
Jill D. Sharkey, Erin Dowdy, Jennifer Twyford, and Michael J. Furlong
UNIVERSITY OF CALIFORNIA, SANTA BARBARA

Abstract
Precise and accurate measures are needed for research to advance knowledge about school violence and safety concerns. This chapter provides an overview of school violence and safety measurement issues. After introducing the history of measurement of school violence and safety, we provide a review of recent studies of school violence with a focus on their measurement approaches and practices. We then describe procedures commonly used to measure school safety- and violence-related variables, review prominent threats to the reliability and validity of these procedures, and suggest mechanisms for improvement.

Progress has been made in understanding school violence and safety measurement issues; however, despite calls to action (e.g., Furlong, Morrison, Cornell, & Skiba, 2004; Sharkey, Furlong, & Yetter, 2004), school violence and safety research continues to be hindered by the limited scrutiny given to the psychometric properties of its measurement tools. This chapter focuses on how researchers have measured school violence and safety variables and the implications of their practices for data validity. After a historical summary, recent school violence research is examined with a focus on these studies’ measurement practices. Existing methods to measure school violence, and crucial measurement considerations, are described. We conclude with a list of important measurement considerations for school violence research.

School Violence and Safety Measurement in Historical Context
School violence and safety instruments predominately include items developed to estimate population trends to monitor public health conditions. Although items on population-based surveys, such as the Youth Risk Behavior Survey (YRBS; Centers for Disease Control and Prevention [CDC], 2004, 2009, 2010), have undergone content scrutiny by researchers, they are not typically subjected to rigorous psychometric analyses. This is because, in part, their original purpose was to obtain population estimates of behaviors and not to assess individual differences, or to
provide information about the root causes of school violence. Nonetheless, researchers often use these surveillance surveys to examine individual differences and the association among risk and health behaviors. However, because the items were not developed for this specific purpose, their sensitivity and validity for complex measurement purposes are not established.

To date, only the Survey of School Crime and Safety was originally and specifically designed to assess school safety and violence conditions; however, it is completed by school principals and does not produce student, staff, or parent-level information. Although many of the most prominent sources of information about school safety and violence have come from sources that were not specifically designed for this purpose, modifications of instruments to include school-context items still provide the best information available about school violence and safety and is used for the reports, Indicators of School Crime and Safety (e.g., Dinkes, Kemp, & Baum, 2009) and Crime, Violence, Discipline, and Safety in U.S. Public Schools (Neiman & DeVoe, 2009). There is overwhelming momentum to suggest that surveillance instruments will continue to be widely used because they have established baseline information that is being used to inform public policy.

Measurement Practices of Recent School Violence and Safety Research

To provide a perspective on recent school violence research measurement practices, we conducted a search of peer-reviewed journal articles with the descriptor of “school violence” for 2009 in the PsycInfo database (a list of these studies and review summary table is available from the authors). This search yielded 59 articles focused on youth in kindergarten through Grade 12 in the following categories: (a) qualitative or theoretical discourse regarding school shootings (n = 18), (b) relations between variables (n = 16), (c) bullying and victimization (n = 7), (d) school violence prevalence (n = 6), (e) other theory/review (n = 4), (f) prevention and intervention studies (n = 4), and (j) book reviews and journal editorials (n = 4). Approximately half (n = 33) of these articles were empirical studies. Although this review was designed to be inclusive, it yielded only a sampling of articles on the topic of school violence. For example, the search only identified 12 of 29 articles published in the Journal of School Violence in 2009. Identifying critical scholarship in the field of school violence would be facilitated with commonly applied keywords and descriptors such as “school violence,” “school safety,” and where appropriate “measurement.”

