Getting Precise and Pragmatic About the Assessment of Bullying: The Development of the California Bullying Victimization Scale

Erika D. Felix¹*, Jill D. Sharkey¹, Jennifer Greif Green², Michael J. Furlong¹, and Diane Tanigawa¹

¹University of California, Santa Barbara, California
²Boston University, Boston, Massachusetts

Accurate assessment of bullying is essential to intervention planning and evaluation. Limitations to many currently available self-report measures of bullying victimization include a lack of psychometric information, use of the emotionally laden term “bullying” in definition-first approaches to self-report surveys, and not assessing all components of the definition of bullying (chronicity, intentionality, and imbalance of power) in behavioral-based self-report methods. To address these limitations, we developed the California Bullying Victimization Scale (CBVS), which is a self-report scale that measures the three-part definition of bullying without the use of the term bully. We examined test–retest reliability and the concurrent and predictive validity of the CBVS across students in Grades 5–12 in four central California schools. Concurrent validity was assessed by comparing the CBVS with a common, definition-based bullying victimization measure. Predictive validity was examined through the co-administration of measures of psychological well-being. Analysis by grade and gender are included. Results support the test–retest reliability of the CBVS over a 2-week period. The CBVS was significantly, positively correlated with another bullying assessment and was related in expected directions to measures of well-being. Implications for differentiating peer victimization and bullying victimization via self-report measures are discussed. Aggr. Behav. 37:234–247, 2011.

Keywords: bullying; victimization; power imbalance; measurement; self-report; psychometrics

INTRODUCTION

Accurate assessment of bullying is essential to intervention planning and evaluation [Furlong et al., 2010]; however, there are long-standing concerns about its measurement [Cornell et al., 2006]. Bullying is recognized as a subset of peer victimization that is intentional, chronic, and characterized by an imbalance of power between victim and aggressor [Olweus, 1993]. Researchers worldwide have long struggled to measure bullying in ways that facilitate cross-national comparisons and to accurately estimate prevalence rates. Past efforts have produced equivocal results with considerable differences of prevalence rates across studies [Smith et al., 2002], which raises the issue of whether rates of bullying differ dramatically across samples or if differences reflect measurement imprecision. Measurement concerns include: (a) whether or not to provide an a priori definition of bullying to respondents [Espelage and Swearer, 2003]; (b) variations in definitions and time frames used; (c) whether to use self-report, peer nomination, or teacher-report methods [Cornell et al., 2006; Solberg and Olweus, 2003]; and (d) if available measures actually assess the subset of peer victimization that is intended to be captured by the scientific definition of bullying [Greif and Furlong, 2006]. The need for a sophisticated bully victimization assessment tool has prompted scholars to rely on prominent existing measures for more advanced purposes than originally intended. Olweus [1996] developed one of the first and most widely used and adapted measures. His measure, and similar bullying assessment instruments, was developed to estimate prevalence—examine base rates at the school level.


*Correspondence to: Erika D. Felix, Department of Counseling, Clinical, and School Psychology, Gevirtz Graduate School of Education, University of California, Santa Barbara, CA 93106-9490. E-mail: efelix@education.ucsb.edu

Received 20 April 2010; Revised 16 September 2010; Accepted 13 January 2011

Published online 14 March 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/ab.20389
Scholars have adapted these measures from their original purpose in order to examine individual differences through methods of classifying children by bullying experience and correlating antecedents and consequences to this classification. Bully assessment tools are also needed to evaluate intervention programs at the individual- vs. school-level [Greif and Furlong, 2006]. It is only recently that scales developed for epidemiological purposes have been evaluated for the psychometrics needed for individual differences research [e.g., Kyriakides et al., 2006; Lee and Cornell, 2010], but with limited, unique samples. Thus, assessing the psychometrics of self-report bullying victimization measures remains crucial for research and applied practice to continue to advance. In this article, we describe the development and psychometric properties of the California Bullying Victimization Scale (CBVS). We begin by reviewing the state of self-report bullying assessment, different approaches within self-report, and the need for further refinement of self-report assessment.

Self-Report Bullying Measurement Approaches

Self-report assessments are the most common method to measure bullying victimization, in part because they are easier and less costly for researchers and educators to implement than other methods, such as intensive behavioral observations of schoolyard interactions. Researchers have debated the relative benefits of self-report vs. other methods. Solberg and Olweus [2003] argued that self-report is the best method for ascertaining prevalence estimates, which is information that educators and policy makers often need. On the other hand, Cornell and colleagues [Cornell et al., 2006; Lee and Cornell, 2010] pointed out that many available measures do not have adequate reliability and validity. For example, Henry [2006] found that self-reports of aggressive behavior lacked criterion-related validity, whereas teacher- and peer-report methods were significantly correlated to observations of aggressive behavior. However, research has not demonstrated a relative lack of validity for self-reports of victimization, especially for indirect forms of victimization that are not readily observable. Thus, self-reports remain a viable way to assess the prevalence of bullying victimization and further research is clearly warranted.

