)	0.17	.432

Bully Bullying subscale, *Victim* Victimization subscale

* Bonferroni correction applied of $p < .006$

Table 9 Comparisons of within-trait and cross-trait correlations for the total sample: evidence for strong discriminant validity

Correlation 1 (Within-Trait)	Correlation 2 (Cross-Trait)	Steiger's <i>t</i>	One-tailed <i>p</i> -value
OBVQ bullying with BVS bullying ($r = .59$)	OBVQ bullying with OBVQ victimization ($r = .30$)	6.56*	<.001
OBVQ bullying with BVS bullying ($r = .59$)	OBVQ bullying with BVS victimization ($r = .26$)	7.97*	<.001
OBVQ bullying with BVS bullying ($r = .59$)	OBVQ victimization with BVS bullying ($r = .23$)	8.48*	<.001
OBVQ bullying with BVS bullying ($r = .59$)	BVS bullying with BVS victimization ($r = .33$)	6.03*	<.001
OBVQ victimization with BVS victimization ($r = .69$)	OBVQ bullying with OBVQ victimization ($r = .30$)	9.84*	<.001
OBVQ victimization with BVS victimization ($r = .69$)	OBVQ bullying with BVS victimization ($r = .26$)	11.14*	<.001
OBVQ victimization with BVS victimization ($r = .69$)	OBVQ victimization with BVS bullying ($r = .23$)	12.22*	<.001
OBVQ victimization with BVS victimization ($r = .69$)	BVS bullying with BVS victimization ($r = .33$)	8.95*	<.001

* Bonferroni correction applied of $p < .013$

groups. In general, students admitted to more victimization than bullying. Further, internal consistency analyses indicated that the BVS had better reliability than the OBVQ which is likely a result of more items on BVS than OBVQ subscales. A number of interesting sex and grade differences were also identified.

Levels of Self-Reported Bullying and Victimization

For both measures students, on average, reported being a victim of bullying significantly more often than being a perpetrator of bullying. This pattern has been found in other studies (e.g., Austin & Joseph, 1996). In the current study, students of both sexes and all three grades reported more victimization than bullying. Contemplating this finding may lead one to question the accuracy of students' reports. For each incidence of bullying/victimization there must be at least one perpetrator and at least one victim. Therefore, one might argue that the number of incidents of bullying and victimization should be equal. There are several possible explanations for students reporting more incidents of victimization than bullying. First, students may

be more honest about victimization than bullying. Thus, students may be accurately reporting levels of victimization and under-reporting levels of bullying. Next, although less intuitive, students may be more honest about bullying than victimization. In this case, students may be accurately reporting levels of bullying and over-reporting levels of victimization.

Another possibility is that a few students are reporting higher levels of bullying than victimization but when the data are averaged across a large number of students the resulting means indicate higher levels of victimization than bullying. In order to test the accuracy of this last hypothesis, post-hoc analyses were conducted by calculating difference scores between levels of victimization and bullying for each student. Differences scores were calculated based on the BVS since these subscales had better reliability. We were interested in exploring whether some students would have negative scores, indicating higher levels of bullying than victimization. As expected, most students had positive difference scores indicating they reported more instances of being the victim of bullying than of being the perpetrator. Only 6.4% of students in this

sample had negative difference scores. Specifically, 34 out of 534 students admitted to being the perpetrator of bullying more often than being the victim. Thus, it is possible that students are accurately reporting bullying and victimization but that this small group of children is responsible for the majority of perpetration.

It is also possible that all three of these explanations contribute to the uneven reports of bullying and victimization. Thus, students may (a) under-report bullying and (b) over-report victimization and (c) a small number of students may be responsible for most of the bullying. Future research should include objective measures for comparison to student self-reports to help assess the accuracy of self-reports. Objective measures might include teacher reports, behavioral observations, electronic diaries (Suveg, Payne, Thomassin, & Jacob, 2010) and participant event monitoring (Peterson, Brown, Bartelstone and Kern, 1996). Suveg et al. (2010) found that electronic diaries were useful in assessing emotional states in school-age children. This model could be adapted for monitoring incidents of bullying and victimization in school-age children.