Standards for reporting quantitative methods in the social sciences necessitate a theoretical rationale for the selection of variables, clearly defined variables, and established psychometric characteristics of measures (e.g., Osborne, 2010). To examine the degree to which recent studies use psychometrically sound measures, we examined 32 (one was inaccessible) of the empirical articles. First, we counted the different types of measures used in each study. Each study could only contribute one count per measure, but multiple measures could be counted per study. The most commonly used type of measurement was self-report (n = 24; e.g., Rose, Espelage, & Monda-Amaya, 2009), distantly followed by teacher report (n = 4; e.g., Yavuzer, Gundogdu, & Dikici, 2009), and neighborhood crime records (n = 2; e.g., Cornell, Sheras, Gregory, & Fan, 2009). The following measurement strategies were used in only one study: mapping tools, interviews and focus groups, observations, sociometric questionnaire, card sort, disciplinary records, trained professional rating, and participatory action research. Self-report was by far the most popular type of measure employed, particularly when examining relations between variables. Although most self-report studies focused on youth, one of the self-report measures was for teacher reports of their bully experiences and how these experiences impacted their school’s bullying interventions (Kokko & Pörhölä, 2009).

Second, we examined the methods section of each article to evaluate the psychometric properties of measures used. Although the majority of the studies used established measures (e.g., California School Climate & Safety Survey [Furlong, 1996; Furlong et al., 2005] or the YRBS)
only 2 studies (Rose et al., 2009; Saylor & Leach, 2009) provided both reliability and validity information of the school violence scales used. Most (n = 17) reported reliability, but no validity information (e.g., Cornell et al., 2009). In a few cases (n = 3; e.g., Estévez, Murgui, & Musitu, 2009), measures were modified for the specific study and the resulting reliability, but no validity, data were reported. Several studies reported no psychometric information (n = 5; e.g., Chen & Astor, 2009). Other scholars developed their own measures as part of the study (n = 8) but less than half (n = 3; e.g., Sela-Shayovitz, 2009) reported psychometric properties of the resulting scale. One study adapted a measure by translating it into different languages, but did not report any psychometric data regarding the final scales (Linares, Dias, Fuentes, & Acién, 2009). Overall, it was clear that some scales used in school violence research are well developed with multiple citations reflecting prior development and use. Yet, in many cases, scales were developed specifically for one study, at times without any reliability or validity analysis presented.

This examination of recent studies in the area of school violence indicated limited consensus about the use of common measures to examine school violence. In 32 studies, 31 different approaches or scales were used to measure school violence. When established measures were used, oftentimes, psychometric properties for the study sample were underreported (e.g., Turner, Powell, Langhinrichsen-Rohling, & Carson, 2009). Measures were adapted for unique purposes without revalidation and even translated without documenting new psychometric components (e.g., Linares et al., 2009). These recent research practices demand attention because they suggest that the validity of school violence scholarship is questionable due to its measurement practices. Our analysis also revealed popular and neglected areas of study. Discourse regarding school shootings was highly popular and dominated more than a quarter of the articles. The study of bullying and victimization was also popular. Neglected, however, were several areas needing additional scholarship such as school violence prevention, crisis intervention, gang interdiction, school discipline practices, and the influence of school climate on school safety.

School Violence and Safety Measurement Instruments

Despite the need for improved measurement practices, a body of evidence is available for some commonly used instruments. This section provides an overview of procedures to assess school violence and safety. The most frequent way of measuring safety and violence-related concerns is through self-report surveys, including population-based surveillance surveys. Other common methods include mandated districtwide reporting of school discipline referrals. Additional methods such as mapping tools and sociometric questionnaires are used less frequently, and thus, are not detailed here.

Comprehensive School Violence and Safe School Instruments

National surveys of students’ self-reported behaviors and experiences are the most common form of data collection regarding school violence and provide information about the widespread prevalence of a problem. However, because this information is aggregated over broad contexts, it is useful for estimating state or national trends, but its applicability to local conditions is restricted. At the school level, students and teachers may be more interested in how their school compares to similar schools in the community or to examine changes over time in response to interventions. School- and district-level assessment of school violence indicators is crucial to indicate specific local needs, as well as to provide a baseline against which treatment efficacy can later be evaluated (Benbenishty, Astor, & Zeira, 2003). Moreover, individual development occurs within the context of school and community influences, and large-scale surveys rarely account for contextual variables (other than geographic region and city size). Therefore, a person’s
characteristics inherently reflect many environmental influences over a protracted time period, so that attempts to partition out the effects of environment from individual-level variables are unlikely to yield large effects. Oftentimes, the sophistication of the analytic methods used to examine predictors of aggression and violence (such as hierarchical linear modeling) supersedes the psychometric sophistication of the measures used to gather the data (Sullivan, 2002).