Considering concerns about the reliability and validity of bullying self-reports, there have been calls to use other assessment procedures such as peer nominations [e.g., Chan, 2006; Chan et al., 2005] and behavioral observations [e.g., Cornell and Brockenbrough, 2004; Craig et al., 2000; Hawkins et al., 2001]. Studies using these methodologies have provided valuable data to answer questions that could not be addressed by self-report data alone. However, there could be ethical and/or logistical challenges with peer nominations and behavioral observations that make it more difficult to obtain research review board or school approval and written guardian consent [Espelage and Swearer, 2003]. For example, in playground observations, researchers face an ethical dilemma when they witness victimization as it would be unethical to observe a physical attack without intervening. Likewise with peer nominations, if parents do not consent for their children to participate, researchers may not collect peer nominations of their sociometric statuses. A single missing participant could affect the validity of all classroom or grade-level reports. Although researchers have successfully negotiated these potential dilemmas to produce valuable research, these considerations may be limiting factors in some schools, communities, and universities, making self-reports a favorable alternative.

In fact, self-report has been the primary approach used by researchers. Hence, it is critical to examine the status and quality of bullying self-report measures [see Furlong et al., 2010 for a review]. The self-report measures used by researchers generally either (a) provide a working definition of bullying and ask a youth if they had experienced this type of victimization or (b) present a list of victimization-related behaviors and ask how often the youth has experienced or committed them. We will review both approaches.

Definition-based self-report strategy. Assessments of bullying and bullying victimization often provide a definition of bullying to acquaint students to the idea of bullying as defined by the researchers, with the intent of establishing a shared meaning of the bullying experience across all study participants. Considerable effort has been devoted to developing a consensus on how to define and measure bullying [e.g., Arora, 1996; Naylor et al., 2006; Smith et al., 2002]. Smith et al. [2002] point out that the issue of definition is crucial for accurate reporting of bullying prevalence cross-nationally and that the word “bully” is not easy to translate. To overcome terminology challenges, they used 25 stick-figure drawings depicting several types of aggressive experiences to assess the words used to describe these experiences across a 14-nation sample. They found significant variation in the number of drawings assigned to each term, demonstrating different victimization scenarios implied by each term.
Researchers have debated the merits of defining bullying before asking students’ about their experiences. The definition of bullying has evolved over the past 30 years to include a broad range of victimization; hence, studies have varied in measurement practices. These practices have led to varying prevalence estimates and consequently it is difficult to evaluate the nature and extent of cross-sample comparability. Also, studies have repeatedly shown that different informants have different views of the definition. Older youth have a more nuanced and complex view of bullying than younger children [Smith et al., 2002]. Similarly, teachers’ definitions of bullying tend to correspond more to the three-part definition of intentionality, repetition, and imbalance of power than students [Menesini et al., 2002].

In Europe, a workgroup reviewed self-report and peer nomination bullying assessment methods in an effort to improve bullying intervention [Ortega et al., 2001]. In a comprehensive review, they assessed the content, items, age ranges, and advantages and disadvantages for both self-report and peer nomination methods to propose core questions for a general survey questionnaire. They concluded that the ideal time frame to explore is the last 3 months, and they recommended using a definition, with cartoons to illustrate examples of physical, verbal, and indirect bullying. However, it is not clear the extent to which these practices have been adopted.

Even if definitional consensus were possible, asking a youth to label himself or herself as a victim or a bully may provoke emotional reactions that could influence victims or bullies to not endorse experiences associated with the label [Cornell and Brockenbrough, 2004; Greif and Furlong, 2006; Hamby and Finkelhor, 2000]. One study found that when students are provided with a definition of bullying, and are repeatedly exposed to the term “bully,” they report significantly less bullying than those not exposed to a definition or the word “bully” [Kert et al., 2010]. In this circumstance, bullying prevalence would be underestimated and intervention outcomes would be affected by imprecise measurement. Likewise, in studies of adult sexual harassment, participants commonly endorse the critical behavioral experiences associated with the legal definition of sexual harassment, but at the end of the questionnaire, when asked if they had been “sexually harassed” many denied it, perhaps not identifying with the label [Cortina et al., 1998; Fitzgerald and Ormerod, 1991].

Others argue that using a definition-first strategy that includes multiple forms of aggression [e.g., presenting a definition listing various types of aggression and then asking the youth if he or she has been bullied] may produce heterogeneous data that mask trends and correlations among subtypes of bullying experiences [Cornell et al., 2006]. A broad definition does not allow for assessing the relative contribution of each form of bullying victimization to psychosocial adjustment [Cornell et al., 2006; Felix and McMahon, 2006]. Even if more than one question about victimization is asked following a bullying definition, it is unclear how well students remember long, multipart definitions when responding to a series of items. Smith et al. [2002] point out that even if a student is initially given an extended definition of bullying, the term “bullying” is used in most of the following questions, and children may use their natural understanding of this term in their culture, rather than the originally presented definition. Menesini et al. [2002] noted that students in their sample were able to take into consideration repetition, intention, and imbalance of power, but their measure used visual aides in the form of cartoons, which most definition-first approaches do not do. Given these concerns, some researchers have adopted a behavior-based strategy.

**Behavior-based self-report strategy.** In a behavior-based approach, specific bully-related behaviors are presented (e.g., hitting, threatening, or spreading rumors) and youth are asked to specify if they have committed or experienced them without using the term bullying. The behavioral method breaks down victimization into specific experiences in order to avoid individual perceptions, stigma, or bias associated with using the term “bullying” or “victim” in the measure itself, and to allow researchers to examine the frequency of each type of victimization. However, this approach does not tend to assess for an imbalance of power, and thus fighting between peers of equal strength, which is considered to be aggression but not bullying, would likely be included [Smith et al., 2002].