In addition, Peterson et al. (1996) argued that participant event monitoring was an appropriate method for measuring low base-rate events in children. These authors studied minor injuries in children as a model for the use of this method. Second grade children and their mothers participated in their study. Both children and mothers were trained in defining injuries and keeping detailed records regarding the type of injury, location of injury, and reactions to the injury. In addition to keeping detailed records of these low base-rate events over a 12-month period, children and mothers were interviewed individually every two weeks. During these interviews, participants were asked a series of scripted questions about injuries that may have occurred during the past two weeks. The authors concluded that participant event monitoring is a promising method for collecting data about low base-rate events in children. This procedure has the potential to be adapted for measuring bullying and victimization.

Internal Consistency Reliability

The results of the current study suggest that the BVS subscales are more reliable than the OBVQ subscales. This is most likely a function of the BVS subscales having more items than the OBVQ subscales. In addition to bullying and victimization subscales, the OBVQ includes items regarding sex of perpetrator, location of bullying, and responses to bullying from victims, teachers and parents. The BVS only includes items that comprise the bullying and victimization subscales. The additional items on the OBVQ that do not comprise the bullying and victimization

subscales were not analyzed for the current study. Nonetheless, this additional information may be very helpful for school administrators when assessing bullying and victimization in their schools and planning intervention programs. Although the OBVQ subscales are much shorter than the BVS subscales, the inclusion of the additional items on the OBVQ results in the overall length of the two measures being similar. An advantage of the BVS is that the bullying and victimization subscales have better reliability. An advantage of the OBVQ is the inclusion of the additional items regarding sex of perpetrator, location of bullying and responses to bullying (Seals & Young, 2003). Although these additional items provide important information, reliability is crucially important. Therefore, it may be useful to add more items to the OBVQ bullying and victimization subscales. Nonetheless, it may also be useful to consider adding items to BVS to examine sex of perpetrator, location of bullying and responses to bullying. Another option would be to develop a measure that has the strengths of both instruments.

Sex Differences

Several interesting sex differences were identified in the current study. First, boys' reports on the BVS Bullying subscale resulted in significantly higher internal consistency reliability than girls' reports. This finding is likely a function of boys' participating in more bullying and, therefore, having higher scores on this subscale resulting in the opportunity for better internal consistency reliability. Next, boys admitted to slightly more bullying than girls based on the BVS but not the OBVQ. This sex difference was not as strong in the current study as in other studies (e.g., Glew, Fan, Katon, Rivara, & Kernic, 2005; Kyriakides et al., 2006; Seals & Young, 2003). Another sex difference was found when examining convergent and discriminant validity. Specifically, for girls, convergent validity was weaker for bullying than for victimization. However, convergent validity was similar across traits for boys. Further, discriminant validity was weaker for bullying than for victimization in girls but not boys. Again, given that girls' reports of bullying were lower than boys, reliability and validity may have been restricted.

It has been observed that measures of bullying and victimization tend to emphasize physical over relational aggression (Sawyer, Bradshaw, & O'Brennan, 2008). The sex differences that have been found on these measures suggest a need to include items that address behaviors that are more common among girls (e.g., verbal and relational aggression). It may be prudent to emphasize physical aggression less and verbal and relational aggression more

because boys appear to display all types of bullying and aggression but girls are much less likely to display physical aggression than boys (Côté et al., 2007; Crick, Ostrov, & Werner, 2006; Murray-Close et al., 2007; Nansel et al., 2001; Owens, Shute, & Slee, 2000; Underwood, Scott, Galperin, Bjornstad, & Sexton, 2004; van der Wal et al., 2003). Given that girls are known to perpetrate relational aggression more often than physical aggression, the addition of more verbal and relational aggression items should be considered. This might result in an increase in the levels of reported bullying in girls and, as a result, increased power for reliability and validity.

Given that sex differences in levels of bullying may have resulted from girls being more likely to endorse verbal and relational aggression items, a closer examination of the BVS Bullying and OBVQ Bullying items was conducted. For BVS Bullying, 8 of 23 items were endorsed at significantly higher levels for boys than girls. For the remaining 15 items there was no difference between boys' and girls' reported levels. Of the items that resulted in sex differences, 5 of 8 involved physical aggression (e.g., I started fights with other kids; I beat up someone; I threw something at other kids to hurt them). Of the items that did not result in sex differences, 11 of 15 involved verbal aggression (e.g., I picked on younger kids; I made other kids do things for me; I called other kids names). Thus, it appears that boys and girls reported similar levels of verbal aggression whereas boys reported more physical aggression than girls. An examination of the BVS items did not reveal sex differences in relational aggression because few, if any, of the BVS items measure this type of aggression.