**Youth Risk Behavior Surveillance Survey (YRBS)**

The YRBS was originally developed in the late 1980s as a youth health risk surveillance system (e.g., see Kann, 2001; Kann et al., 2000). It is an anonymous self-report survey administered to students within their classroom. The YRBS is administered biennially to a representative sample of United States high school students. Its content focuses on health-risk behaviors that may result in later disability, mortality, morbidity, and/or significant social problems (CDC, 2009) including alcohol and drug use, unintended pregnancies, dietary behaviors, physical activity, and sexual activity. Results are used to monitor health-risk behaviors among high school students, evaluate the impact of various efforts to decrease the prevalence of health-risk behaviors, and monitor the progress of national health objectives (e.g., CDC, 2010). However, there is limited reliability and validity data to support the use of the YRBS (beyond its epidemiological purpose) to describe characteristics of youths who engage in high-risk behaviors (see CDC, 2004). For instance, the stability of weapon-carrying items, as measured through test-retest reliability, was low, and reliability analyses did not address the different time frames of items (i.e., 30-day vs. lifetime; Brener et al., 2002). Finally, the validity of the responses of youth who selected the most extreme response options was not supported (see Furlong, Sharkey, Bates, & Smith, 2004).

**California Healthy Kids Survey (CHKS)**

The CHKS is a series of assessment modules developed by WestEd’s Human Development Program in collaboration with Duerr Evaluation Resources for the California Department of Education (California Department of Education [CDE], 2010). Items for the CHKS were taken from the YRBS and the California Student Substance Use Survey. The CHKS provides schools with a method of collecting ongoing youth health and risk behavior data and, since 1998, has been successfully administered across the United States and internationally. Three separate versions of the CHKS are available for use at the elementary, middle school, and high school levels (CDE, 2010) and supplemental topics may be added to the core surveys. The surveys contain items concerning youth nutrition and physical activity, sexual behavior, exposure to prevention and intervention activities, risk and protective factors, alcohol, tobacco and other drug use as well as items specific to violence, school safety, gang involvement, and delinquency (CDE, 2010). The surveys include items to assess the truthfulness of each respondent’s answers. The results can be compared to results from the YRBS and various health-risk surveys. Benefits of using the CHKS include: meeting categorical program requirements (such as Title IV), identifying program goals and high-risk groups (e.g., CDE, 2010), and identifying risk and protective factors (through its Resiliency and Youth Development Module), which has an extensive report of its psychometric properties (Hanson & Kim, 2007).

**Mandated Districtwide Reporting**

The federal Gun-Free Schools Act of 1994 (GFSA; 1994) imposed specific reporting requirements for each state regarding the number of students who engage in a variety of violent behav-
School Violence Measurement Issues

Unfortunately, data submitted by districts were not comparable (Kingery & Coggeshall, 2001) because (a) some states reported more detailed information, (b) districts differed in how they categorized violent behavior, (c) inconsistencies in time periods assessed interfered with comparing data, and (d) federally-mandated definitions of the classifications of weapons and violent behavior were often not specified in student surveys (for instance, the GFSA definitions included bombs and gas as firearms). Kingery and Coggeshall (2001) also cited other problems that interfered with a clear understanding of the data. They noted that (a) students may hide serious incidents from staff, (b) school personnel may fail to detect many violence incidents or only document observed infractions, (c) staff are not often trained in a standardized fashion, and (d) school personnel may feel pressured to underreport violence so not to reflect negatively on their school or district.