Behavioral strategies generally classify the group with the highest level of victimization experiences as “victims” and those with the highest levels of perpetration as “bullies” [Espelage and Swearer, 2003]. Some issues associated with this approach are: (a) the criteria used to select extreme responders (e.g., one standard deviation above the mean, less, more) vary across studies [Solberg and Olweus, 2003]; (b) educators conducting school-based surveys are unlikely to be able to complete the complex computations needed to classify students with statistical methods; and (c) unless additional items are coadministered, the behavioral method assesses only one aspect of the definition of bullying, which is
the repetitive nature of the experience. This neglects intentionality and imbalance of power, which also differentiate bullying from general peer victimization. Although it may be argued that if a child is repeatedly victimized, there must be an imbalance of power, as the child is unable to stop the victimization, this is not always the case. For example, two boys of the same strength and social standing may dislike each other and repeatedly fight, but this would not traditionally be considered bullying.

**Advantages of Self-Report Strategies and Moving Forward**

Self-reports address many complex issues of understanding bullying behavior. First, self-report procedures are efficient and can be conducted in school, community, and laboratory settings with minimal cost. Second, self-reports can assess the diverse and important subtypes of bullying behavior, including direct, indirect, and relational aggression. This includes behaviors that are not readily observable by others. Third, they have the potential to assess the perceived power imbalance from the perspective of the bully, victim, or bully-victim. Fourth, they do not rely on consent from other participants (as in peer nomination procedures). Fifth, they assess the perspective of the participant (both bully and victim), which is critical when attempting to understand intentionality and impacts on psychosocial functioning. Finally, self-report can simultaneously assess and distinguish between different forms of bullying behavior to better understand differential impact.

Although there are advantages to using self-report, there are also unresolved issues to be addressed in order to move research and evaluation of prevention programs forward. Bullying is a subset of peer victimization [Björkqvist et al., 1982] and a challenge of bullying assessment methods is to distinguish bullying from other types of peer victimization including playful behavior [Cornell et al., 2006]. Indeed, some researchers have gone beyond the bully, victim, bully-victim, and nonvictim classification to explore the roles of all students in bullying encounters [e.g., Salmivalli, 1998, 1999, 2010]. Salmivalli [1998] articulated multiple participant roles, in addition to bully and victim, which included the assistant to the bully, the reinforcer, the defender, and the outsider. As many have noted, friends or acquaintances can engage in teasing and horseplay that may look like bullying to an outside observer [Cornell et al., 2006; Rigby, 2004]. Given various bully victimization roles, it is critical to evaluate the importance of nuances in types of peer victimization to accurately understand the impact of these diverse experiences.

Research supports the important distinction between the larger group of peer victims and the subset of bully victims [You et al., 2008]. Thus, there is a need to explore alternative strategies to assess power imbalance, one of the three key elements of bullying. Namely, youth who have experienced repeated bullying have worse psychological health than youth who have experienced peer victimization that was not bullying. Thus, relying solely on the repetitive nature of victimization may be inadequate to truly understanding the bully experience and research needs to implement and test methods of assessing the power imbalance.

Lastly, researchers vary on defining chronicity. Studies using a behavioral approach have identified extreme responders to classify “bullies” or “victims” [Espelage and Swearer, 2003], which necessitates a cumbersome statistical procedure. However, Solberg and Olweus [2003] argued that using a definition of bullying and the empirically derived frequency criteria of “2–3 times per month” or more derives the best estimate of bullying victimization based on its relation with negative conditions such as depression. As there are advantages and disadvantages to both definition-first and behavioral approaches, and there is a need for identifying the psychometric properties of a self-report bullying assessment specifically designed for individual differences research, we developed the CBVS aiming to address current needs in bullying assessment.

**California Bullying Victimization Survey**

The CBVS was developed as a self-report measure of multiple forms of bullying victimization, and to distinguish bullying from peer victimization, without the use of a definition. The CBVS avoids the label “bully,” on the premise that this term may be emotionally laden to respondents, and influence victims or bullies not to endorse experiences associated with the label [Cornell and Brockenhroug, 2004; Greif and Furlong, 2006]. As one method to differentiate bully victims from peer victims, we specify in each question that the behavior must be done on purpose in a mean and hurtful way. Anecdotally, we observed in student interviews that some children who initially endorsed a behavioral description later stated that it was not done on purpose in a mean or hurtful way. Thus, without this specifier, prevalence rates may be overestimated. The aggressor and victim may have different opinions as the aggressor may deny that he or she intended to hurt the other person or may minimize
the harm done. We took the perspective of the victims, as ultimately it is their appraisal of the situation that will likely affect their well-being.\(^1\)

Another way to differentiate bullying from peer victimization is through the presence of a power imbalance. Without using a definition, the CBVS was designed to assess whether the person who perceives a power imbalance between self and the aggressor. We assess power imbalance in terms of how popular, smart, and strong was the person who did mean things on purpose to the respondent. These power imbalance options are similar to those in the Swearer Bullying Survey [Swearer, 2001].

An additional measurement consideration addressed by the CBVS is the time frame and frequency scale. Given that the accuracy of recall diminishes with time, the CBVS uses a past 30-day time frame and the same frequency scale used in the Olweus Questionnaire, which allows comparison across studies using his measure. The Olweus frequency scale classifies victims as those experiencing bullying 2–3 times a month or more, a cutoff that is easy for schools to use and understand.