A similar examination of the OBVQ Bullying items provides additional support for adding verbal and relational aggression items to measures of bullying and victimization. Of the nine OBVQ Bullying items that were included in the current study, two of them are considered relational aggression items (i.e., I excluded someone from my group of friends; I spread false rumors about someone and tried to make others dislike him/her). Although there were no statistically significant sex differences on these two items, they were 2 of only 5 bullying items, on either measure, that resulted in higher levels of endorsement from girls than from boys. Other items that were endorsed at higher levels by girls, albeit not significantly, were verbal bullying items (i.e., I made other kids do things for me; I was with a group of kids who picked on other kids; I bullied someone with mean names or comments about his/her race). This suggests that these relational and verbal bullying items are particularly relevant for girls and still relevant for boys. These findings provide support for adding more verbal and relational aggression items to measures of bullying and

victimization to improve the psychometric properties for girls.

Grade Differences

A few grade differences also emerged from the current study. First, as expected, levels of reported victimization were negatively related to grade. Thus, third graders reported the highest levels of victimization and fifth graders reported the lowest levels. There were no grade differences for levels of bullying reported. As with the primary finding regarding higher levels of self-reported victimization than bullying, this grade difference for victimization may also be impacted by the accuracy of student self-reports. As mentioned earlier, students may be prone to under-report being the perpetrator of bullying and over-report being the victim of bullying. Younger students could be even more likely to over-report victimization. Again, further research should include objective comparison measures in addition to student self-reports to help us better understand the accuracy of student self-reports and response styles.

Another grade difference was related to reliability and validity. For BVS Bullying, reports from third graders resulted in significantly lower reliability than those from fourth and fifth graders. For OBVQ Victimization, in contrast, reports from third graders resulted in significantly higher reliability than those from fourth and fifth graders. Further, reports from third graders resulted in significantly stronger convergent validity for victimization than for bullying. Next, reports from third graders resulted in significantly stronger convergent validity for bullying than reports from fifth graders. Finally, discriminant validity for bullying was not as strong for third graders as it was for fourth and fifth graders. Taken together, these findings suggest that third graders' reports of bullying may not be as reliable and valid as (1) their reports of victimization and (2) fourth and fifth graders' reports of bullying. Both of the measures used in the current study were designed for use with 3rd through 12th graders. The current findings suggest that third graders' reports of bullying may not be adequately reliable and valid. Therefore, third graders' reports of bullying should be interpreted with caution. However, their reports of victimization appear to have adequate psychometric properties.

Limitations and Future Directions

In future studies, the inclusion of objective comparison measures, such as teachers' ratings, behavioral observations, participant event monitoring (Peterson et al., 1996) and electronic diaries (Suveg et al., 2010), may be highly informative. The use of measures obtained from different

reporters or procedures might result in a better understanding of the accuracy of students' self-reports. Given the sex and grade differences found in the current study, these variables should also be explicitly examined in future studies. It is important to ensure that measures of bullying and victimization have adequate psychometric properties across sex and grade. Also, in the future, it would be useful to conduct test–retest reliability analyses of these measures. This is particularly important given that many schools are implementing prevention and intervention programs and, therefore, are measuring bullying and victimization at multiple time points.

Participants in the current study were from a small college town and were predominately European American. As a result, the extent to which the results may generalize to other geographic regions and more ethnically diverse groups is unknown. The consent procedure resulted in participation from 60% of eligible students. It is not known whether certain characteristics such as IQ, ethnicity, SES, or bully-victim status might have predicted which parents allowed their children to participate.

Another limitation of the study was the removal of the OBVQ items about sexual bullying and victimization. Only 9 of 10 items for each OBVQ subscale were administered. This may have affected the reliability and validity of the OBVQ since the number of items on each subscale was already limited. Although there are currently no clear guidelines regarding the assessment of sexual bullying in children, research has suggested an increase in sexual victimization with 3% of 6- to 9-year olds reporting this type of victimization compared to 10% of 10- to 13-year olds (Finkelhor, Ormrod, & Turner, 2009). Based on the Finkelhor et al. study, researchers might consider including these items for 5th graders but omitting them for 3rd and 4th graders in future studies.

A final limitation of the current study was the reading of items to participants. The administration recommendations for the BVS suggest reading items aloud for children with learning or intellectual disabilities but not for typical children. For the school district in which the current study took place, children with disabilities were integrated into regular classrooms making reading items aloud to a subset of students impractical. It is possible that reading the items aloud increased attention to the task; however, it is also possible that reading the items aloud increased self-consciousness. If this were the case, students may have been more hesitant to admit to bullying.