Discipline Referrals

School discipline data, such as office referrals, suspensions, and expulsions, have the potential to provide valuable information about students’ risk for future infractions. Sugai, Sprague, Horner, and Walker (2000) observed that all schools collect discipline data to some extent, so discipline data may be the most effective way for school systems to understand the relationships and behavioral contexts that disrupt schools. Discipline data can show (a) which student behaviors are of greatest concern; (b) whether or not there are disproportionate referrals by gender, ethnicity, or special education status; and (c) whether or not there are disproportionate referrals made by certain teachers or during certain periods of instruction (e.g., recess, P.E., reading, and math). This information may be used to identify targets for intervention. However, methods for collecting discipline data need careful consideration and procedures need to be standardized across classroom, playground, and other school settings.

Morrison, Peterson, O’Farrell, and Redding (2004) note that although these data are easily obtained, little is known about the reliability or validity for predicting future aggressive acts with either school-level or individual-level data. Additionally, Morrison and Skiba (2001) caution that school discipline data are not as straightforward as they might appear. Behavior referrals, suspensions, and other disciplinary actions reflect not only students’ behavior, but also teachers’ tolerance for disruptive behavior; teachers’ skills in classroom management; administrative discipline policies; and other classroom, school, and community factors, although they often fail to document the contribution of these environmental influences. Thus, predicting disruptive and violent behavior from school discipline data is problematic because it must account for these multiple levels of influence.

Persistent School Violence and Safety Measurement Issues

Although numerous studies examine factors affecting the measurement accuracy of other health-risk behaviors, such as substance use and dietary behavior, studies examining behaviors related to violence are scarce (Brener, Billy, & Grady, 2003). Researchers have identified threats to the validity of school violence survey data and recommended a variety of design practices for enhancing survey accuracy. These strategies include maintaining confidentiality or anonymity of responses (Ong & Weiss, 2006), verifying the survey’s reading level (Stapleton, Cafarelli, Almario, & Ching, 2010), checking for the honesty of answers (Rosenblatt & Furlong, 1997), including sufficient numbers of items to measure a given construct (Bradley & Sampson, 2006), and asking questions about past experiences in ways that are most likely to elicit accurate recall (Coggeshall & Kingery, 2001). Despite the need for enhanced methodological rigor, information about the reliability and validity of methods used to collect self-report surveys is limited.
Implausible Responding Patterns

Investigators need to address the possibility that students may not respond honestly. This issue is related to, but different than, the occurrence of missing or spoiled survey responses to specific questions. For example, with the YRBS, one report (Brener, Kann, et al., 2004) indicated that unusable responses were obtained for 0.4% (age) to 15.5% (suicide attempt) of all 2003 YRBS responses. Data screening methods should detect response inconsistencies or implausibly extreme patterns of responding, which may indicate not just that specific item responses were spoiled, but that the validity of the entire set of responses is questionable. Although data validity screening should be an essential element of all large-scale student surveys, unfortunately, few school violence surveys have included such safeguards (Cornell & Loper, 1998; Kingery & Coggeshall, 2001), and many others do not report whether or how they evaluated the quality of student responses (Furlong, Morrison, et al., 2004). To investigate this issue, researchers from the University of Oregon examined dubious and inconsistent responders to a state survey and found that 1.9% of the cases were eliminated (Oregon Healthy Teens, 2004). This finding supports the general utility of using students’ self-reports; however, it suggests that surveys that do not apply such standards may produce results that overestimate very low incidence risk behaviors because inconsistent responders report higher rates of risk behaviors (Cornell & Loper, 1998; Rosenblatt & Furlong, 1997).