**Study Purpose**

The purpose of this study is to examine the psychometric properties of the CBVS for use with individual students to determine its functioning as a new measure of bullying victimization. First we examine item-level test–retest reliability of the CBVS victimization items. We subsequently use the CBVS to classify students as nonvictims, peer victims, or bully victims and examine the test–retest classification reliability of these categories. We next turn to measure validity and compare the CBVS classification of bully victims to respondents who are classified as bully victims using the “definition-first approach.”

Finally, we examine the predictive validity of the CBVS to determine the extent to which classification on this measure is associated with theoretically related constructs and measures of psychological well-being. As there are gender differences in the type of victimization girls and boys experience most frequently [e.g., Crick and Grotpeter, 1995; Olweus, 1993], initial analysis was by gender. In addition, given the developmental range covered in this study, analyses were also separated by grade groupings (i.e., elementary, middle school, high school).

\(^1\)Research on sexual harassment notes a significant difference in opinion between the perpetrator and victim on whether a behavior constituted harassment and legally, courts uphold the view that it is not the intent of the perpetrator that matters, but rather the effect on the victim [Paludi, 1997].

**METHOD**

**Preliminary Development of CBVS**

In response to the need for improved self-report assessments, as well as a request for help with a school-wide screening by a local junior high school, we created a bullying victimization survey that served as the basis for the CBVS. This initial draft of the survey was administered to 463 junior high school students of whom 329 were in seventh grade (71%) and 134 were in eighth grade (29%) in June 2005. The sample was 55% White, 26% Latino/a, and 19% representing other ethnic groups, and was evenly divided on gender (51.6% female). To further develop the survey, we analyzed student responses to items for any inconsistent responding and queried students for qualitative feedback on item content. We delete revised items to increase clarity and added an item on sexual harassment for students in Grade 7 or higher, in recognition that it is a common form of peer victimization starting around puberty [Felix and McMahon, 2006, 2007]. Subsequently, our research team thoroughly reviewed several drafts of the CBVS for consensus on item content, wording, and layout.

To further refine the CBVS, focus groups were subsequently conducted at junior high and elementary school levels to review a draft of the survey and provide feedback on item wording and clarity. At the junior high school level, we recruited two classrooms at a midsize school with diverse socioeconomic status (37% disadvantaged) and primarily European American (42%) and Latino/a American (40%) ethnic background: an eighth-grade leadership class and a seventh-grade general science class. At the elementary level, we selected the single fifth/sixth-grade classroom at a small school that serves primarily Latino/a American (84%) and socioeconomically disadvantaged youth (74%). Consent forms were sent home to parents at both schools with over 90% positive consent across classrooms. All students present on the day of the focus group participated for one class period, which was 50 min. More females than males participated, and the groups were roughly evenly divided between grades, with the exception of a lower number of sixth graders. One of the principal investigators and a graduate student research assistant conducted each focus group. Students completed an item-by-item review of the CBVS and were asked about the clarity of instructions and item wording. The research assistant documented feedback on each item and notes on each focus group were reviewed by the research team and common themes were identified. For example, students questioned the
term “washroom” and suggested “bathroom” instead. The survey wording was changed where there was consistent feedback or a consensus between researchers, and the research team approved a final version.

In October 2006, two focus groups were conducted with a convenience sample of high school students in their elective class to review the wording of the survey to make it appropriate for an older student population. All students present for that class on the day of the focus group participated. The student population for the school is 53% Latino/a, 42% Caucasian, and the remaining is a mix of other ethnicities. Approximately 27% of students were considered socioeconomically disadvantaged and 38% were English Learners. One of the principal investigators and a graduate student research assistant conducted each focus group. Students were led on an item-by-item review of the CBVS and asked if the instructions and item wording were understandable. No wording changes were suggested for the items assessing victimization.

Participants

Criteria for retention in the study. We reviewed the data to find cases that had such high levels of missing data that it was not acceptable to retain them in the study. Using the following criteria, we maintained 92% of all participants in the final dataset: (a) no more than five items missing from the validity scales and (b) at least one victimization item completed. Although the complete analysis required that the youth complete all victimization items, we retained cases with incomplete victimization responses for the item-by-item test–retest analysis. For this reason, the number of cases varies in the analyses shown in the following sections.

Sample 1. Three schools in the central coast region of California participated in the survey. One school was a junior high school, one was a K-8 school, and one was a K-6 elementary school. A total of 330 students had valid responses: 46% boys, 14% fifth graders, 12% sixth graders, 40% seventh graders, and 34% eighth graders. The ethnic composition of the students was 51% White, 29% Hispanic/Latino(a), and 12% represent other or mixed ethnicity. Institutional Review Board concerns for minority ethnicity. Institutional Review Board concerns for Hispanic/Latino(a), and 20% other or mixed ethnicity.

Sample 2. In study 2, we extended our descriptive study and exploration of the validity of the CBVS to a high school sample. One high school in the central coast region participated, and a total of 354 students provided valid surveys, representing 14% of the total student population. Over half were female (53%). For grade representation, there were 39% in ninth grade, 22% in tenth grade, 22% in eleventh grade, and 17% in twelfth grade. The ethnic composition was 54% White, 34% Hispanic/Latino(a), and 12% representing other or mixed ethnicities.

Measures

A single questionnaire was created and titled, My Experiences with Classmates at School. After brief instructions, the questionnaire consisted of the following scales in the order described.