Conclusions

In summary, the current study provided additional evidence for adequate psychometric properties of two commonly-

used measures of bullying and victimization. Further, this study identified a number of notable sex and grade differences that warrant further study. In particular, current measures may not be optimal for measuring perpetration of bullying in girls. The inclusion of more items designed to measure verbal and relational bullying might improve the accuracy of measures of bullying and victimization for girls. This could be accomplished by creating new measures or adding items to existing measures. Finally, the current study, like previous studies, found that students tend to admit to more victimization than bullying. It is not clear if this is a result of response styles or real differences in rates of bullying and victimization. Additional studies are needed that include student self-reports as well as other measures of bullying to determine the accuracy of self-reports. Novel techniques, such as electronic diaries (Suveg et al., 2010) and participant event monitoring (Peterson et al., 1996), could potentially be applied to the assessment of bullying and victimization. Given that bullying appears to be an ongoing, if not escalating problem in our schools, continued research is warranted.

Acknowledgments We thank the following individuals for their contributions to this project: Vanessa Brown, Christina David, David Estes, David Fedele, Melissa Jackson, Jamie Johnson, Elizabeth Lefler, Cynthia Muhamedajic, Megan Rose, Douglas Scambler, Benjamin Sigel, Meredith Sligh, Brooke Whitesell, and Benjamin Wilkowski.

References

- Austin, S., & Joseph, S. (1996). Assessment of bully/victim problems in 8 to 11 year-olds. *British Journal of Educational Psychology*, 66, 447–456.
- Campbell, D. T., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Casey-Cannon, S., Hayward, C., & Gowen, K. (2001). Middle-school girls' reports of peer victimization: Concerns, consequences, and implications. *Professional School Counseling*, 5, 138–147.
- Charter, R. A. (2003). A breakdown of reliability coefficients by test type, reliability method, and implications of low reliability. *Journal of General Psychology*, 130, 290–304.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Côté, S. M., Vaillancourt, T., Barker, E. D., Nagin, D., & Tremblay, R. E. (2007). The joint development of physical and indirect aggression: Predictors of continuity and change during childhood. *Development and Psychopathology*, 19, 37–55.
- Crick, N. R., & Grotpeter, J. K. (1996). Children's treatment by peers: Victims of relational and overt aggression. *Development and Psychopathology*, 8, 367–380.
- Crick, N. R., Ostrov, J. M., & Werner, N. E. (2006). A longitudinal study of relational aggression, physical aggression, and children's social-psychological adjustment. *Journal of Abnormal Child Psychology*, 34, 131–142.
- Crothers, L. M., & Levinson, E. M. (2004). Assessment of bullying: A review of methods and instruments. *Journal of Counseling & Development*, 82, 496–503.