The CHKS uses high quality data screening and is a carefully developed instrument, having undergone more than six years of rigorous development and review by a standing panel of independent experts. Analysis of CHKS data includes a data-validation procedure that screens records from the data set that meet a combination of criteria, including inconsistent patterns of responding, implausible reports of drug use, the endorsement of a fictitious drug, or failure to assent to having answered survey questions honestly. The CHKS procedure is rigorous and typically produces a listwise case elimination of about 2%–3% of cases (e.g., O’Brennan & Furlong, 2010). In their examination of responses to the California School Climate and Safety Survey, Rosenblatt and Furlong (1997) compared a group of students who failed reliability and validity checks to a matched control group. They found systematic bias in the way failed students responded to the survey, including higher ratings of school violence victimization, campus danger, poor grades, and few good friends. By contrast, comparison students were more likely to be detected by the social desirability item.

Although surveyors are often concerned with participants responding in a socially desirable manner, one concern is that youths involved with antisocial and aggressive peers will exaggerate their involvement in delinquent activities as an alternative form of social desirability. For instance, Cross and Newman-Gonchar (2004) examined responses on three surveys: the Colorado Youth Survey, the Safe Schools/Healthy Students survey (SS/HS), and a Perception Survey. When examining the pattern of extreme and inconsistent responses, they found that not only did rates vary substantially by survey and by school, but extreme and inconsistent responses inflated rates of violent behavior much more at one school than at the other school. Even small percentages of questionable responses had the ability to inflate estimates of risk behavior substantially. For example, by excluding the fewer than 3% of responses to the Colorado Youth Survey that were suspect, Cross and Newman-Gonchar eliminated 70% of reported incidents of gun-carrying. It is important to note that such small differences in rates are important for low incidence behaviors (such as school gun possession) because even high-quality large-scale surveys such as the YRBS are designed to have a 5% confidence interval (Brener, Kann, et al., 2004). However, school violence researchers rarely consider the effects of implausible responses on their study results.
Survey Administration Procedures

Cross and Newman-Gonchar (2004) examined the impact of survey administrator training on student response patterns. District prevention specialists coached a group of teachers about the importance of collecting quality data, and they were taught to explain the uses and importance of the data to their students and to ask them to respond honestly. Rates of invalid responses to surveys administrated by trained versus untrained teachers were compared and results indicated that the trained administrators obtained far lower rates of highly suspect responses (3%) than the untrained administrators (28%). Related to this finding, researchers using large sample databases rarely present data related to adherence with a standard administration protocol by school staff.

In addition to administration procedures, responses may be influenced by survey format. Turner et al. (1998) provided one of few investigations regarding how traditional paper-and-pencil formats and computer assisted presentation influence response rates. Based on similar studies about youth substance use, computer presentations were expected to produce higher self-report rates. Turner and colleagues found that the computer format produced substantially higher prevalence rates than did the paper-and-pencil format for weapon-carrying, acts of violence, and threatened violence. The authors explained that the computer format, with the benefit of audio presentation and computerization, was likely to promote more accurate responding for sensitive questions. This explanation is supported by evidence of similarly high rates from studies that rely on retrospective accounts by adults regarding sensitive adolescent behaviors.

In a related study, researchers from the CDC compared paper-and-pencil versions to several online versions of the YRBS (Denniston et al., 2010). They found that seeking passive parental consent for both the hard copy and the online options produced similar usable response rates. One difference was that when the survey was completed online the youth were more likely to say that they did not think the answers were private, given the visibility of monitors in the computer lab administration setting. The students were also given one other option. They were given a card with the URL for the survey and asked to complete it on their own time over the following two weeks. They received a reminder after one week. When given the option and asked to do it on their own time, only 24% of the students completed the survey. Although not mentioned by these researchers, this was interesting because this latter finding suggests that the level of intrinsic motivation to complete school violence and safety surveys among adolescents may be low.

In an important study, Hilton, Harris, and Rice (2003) examined the consistency of youth self-reports of violence victimization and perpetration (not school violence) and compared prevalence rates derived from traditional paper-and-pencil reports to those provoked by the same experience modeled in an audio vignette. They found that the same youths reported 2 to 3 times more violence perpetration and victimization using the self-report format. This finding, considered in relation to other studies (e.g., Hilton, Harris, & Rice, 1998), led these researchers to conclude that, “...although we would not suggest that standard paper-and-pencil surveys yield completely inaccurate data, the accumulated evidence, including these results, calls on researchers in the field of interpersonal violence to redouble efforts to demonstrate and improve the factual accuracy of their primary dependent measures” (Hilton et al., 2003, p. 235).