California Bully Victimization Scale. The final version of the CBVS includes items asking about six (elementary version) to seven (secondary version) forms of victimization that respondents may have experienced at school: been teased or called names by another student; rumors spread or gossip behind someone’s back; left out of a group or ignored; hit, pushed, or physically hurt; been threatened; had your things stolen or damaged; or had sexual comments or gestures directed at them (this question is only included in the secondary school versions). Students are asked to rate the frequency of each of these experiences on a five-point scale (0 = Never, 1 = once in the past month, 2 = 2 or 3 times in the past

For Study 2, we also piloted a question on cyber-victimization that was not originally included in Study 1 with the middle school students. We found 11% of the high school students we sampled reported some cyber-victimization. We decided not to include that one item in this study because (a) it was only in Study 2, whereas sexual harassment was assessed in both studies (Grade 7 or higher), and (b) it seems to be more of a means of victimization than a type, in our estimation. For example, texting can be considered akin to passing mean notes; threatening a fight via a social network sites is similar to threatening via “friends” or other associates. Other research has supported that cyber-bullying is more of a modality than a distinct type [e.g., Varjas et al., 2009].
month, 3 = about once a week, and 4 = several times a week). In the next question, respondents are asked to indicate if these behaviors were intentional and done in a mean way. They select from among the following responses in reference to the previously rated victimization experiences broadly: They were almost never mean (just joking), they were sometimes mean, they were almost always mean. Next, the CBVS assesses power imbalance through a series of items asking students to compare themselves to the “main person who did these things to you.” Students rated victimization experiences broadly: They select from among the possibilities in a mean way. They select from among the possibilities in comparison to themselves. These choices are similar to the ones Swearer [2001] used in her survey. The CBVS includes several other questions not used in this study that are designed to guide intervention planning including types of bullying witnessed (as a bystander), information about students doing the bullying, where on the school campus bullying occurs, when during the school day it is most frequent, and who students talk with about bullying.

Swearer Bullying Survey. The CBVS was compared with an item from the Swearer [2001] Bullying Survey that assessed bullying by providing a definition and asking students how frequently they have been victimized. The definition states: “Bullying happens when someone hurts or scares another person on purpose and the person being bullied has a hard time defending himself or herself. Usually, bullying happens over and over.” The following are listed as possible examples of bullying behavior: punching, shoving and other acts that hurt people physically; spreading bad rumors about people; keeping certain people out of a “group”; teasing people in a mean way; and getting certain people to “gang up” on others. Students are asked to indicate if they were bullied this month and if yes, how often (i.e., once in the past month, 2–3 times in the past month, once a week, several times a week).

Students’ Life Satisfaction Scale (SLSS). This seven-item measure uses a six-point Likert scale (strongly disagree to strongly agree) to assess overall well-being (e.g., “My life is going well” and “My life is just right”) for students ages 8–18 years old. According to prior research, internal consistency ranges between .73 and .86 [Huebner et al., 2005] and test–retest reliability is .76 across one to two weeks [Terry and Huebner, 1995]. Correlations of the SLSS with other life satisfaction scales meet expectations [Huebner, 1991], and studies support its construct, discriminant, and predictive validity [Huebner et al., 2005]. For this study, the \( \alpha \) coefficient was .88.

School Connectedness Scale (SCS). The SCS measures the bond felt by students toward their school and the quality of the relationship between students and teachers [McNeely, 2005]. Responses are measured using a five-point Likert scale (strongly disagree to strongly agree). This scale was constructed out of items originally included in the National Longitudinal Study of Adolescent Health. McNeely [2005] noted that three versions of the SCS have been used from the NLSAH Study. The version used in this study is the one employed by McMeely et al. [2002], and administered in the Resilience Youth Development Module of the California Healthy Kids Survey (see WestEd: www.wested.org/hks). An \( \alpha \) of .81 was obtained in this study.

Children’s Hope Scale (CHS). The six-item CHS is designed for children between the ages of 7 and 15 years. The CHS measures two aspects of hope—the cognitive capacity to formulate plans to achieve set goals (three items; e.g., “I can think of many ways to get the things in life that are most important to me”), and self-efficacy, the belief that one can achieve set goals through effort (three items; e.g., “I think I am doing well”). Children respond to the items using a six-point scale (1 = none of the time to 6 = all of the time). Both the internal consistency and the test–retest reliability are greater than .70 [Snyder, 2005]. The CHS possesses strong concurrent validity and acceptable predictive and discriminant validity [Snyder, 2005]. The \( \alpha \) for this study was .88.

Procedures

Sample 1. Schools participated in the survey in May and June of 2006 after obtaining Institutional Review Board approval. At the junior high school, two social studies teachers agreed to administer the survey to their students in multiple classes. At the K-8 and elementary schools, all fifth- and sixth-grade students were given the opportunity to participate. Teachers were provided with detailed written instructions for administering the survey by the research team. Teachers administered the survey in their classrooms to students with parental consent and student assent. Students were not asked to provide their names; rather, random numbers were used to protect their identity. For test–retest reliability analysis, students received an envelope with two surveys matched by their random identification number. Students wrote their name on the envelope and approximately 2 weeks later, teachers readministered the survey (retest subsample). Students turned in the envelopes with the matched surveys and the envelopes were destroyed.
Sample 2. In November 2006, teachers administered the survey in a paper-and-pencil format during one class period to all students in the high school who provided parental consent and student assent. A detailed instruction sheet was provided for teachers to administer the survey.

