IMPROVEMENT OF THE MEASUREMENT OF BULLYING

IN SCHOOLS

A Dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree
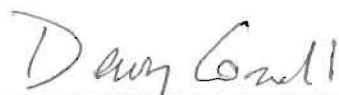
Doctor of Philosophy

---

by

Michael W. Baly, M.Ed.

August 2013

Programs in Clinical and School Psychology
Curry School of Education
University of Virginia
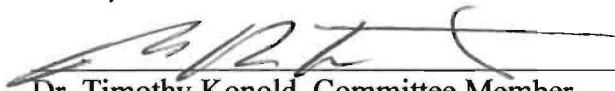Charlottesville, Virginia

APPROVAL OF THE DISSERTATION

This dissertation, "Improvement of the Measurement of Bullying in Schools," has been approved by the Graduate Faculty of the Curry School of Education in partial fulfillment of the requirements for the degree of Doctor of Philosophy.
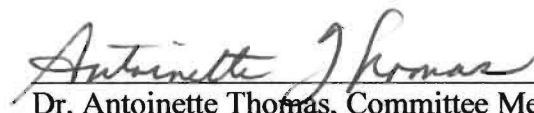

_____

Dr. Dewey G. Cornell, Committee Chair

_____

Dr. Nancy Deutsch, Committee Member

_____

Dr. Timothy Konold, Committee Member

_____

Dr. Antoinette Thomas, Committee Member

8/22/12 Date

Abstract

This dissertation is comprised of three manuscripts and presents a line of research aimed at improving the measurement of bullying in schools.

The first manuscript investigated the impact of an educational video on self-reports of bullying. A sample of 1,283 middle school students in randomly assigned classrooms either watched or did not watch an educational video about bullying prior to completing a self-report bullying survey. Students who watched the video reported 32% less social bullying victimization and boys who watched the video reported 54% less physical bullying victimization and 68% less physical bullying of others. These results indicate that self-report surveys could yield inflated estimates of the prevalence of bullying if students are not adequately educated about the distinction between bullying and other forms of peer conflict.

The second manuscript examined the use of validity screening items on longitudinal adolescent survey data. This longitudinal study examined the response patterns of 382 students who completed self-report surveys each fall and spring for three years of middle school (grades 6-8). Approximately 10% of students in each wave indicated on validity screening questions that they were either not telling the truth or paying attention. Hierarchical linear modeling analyses found that invalid responding students were more likely to be referred for disciplinary infractions, made poorer grades, and performed less well on standardized tests than other students. This study provides new information about student survey validity and appears to be the first to identify characteristics of students who generate invalid response patterns.

The third manuscript examined the longitudinal stability and cumulative impact of bullying victimization through both peer- and self-report measures. Using the same sample as the second study, this study assessed victimization experiences in the fall and the spring of grades 6, 7, and 8. Students who had higher levels of victimization throughout middle school were more likely to be referred for disciplinary infractions, make poorer grades, perform less well on standardized tests, engage in more risk behaviors, and have less favorable perceptions of school climate. This study provides new information about the cumulative impact of peer- and self-reported bullying across middle school.

DEDICATION

I dedicate this dissertation to my fiancé and best friend, Sage. I could not have done this without you.

ACKNOWLEDGEMENTS

thank you to grandmother and Hank for encouraging me to follow my heart and do what I love.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Overview of Three Manuscripts Examining the Improvement of the Measurement

of Bullying in Schools

Michael W. Baly

University of Virginia

Overview of Three Manuscripts Examining the Improvement of the Measurement

of Bullying in Schools

The well-documented negative effects of bullying on children and adolescents (e.g., Boulton, Trueman, & Murray, 2008; Lopez & Dubois, 2005; Swearer, Espelage, Vaillancourt, & Hymel, 2010) have prompted an array of intervention programs aimed at decreasing the rates of bullying in schools (Ttofi & Farrington, 2009; Vreeman & Carroll, 2007). As a result, numerous studies have focused on measuring the prevalence of bullying in order to assess the effectiveness of intervention efforts (Wang, Ianotti, & Nansel, 2009; Ttofi & Farrington, 2009). A recent national study found that 36% to 43% of middle school students reported being bullied at school during the school year (DeVoe & Bauer, 2010).

Despite the widespread use of anonymous self-report measures of bullying, there is only modest agreement across instruments and little evidence of criterion-related validity using independent sources of information (Branson & Cornell, 2009; Lee & Cornell, 2010). Studies have reported mixed effectiveness of intervention programs (Ttofi & Farrington, 2009), but how can researchers be sure of the effectiveness of interventions if they cannot trust the accuracy of their assessment methods?