Item Wording

It is also important to evaluate if all respondents understand school violence survey questions in the same way. Brener, Grunbaum, et al. (2004) found that differences in item wording across three national surveys resulted in significantly different rates of behavior. One challenge with surveys of school violence and safety is that questions are worded so broadly as to leave a great deal of interpretation to individual responders (Cornell & Loper, 1998). A survey question that
asks students if they carry a weapon without defining the word “weapon” has the potential to result in students reporting their hunting rifles or pocketknives as weapons. For example, the SS/HS (Safe Schools Healthy Students, 2004) item, “During the past 12 months, how often have you been picked on or bullied by a student on school property?” is ambiguous for several reasons. First, it is unclear whether being picked on and being bullied are to be interpreted as separate experiences, or if they are implied to be equivalent. Beyond this, the item does not define “bullying” in a way that is consistent with best research practices. It could be argued that this item does not measure “bullying” victimization per se, but rather each student’s understanding of this word. To promote consistent and accurate responses, items should be unambiguous, easy to read, and all terms should be clearly defined (Fowler, 1993).

**Items Response Time Frames**

Many commonly used school violence and safety instruments include items that refer to past behavior using a variety of time frames (e.g., 30 days, 6 months, 1 year). It would seem logical to presume that survey responses in any given month provide a cross-section (in time) of students’ behaviors and experiences and that the reported incidence of these behaviors would be higher over a much longer period of time. For example, if 10% of students report carrying a weapon to school in the past 30 days, one might expect this percentage to be higher if the same sample of students were asked to report about past year weapon possession. However, no published research has confirmed this. According to Schwartz (1999), asking about past month incidents is likely to convey to respondents that researchers want to know about less serious but common events. In contrast, students might interpret asking about fights in the past year as seeking information about less frequent but more serious fights (Schwartz, 1999).

To explore this issue, Hilton and colleagues (1998) examined differences in self-reports across 1-month, 6-month, and 1-year time periods. They reported “… standard self-reports of interpersonal violence were insensitive to the specified time frame; for example, participants reported almost the same number of violent acts in the past month as in the past year, something that could not be factually true” (p. 234). Similarly, based on their review of what is known about influences of response time referents, Brener et al. (2003) concluded that multiple factors affect student recollection of school violence experiences.

Survey designers should be mindful that respondents often find it difficult to accurately remember past events (Cornell & Loper, 1998; Fowler, 1993). For this reason, survey questions that ask about past behavior or experiences can incorporate any of a variety of techniques for enhancing recall. Converse and Presser (1986) recommend: (a) using simple language; (b) asking about experiences within a narrow reference period (within no more than a 6-month period); (c) using memory landmarks (e.g., asking about behavior “since the beginning of the school year” or “since New Year’s”); and (d) stimulating recall by describing concrete events (“Instead of asking respondents if they have experienced ‘assaults,’ for instance, … ask if anyone … used force by grabbing, punching, choking, scratching, or biting,” p. 22). Unfortunately, school violence and safety research has not yet thoroughly attended to these measurement concerns.

**Item-Response Options**

Not only do surveys differ in the response time frame, but they also vary in the number and type of response options offered. Often, items that appear on one survey are included on other questionnaires with a different number of response options, without stating a rationale for the change in response options. For example, the YRBS asks, “During the past 30 days, on how many days did you carry a weapon such as a gun, knife, or club on school property?” The following five
response options are offered; 0 days, 1 day, 2 or 3 days, 4 or 5 days, and 6 or more days. However, when applied by Farrell and Meyer (1997) to evaluate the effectiveness of a violence prevention program, the item included six response options: Never, 1–2 times, 3–5 times, 6–9 times, 10–19 times, and 20 times or more. Although it might appear that similar items with different numbers of response options yield equivalent results, in fact, this is not the case (Schwartz, 1999). In addition, Schwartz noted that survey respondents also are most likely to choose response options near the middle of response scales.