RESULTS

Rates of Peer Victimization

Descriptive statistics were calculated to determine the percentage of boys and girls at each school level (elementary, junior high, high school) who experienced each type of victimization with the frequency criteria set at 2–3 times per month or more. Chi-square tests of significance revealed if rates were different by gender. Results indicate that peer victimization is a common experience at school (Table I). Teasing was the most frequent form of victimization reported by male and female students across grade levels. Females reported higher rates of victimization through rumors and social exclusion, as compared with males. Males in Grades 7–12 reported that they were hit significantly more often than females, although the opposite trend was found in Grades 5 and 6. In general, there was a decline in frequency of reported victimization from elementary to high school.

Students were also classified as nonvictims, peer victims, and bullied victims. Nonvictims were students who reported no victimization experiences on the CBVS. Peer victims reported at least one victimization experience of any frequency, but no power imbalance. Bully victims were students who reported at least one type of victimization at least 2–3 times per month, reported that this victimization was intentional at least some of the time, and reported at least one form of power imbalance in relation to the main person bullying them (Table II). A hierarchical log-linear analysis with backward stepwise model selection was conducted with SPSS 18.0 to test the higher-order relations between the three classification variables of gender, age, and victimization status. In the backwards elimination steps, the first interaction eliminated was for gender by grade, \( \chi^2 = 0.71, df = 2, P = .70 \), followed by the three-way interaction, \( \chi^2 = 4.64, df = 6, P = .59 \), and finally the gender by victimization interaction, \( \chi^2 = 4.35, df = 2, P = .11 \). There was a remaining significant interaction for grade by victimization, \( \chi^2 = 163.41, df = 4, P < .001 \). The percentages of students categorized as peer and bullied victims declined with age; this decrease was most pronounced for the general class of peer victims, which decreased from 45.3% and 52.2% among elementary and junior high school students, respectively, which compared with only 11.3% reported by high school students. Almost three-fourths of high schools students (73.4%) were in the nonvictim group, compared with 28.2% of junior high and 31.0% of elementary school students. The decline in bullied victims was smaller across the grade levels: elementary school (23.8%), junior high school (19.6%), and high school (15.3%).

Test–Retest Stability

Two-week stability of the CBVS was evaluated to determine test–retest reliability in a subsample of students \( n = 131 \). First, as one test of response stability, a total victimization score was calculated as a sum of the six core victimization items (all but sexual harassment). Correlations between the total victimization scores at time 1 and time 2 indicated a high degree of agreement across points in time.
Cohen’s kappa coefficient was examined for each item (Table III). For individual types of victimization, Cohen’s $k$ ranged from .46 to .64, and were all significant, indicating that agreement exceeds chance levels across victimization experiences. As a final analysis of test–retest stability, we examined the consistency in classifications of students as nonbullied and bullied across two points in time (percent agreement = 89.6%, $\kappa$ = .71; $\chi^2$ = 65.6, $P < .001$).

**Concurrent Validity**

Evidence for the concurrent validity of the CBVS was evaluated in comparison to a “definition-first” approach assessing bullying using the Swearer Survey [Swearer, 2001; Table V]. The rates of students classified as being bullied were relatively similar in the definition-first measure (20.2% of fifth/sixth graders, 17.5% of seventh/eighth graders) and the CBVS (23.8% of fifth/sixth graders, 20.0% of seventh/eighth graders). However, there were substantial differences in exactly which students were identified as being bullied by the two measures ($\kappa$ = .34 for fifth/sixth graders, $\kappa$ = .49 for seventh/eighth graders). This indicates that the different measurement strategies significantly influenced classification status of individual students (i.e., they identified somewhat different students).

**Predictive Validity**

The predictive validity of the CBVS was analyzed per grade-level grouping by examining the relation between victimization experiences and measures of life satisfaction, school connectedness, and hope (Table VI). The total victimization score had a moderately negative relationship ($r = -.14$ to $-.50$) with each of these indicators across age groups, which were statistically significant in eight out of nine comparisons.
Analysis of variance of victimization classification with measures of life satisfaction, school connectedness, and hope indicated a significant decrease in positive functioning among students experiencing peer victimization and bullying (Table VII). Students who were specifically classified as bullied victims by the CBVS had significantly lower scores on these indicators of positive functioning than students who were peer victims but did not meet criteria to be classified as bullied. The association of victimization and positive functioning were most pronounced for junior high and high school students. Life satisfaction and school connectedness seemed to suffer, in particular, among students who were classified as peer victims and bullied victims.

**Power Difference and Victimization Frequency**

When applying the theoretical definition of bullying to assign groups, there is a possible confound between frequency of victimization and reporting a power differential. Although our objective was to determine whether strictly defined bullied victims differed from other types of peer victims, we conducted a contingency table analysis to determine the level of association between power differential and frequency of victimization. Results are displayed in Table VIII. Although there was a significant relation between frequency of victimization experience and endorsement of a power differential, $f^2 = 7.53$, df = 326, $P < .05$, practically, there were several students who reported frequent victimization yet did not endorse a power differential, suggesting that assessing a power differential can more accurately identify bullied victims.