- Dennis, M. J. B., & Satcher, J. (1999). Name calling and peer beliefs of elementary school children. *Professional School Counseling, 3*, 76–80.
- Embry, S. L., & Luzzo, D. A. (1996). The relationship between name-calling and peer beliefs among elementary school children: Implications for school counselors. *Elementary School Guidance & Counseling, 31*, 122–130.
- 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*, 365–383.
- Finkelhor, D., Ormrod, R. K., & Turner, H. A. (2009). The developmental epidemiology of childhood victimization. *Journal of Interpersonal Violence, 24*, 711–731.
- Glew, G. M., Fan, M., Katon, W., Rivara, F. P., & Kernic, M. A. (2005). Bullying, psychosocial adjustment, and academic performance in elementary school. *Archives of Pediatrics and Adolescent Medicine, 159*, 1026–1031.
- Great Schools (2009). Retrieved May 4, 2009 from <http://www.greatschools.net>.
- Greif, J. L., & Furlong, M. J. (2006). The assessment of school bullying: Using theory to inform practice. *Journal of School Violence, 5*, 33–50.
- Hartung, C. M., & Scambler, D. J. (2007). Approaches to measuring, preventing, and decreasing bullying and victimization in schools. In S. Evans, M. Weist, & Z. Serpell (Eds.), *Advances in school-based mental health interventions* (Vol. 2, pp. 13:1–13:17). New York: Civic Research Institute.
- Henson, R. K. (2001). Understanding internal consistency reliability estimates. *Measurement and Evaluation in Counseling and Development, 34*, 177–189.
- Keppel, G., & Wickens, T. D. (2007). *Design and analysis: A researcher's handbook* (5th ed.). New Jersey: Prentice Hall.
- Kyriakides, L., Kaloyirou, C., & Lindsay, G. (2006). An analyses of the Revised Olweus Bully/Victim Questionnaire using the Rasch measurement model. *British Journal of Educational Psychology, 76*, 781–801.
- Limber, S. P., & Small, M. A. (2003). State laws and policies to address bullying in schools. *School Psychology Review, 32*, 445–455.
- Lindeman, M., Harakka, T., & Keltikangas-Jarvinen, L. (1997). Age and gender differences in adolescents' reactions to conflict situations: Aggression, prosociality, and withdrawal. *Journal of Youth and Adolescence, 26*, 339–351.
- Murray-Close, D., Ostrov, J. M., & Crick, N. R. (2007). A short-term longitudinal study of growth of relational aggression during middle childhood: Associations with gender, friendship intimacy, and internalizing problems. *Development & Psychopathology, 19*, 187–203.
- Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment. *Journal of American Medical Association, 285*, 2094–2100.
- Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Cambridge, MA: Blackwell Publishers.
- Olweus, D. (2004). Brief psychometric information. Unpublished Manuscript.
- Ortega, R., Mora-Merchan, J. A., Singer, M., Smith, P. K., Pereira, B., & Menesini, E. (2001). Final report of the working group on general survey questionnaires and nomination methods concerning bullying. *Nature and Prevention of Bullying*. Retrieved November 12, 2008, from http://old.gold.ac.uk/tmr/reports/aim2_seville1.html.
- Owens, L., Shute, R., & Slee, P. (2000). “Guess what I just heard!”: Indirect aggression among teenage girls in Australia. *Aggressive Behavior, 26*, 67–83.
- Pepler, D. J., & Craig, W. M. (1995). A peek behind the fence: Naturalistic observations of aggressive children with remote audiovisual recording. *Developmental Psychology, 31*, 548–553.
- Pepler, D. J., Craig, W. M., Connolly, J. A., Yuile, A., McMaster, L., & Jiang, D. (2006). A developmental perspective on bullying. *Aggressive Behavior, 32*, 376–384.
- Perry, D. G., Kusel, S. J., & Perry, L. C. (1988). Victims of peer aggression. *Developmental Psychology, 24*, 807–814.
- Peterson, L., Brown, D., Bartelstone, J., & Kern, T. (1996). Methodological considerations in participant event monitoring of low-base-rate events in health psychology: Children's injuries as a model. *Health Psychology, 15*, 124–130.
- Reynolds, W. M. (2003). *Reynolds bully victimization scales for schools manual*. San Antonio, TX: The Psychological Corporation.
- Ross, D. M. (1996). *Childhood bullying and teasing: What school personnel, other professionals, and parents can do*. Alexandria, VA: American Counseling Association.
- Sawyer, A. L., Bradshaw, C. P., & O'Brennan, L. M. (2008). Examining ethnic, gender, and developmental differences in the way children report being a victim of “bullying” on self-report measures. *Journal of Adolescent Health, 43*, 106–114.
- Seals, D., & Young, J. (2003). Bullying and victimization: Prevalence and relationship to gender, grade level, ethnicity, self-esteem, and depression. *Adolescence, 38*, 735–747.
- Silverman, W. K., & Rabian, B. (1999). Rating scales for anxiety and mood disorders. In D. Shaffer & C. P. Lucas (Eds.), *Diagnostic assessment in child and adolescent psychopathology* (pp. 127–166). New York: Guilford Press.
- Smith, P. K., & Ananiadou, K. (2003). The nature of school bullying and the effectiveness of school-based interventions. *Journal of Applied Psychoanalytic Studies, 5*, 189–209.
- Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggressive Behavior, 29*, 239–268.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245–251.
- Suveg, C., Payne, M., Thomassin, K., & Jacob, M. L. (2010). Electronic diaries: A feasible method of assessing emotional experiences in youth? *Journal of Psychopathology and Behavioral Assessment, 32*, 57–67.
- Underwood, M. K., Scott, B. L., Galperin, M. B., Bjornstad, G. J. & Sexton, A. M. (2004). An observational study of social exclusion under varied conditions: Gender and developmental differences.
- van der Wal, M. F., de Wit, C. A. M., & Hirsing, R. A. (2003). Psychosocial health among young victims and offenders of direct and indirect bullying. *Pediatrics, 111*, 1312–1317.