School Violence and Safety Instrument Psychometric Properties

The survey method typically used to assess school safety and violence seeks to identify indicators that increase the likelihood of negative outcomes. This practice differs substantially from that most often used in school psychology, which establishes in advance the psychometric properties of a scale for identification or diagnosis purposes (Cornell & Loper, 1998). Unfortunately, some reports have used survey instruments for identification purposes without adequately examining the psychometric properties of the scales. There is limited research examining the reliability and validity of school violence and safety instruments.

Stability of Responses

As with any self-report measure, measures of violence and safety should have rigorous psychometric testing to establish reliability and validity (Rosenblatt & Furlong, 1997). Given that school violence and safety instruments typically do not measure latent traits, the reliability of greatest interest is the consistency of responses over a brief interval. Unfortunately, testing of response stability is infrequent. Brener et al. (2002) completed the only study to date that has examined the reliability of the YRBS items that inquired about school-associated behaviors. They examined the responses of 4,619 of 6,802 eligible students, derived from a convenience sample, who completed the YRBS twice over a 2-week time period. They subsequently converted the responses of all items into binary format to compute a \( \kappa \) statistic, a measure of response consistency that is corrected for chance agreement. \( \kappa \) coefficients ranged from .41 to .68 (in the moderate to substantial range; Landis & Koch, 1977) for the four YRBS items that directly assess school violence content. The incidence of one behavior (“Injured in a physical fight one time in the past 12 months”) was significantly higher at time 2 compared to time 1. Though Brener et al. concluded that, in general, the YRBS is a reliable instrument and this study is cited often to justify the use of the YRBS, there are several methodological questions with this study, particularly when school violence items are examined. These issues pertain to the appropriateness of excluding inconsistent responses prior to analysis; converting responses to binary format and using the \( \kappa \) statistic; using a 14-day (average) retest period for items with 30-day or 12-month reference timeframes; and interpreting \( \kappa \) coefficients for items with different reference timeframes (e.g., past 30 days and past 12 months). As a result, this analysis likely generated the most favorable possible test-retest reliability. Even so, the reliability coefficients for the school items were only in the moderate range.

Validity of Measures

Skiba et al. (2004) noted that instruments can contribute significantly to the field of school violence and safety only when they include the complex and representative set of factors involved. They recommended using empirical methods such as factor analysis to justify inclusion of items in surveys, rather than relying only on professional judgment, as done in the past. Regarding the
content of school safety and violence measures, Skiba et al. pointed out that many surveys focus on significant acts of violence, such as weapon carrying. However, although extreme violence is the end result within a context of unsafe school climate, it may not provide enough variance to yield meaningful comparisons, particularly for typical schools. For this reason, Skiba et al. stated that researchers need to focus less on serious violent acts, especially as smaller discipline problems are more frequent and are part of a common trajectory towards future antisocial behavior.

A few researchers have developed school violence and safety instruments using psychometric approaches to scale development (Furlong, Sharkey, et al., 2004; Skiba et al., 2004). Although these instruments are not appropriate for surveillance style surveys, they have potential to provide options when used as outcome measures in local evaluations and studies using controlled experimental designs. Recently, Greif and Furlong (2004a, 2004b) extended the psychometric approach by using item response analysis to create a unidimensional school violence victim scale. Such a scale could be used to track victimization experiences across time. Table 20.1