**DISCUSSION**

Despite the many measures that have been developed to evaluate bullying, assessment continues to need improvement in this field of research [Cornell et al., 2006; Lee and Cornell, 2010]. This study evaluated the psychometric properties and validity of a new instrument, the CBVS, which was designed to address some of the weaknesses in existing measures of bullying. In particular, the CBVS differentiates bullying from other forms of peer victimization, a distinction that is rarely studied [Hunter et al., 2007; Schäfer et al., 2002], yet important for identifying students with the most negative peer experiences. The CBVS specifically measures the hallmark characteristics of bullying (intentional, repeated, power differential) to make this differentiation, while aiming to increase the accuracy of classification of peer victims and bullied victims. Several findings from this study indicate that the CBVS demonstrates strong reliability and validity.

**TABLE VI. Validity Coefficients of Total Victimization Score (range 0–24) and Well-being Measures**

<table>
<thead>
<tr>
<th>Well-being measures</th>
<th>Grade</th>
<th>5–6</th>
<th>7–8</th>
<th>9–12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life satisfaction</td>
<td>-0.50***</td>
<td>-0.37***</td>
<td>-0.25***</td>
<td></td>
</tr>
<tr>
<td>School connectedness</td>
<td>-0.17</td>
<td>-0.41**</td>
<td>-0.34***</td>
<td></td>
</tr>
<tr>
<td>Hope</td>
<td>-0.29**</td>
<td>-0.29**</td>
<td>-0.14***</td>
<td></td>
</tr>
</tbody>
</table>

* ***P < .001; **P < .01; one-tailed Pearson’s correlation.

**TABLE VII. Validity of Bully Victimization Classification for Victimization Group by Grade Level**

<table>
<thead>
<tr>
<th>Variability of Victimization</th>
<th>Grades 5–6</th>
<th>Grades 7–8</th>
<th>Grades 9–12</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSS</td>
<td>3.76 (0.9)</td>
<td>3.69 (0.7)</td>
<td>3.64 (1.0)</td>
</tr>
<tr>
<td>SCS</td>
<td>2.88 (0.9)</td>
<td>3.48 (0.9)</td>
<td>2.96 (0.6)</td>
</tr>
<tr>
<td>CHS</td>
<td>3.69 (0.7)</td>
<td>3.55 (1.0)</td>
<td>3.53 (1.1)</td>
</tr>
</tbody>
</table>

* df = 2, 77

**TABLE VIII. Victimization Total Score (range 0–24) and Percent with a Power Differential Endorsement**

<table>
<thead>
<tr>
<th>Power differential</th>
<th>Total</th>
<th>29.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>328</td>
<td></td>
</tr>
</tbody>
</table>

Note: LSS = Life Satisfaction Scale, range 0–5; SCS = School Connectedness Scale, range 0–4; CHS = Children’s Hope Scale, range 0–5. Standard deviations in parentheses.

***P < .001; **P < .01.

Aggr. Behav.
validity to evaluate bullying victimization at the individual level.

First, CBVS prevalence estimates had a similar pattern to other studies that report age-related decreases in rates of bully victimization, as youth transition from junior high to high school [Nansel et al., 2001; Solberg et al., 2007]. The CBVS additionally showed gender-related differences in the type of bullying experienced that are similar to previous research, with males reporting higher rates of physical victimization and females reporting higher rates of relational victimization [Crick and Grotpeter, 1995; Rivers and Smith, 1994].

Second, the CBVS had good test–retest reliability. Report stability is a key indicator of effective bullying evaluation as bullying, by definition, occurs repeatedly. The CBVS demonstrates good reliability over a 2-week period for the total scale score, each form of victimization, and classification of students as bully victims. It is notable that students who reported bully victimization consistently across the two assessments had lower scores on indicators of well-being than non victims or unstable bullied victims. This suggests that consistency across test–retest conditions is not only a methodological issue, but may also indicate the development of stable identification as a bullied victim, which is associated with poor psychosocial outcomes [Rosen et al., 2007].

Third, responses to the CBVS are associated with indicators of well-being. Students who reported higher CBVS scores reported lower life satisfaction, school connectedness, and hope, indicating that the CBVS detects students who are most at-risk for poor psychosocial development at school. Particularly striking is our finding that the group of students who were bullied had lower scores on positive indicators than students who reported more general experiences of peer victimization. This result is consistent with Hunter et al. [2007] who found when they categorized bullied youth separately from other peer victims, those who were bullied were at much greater risk for depressive symptomatology. Although few studies have isolated these groups and evaluated differential outcomes, students who specifically meet criteria for the definition of bullying experience significantly more difficulty than those who report more general peer victimization. In addition, we found that assessing for the power differential helped in the identification of bullied victims than when examining frequency of victimization alone.

Although CBVS prevalence rates were similar to those derived from a definition-first measure that was coadministered in this study, the two measures identified somewhat different sets of students as having been bullied. The lack of concordance between the two measures indicates that the way self-report surveys are structured fundamentally impacts which students will identify themselves as having been bullied. The absence of the word “bully” from the CBVS may have made questions easier to endorse for students who had been victimized. However, the requirement that students explicitly endorse power differential may have had the contrary effect of limiting respondents who did not perceive a power difference or did not identify the particular power difference that they experienced among the list provided by the CBVS. Our choice of what qualities to assess as a power differential represented only a few of potentially many sources, and was guided by previous bullying surveys. “Smart” was included as one source, as some prior research indicated that victims have lower self-concept in terms of intelligence [Salmivalli, 1998]. However, in practicality, only a few students identified smart as being the only power differential. It may be that other sources of power differential are more beneficial to assess.