<table>
<thead>
<tr>
<th>Type</th>
<th>Intended Use</th>
<th>Limitations</th>
<th>Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Surveillance</td>
<td>Obtain population estimates broadly about youth risk.</td>
<td>Not subjected to rigorous psychometric analysis; limited information about school violence and safety; does not inform individual development.</td>
<td>Conduct psychometric analysis for new purpose prior to extending their use.</td>
</tr>
<tr>
<td>School Violence Items in Surveillance Surveys</td>
<td>Obtain populations estimates about school violence and safety.</td>
<td>The most rigorous survey is only administered to school principals and does not produce student, staff, or parent-level information; does not inform individual development.</td>
<td>Include items designed to specifically assess school violence and safety in more broadly administered assessments. Make sure reliability and validity have been tested for intended purposes.</td>
</tr>
<tr>
<td>School Violence and Safety Measurement Instruments</td>
<td>Measure latent constructs of youth violence.</td>
<td>Threats to validity such as reading level, honesty of answers, response options, recall, and extreme responses.</td>
<td>Advance knowledge of how to overcome threats to validity by studying these issues. Create and validate more advanced measurement tools with the knowledge gained.</td>
</tr>
<tr>
<td>Mandated Districtwide Reporting</td>
<td>Track student violence across schools.</td>
<td>Data submitted by districts are not comparable.</td>
<td>Identify an alternative method or implement extensive training, standardization, and incentive for accurate and comparable reporting.</td>
</tr>
<tr>
<td>Discipline Referrals</td>
<td>Examine school safety; evaluate if there are disproportionate referrals; predict students’ risks for future infractions.</td>
<td>Lack of standard procedures within and across schools; lack of psychometric scrutiny.</td>
<td>Limit cross analysis of data from different sources unless discipline procedures, reporting, and data collection are standardized within and across settings. Evaluate data for reliability and validity.</td>
</tr>
</tbody>
</table>
provides a review of the school violence and safety measures along with their limitations and recommendations for improvement.

Conclusion

Efforts to refine school violence and safety measurement practices are needed. Although more sophisticated techniques are being employed, such as web administration with programmed skip patterns, it is essential that researchers continue to examine the fundamental aspects of measurement procedures. Significant limitations currently exist, perhaps most notably the continued and sustained use of items from instruments with little, or no known, validity evidence. Researchers and consumers of research should be aware of the dangers of drawing conclusions based on measures with limited or poor psychometric properties. Similarly, as new measures are being developed it is critical that researchers do not rely on a “bootstrapping” approach, in which they continually try to validate new measures against known inferior measures until enough evidence is accumulated to demonstrate that the newly developed measure is superior. In this chapter, we provided a review of the current state of research and current available instruments and highlighted many issues affecting measurement accuracy. The key recommendations of the chapter are summarized in Table 20.2. While there is still much to be done, it is hoped that the concepts and guidelines discussed here will lead to improved research in school violence and safety measurement.

Table 20.2 Implications for Practice: Recommendations for the Use of School Violence Self-report Procedures

<table>
<thead>
<tr>
<th>Constructing Surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clearly define key terms</td>
</tr>
<tr>
<td>Keep questions concrete</td>
</tr>
<tr>
<td>Use memory aids when asking about past events</td>
</tr>
<tr>
<td>Include multiple questions to measure each main construct</td>
</tr>
<tr>
<td>Check survey’s reading level for the target population</td>
</tr>
<tr>
<td>Include items that screen for honesty and accuracy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Selecting Existing Surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check reliability and validity for the population to be tested</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Administering Surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensure confidentiality / anonymity</td>
</tr>
<tr>
<td>Train staff to adhere to a standard protocol</td>
</tr>
<tr>
<td>Explain purpose of survey to students and how results are used</td>
</tr>
<tr>
<td>Be prepared to read questions aloud to students and to clarify</td>
</tr>
<tr>
<td>Debrief students afterward: What were they thinking?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reporting Survey Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present information separately to students, staff, parents</td>
</tr>
<tr>
<td>Solicit perceptions of survey data</td>
</tr>
<tr>
<td>Brainstorm alternative courses of action</td>
</tr>
<tr>
<td>Maintain a collaborative, team-oriented stance</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Next Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link assessment to planning and intervention</td>
</tr>
<tr>
<td>Choose activities and programs with empirical support</td>
</tr>
<tr>
<td>Reassess students periodically to measure program effectiveness</td>
</tr>
</tbody>
</table>
References


School Violence Measurement Issues