In addition, prevalence rates in this study were higher than rates reported in some of the prior research using definition-first measures [Nansel et al., 2001, 2004; Solberg and Olweus, 2003] and peer-nomination methods [Juvonen et al., 2003]. Higher response rates on the CBVS and the Swearer Survey raise questions about whether the sample in this study truly experienced higher rates of bullying or whether inflated rates were a product of the methodology used in this evaluation. Yet, validity analyses support the categories yielded by the CBVS as students whose responses resulted in the bullied status also reported the lowest levels of well-being.

In developing the CBVS, we were particularly interested to determine whether it would be possible to include all three definitional characteristics of bullying in a self-report measure. We did this by asking a series of questions, so participants did not have to remember all three definitional components at once. Prior research has found that very few children naturally incorporate power differential and repetition in self-generated definitions of “bullying” [Vaillancourt et al., 2008]. Although definition-first measures commonly embed a statement about power differential in their description of bullying and subsequently ask students to rate the frequency of victimization, to our knowledge, only one other study has explicitly incorporated items measuring power differential in bullying assessment [Hunter et al., 2007]. This study found considerably higher rates of depressive symptoms among students who met
criteria for all components of the bullying definition. Similarly, Solberg and Olweus [2003] found that frequent victimization was associated with greater internalizing and externalizing problems.

Preliminary analysis indicated a significant, but not fully overlapping relation between repetition of victimization and power differential endorsement. It is possible that these two groups of victims have different relations to outcome variables. In the future, it will be important to determine whether all three components of the definition of bullying are equally associated with psychological well-being, or whether repetition, intentionality, and power imbalance may be unequally weighted in their impact or differentially associated with outcomes.

LIMITATIONS AND FUTURE DIRECTIONS

This study is an initial demonstration that the CBVS appears to have promise as a psychometrically sound measure, but some limitations deserve mention. First, the CBVS is a self-report measure and there have not yet been any data collected from other informants or through observation to confirm the validity of the measure. Even though other studies of bullying have found limited agreement between informants [Rønning et al., 2009], it would be useful to assess agreement between self-, teacher-, and peer-informant detection of students who were bullied victims vs. peer victims. Self-report alone may be subject to misattribution of aggression, which can lead to overidentification of bullying, or shame and fear of repercussions, which can lead to underestimation of bullying [Card and Hodges, 2008].

Second, the CBVS does not yet distinguish students who are purely victims of bullying from those who are also bullies. This is in development. Identifying bullies is associated with a new set of challenges; however, evidence suggests that students who are both bullies and victims (bully-victims) should be considered a separate group from pure victims as their profiles differ and are associated with worse outcomes [Nansel et al., 2004; Solberg et al., 2007]. Future iterations of the CBVS will explore identification of bullies, using similar strategies to evaluate power differential in relationships. Likewise, bullying can be seen as a peer group process [Salmivalli, 1998, 1999, 2010], and other participant roles, such as reinforcer and defender, deserve assessment and study to better inform intervention efforts.

A third limitation of this study is a question of the generalizability of the sample. As noted earlier, rates of bullying among the students surveyed were higher than is typical in bullying research, even for the definition-first measure that was coadministered with the CBVS. Although data were collected from four different schools, all schools were located within the central coastal region of California, with a predominantly White and Latino/a student population. We have a strong interest to expand our administration of the CBVS to students and schools in other locations.

CONCLUSION

Although there is a significant amount of research dedicated to peer victimization and bullying, this discipline is affected by a lack of measurement precision. Measures that have been thoroughly developed to estimate prevalence within broad populations have been adapted for individual difference analyses without revalidation. For bullying scholarship to advance, it is critical to have consistently high standards for psychometric testing of instruments before implementation in research or in practice. We developed the CBVS to address gaps between the definition and theory of bullying and the way bullying is measured. In this preliminary study, the CBVS reliably assessed the multiple components of bullying and validly distinguish between nonvictims, peer victims, and bullied victims. Results also indicated that these distinctions are important for identifying students most impacted by victimization experiences. Overall, these findings suggest that the CBVS has promise to be included in a new generation of bullying assessments that are valid for use in research and in practice to study the individual-level impacts of bullying.

Ultimately, the aim is for the CBVS to become an early screening in a multigating assessment, by which students with high CBVS scores will be referred for a more in-depth interview to determine the extent to which they are involved in bullying and identify mechanisms of the bullying relationship that can inform interventions. For example, a student who reports being bullied by a single classmate may need an intervention directed at that dyadic relationship in the context of their classroom. However, a student who is repeatedly bullied by multiple individuals across different settings may benefit from the equivalent of a functional behavioral analysis to understand the process by which bullying is occurring and guide intervention for that particular student [Colvin et al., 1998]. In addition, a classroom or school-wide intervention aimed at
addressing bullying behavior and the supporting roles of assistant to the bully, or reinforcer [Salmivalli, 1999] would be indicated. A multigating assessment framework provides school-wide data to derive large-scale prevention and intervention efforts, while offering detailed information about students who are bullied that can have practical implications for the targeted delivery of interventions for individuals.

ACKNOWLEDGMENTS

The project was supported by the Field Initiated Studies Program from the Hamilton Fish Institute with funding from the United States Office of Juvenile Justice and Delinquency Prevention via George Washington University. Additional information about the California Bully Victimization Scale is available from the first author.

REFERENCES


Aggr. Behav.