The purpose of this manuscript-style dissertation is to investigate the limitations of our current methods for measuring bullying and to examine the effectiveness of alternative strategies. Each manuscript is in a different stage of journal publication. The dissertation version of the studies may not be equivalent to the final published version due to corrections recommended by journal reviewers. The first manuscript, a completed study that has been published in the *Journal of School Violence* (Baly & Cornell, 2011),

tested whether students in classrooms that were randomly assigned to watch an educational video about bullying would report less bullying than their peers who did not see the video. The second manuscript examined the use of validity screening items on longitudinal adolescent survey data. It was submitted for publication and provisionally accepted by a peer-review journal. Michael Baly is the third author on this paper and contributed by collecting data, constructing the unique longitudinal data set, conducting descriptive statistical analyses, consulting on the methods, and assisting in the editing process. The purpose of this study was to gain new insight into students who mark their surveys in an invalid manner and examine the value of validity screening. Finally, the third paper examined the stability and cumulative impact of bullying victimization measured through both self- and peer-reports across middle school. This manuscript was accepted for publication by *Psychology in the Schools* and is currently in press (Baly, Cornell, & Lovegrove, In press).

First Manuscript

Bullying can generally be defined as aggressive behavior that is intentional and repetitive, and involves an imbalance of power between the perpetrator and victim (Olweus, 2010). Considering each element – especially the concept of a power imbalance – to correctly identify bullying is a challenging task for both participants and observers. Discrepancies in power can be quite subtle in social bullying, as the aggressor's advantage is implied because of social status or popularity. Students who are not able to distinguish bullying from ordinary peer conflict between individuals of comparable size and status may generate inflated prevalence rates by being overly inclusive in their reporting.

The first manuscript of this dissertation (Baly & Cornell, 2011) investigated whether an educational video might improve students' understanding of bullying. Self-report measures of bullying are limited by how well students distinguish bullying from other forms of peer conflict. A recent study (Cornell & Mehta, 2011) found that many middle school students who read a survey definition of bullying and then identified themselves as victims of bullying were later found by school counselors to be involved in ordinary peer conflict that did not involve the power differential characteristic of bullying. Baly and Cornell (2011) conducted an experiment to determine whether an educational video could affect self-reports of bullying. Randomly assigned classrooms of 1,283 students from three middle schools either watched or did not watch the video designed to distinguish bullying from ordinary peer conflict prior to completing a self-report bullying survey. Compared to the control group, students who watched the video reported 32% less social bullying victimization. Boys who watched the video reported 54% less physical bullying victimization and 68% less physical bullying perpetration. These results indicate that student self-reports could yield inflated estimates of the prevalence of bullying if students are not adequately educated about the distinction between bullying and other forms of peer conflict.

Another limitation of self-report surveys is that students primarily take them on an anonymous basis. Anonymous surveys are used in large part to encourage students to be more forthcoming (Olweus, 2010, Solberg & Olweus, 2003); however, multiple studies have found little or no difference in reported behaviors on anonymous and confidential surveys. Anonymous surveys make it impossible to conduct studies that examine trends in bullying and other behaviors over time (Cornell & Bandyopadhyay, 2010).

Additionally, they prevent researchers from linking survey results from other student information (Cornell & Cole, 2011).

Second Manuscript

Anonymous self-report surveys are widely used to measure the prevalence of adolescent bullying and risk behavior, as well as to assess the effectiveness of school-based prevention and intervention programs (Johnston, O'Malley, Bachman, & Schulenberg, 2010; Robers, Zhang, & Truman, 2010; Sharkey, Furlong, & Yetter, 2006). Despite concerns about the possibility for under-reporting of drug use and delinquent behavior (e.g., Hindelang, Hirschi, & Weiss, 1979), few studies have addressed the issue of over-reporting (Fan et al., 2006; Sharkey et al., 2006). Many surveys do not account for the possibility of rebellious adolescents either inflating reports of risky behaviors or not taking the survey seriously and marking it haphazardly. Research suggests that validity screening questions can identify students who give seemingly inflated reports of their risky behavior and negative views of school climate (Cornell, Klein, Konold, & Huang, 2012; Miller, Fan, Christensen, Grotevant, & van Dulmen, 2000). However, due to the use of anonymous surveys, previous work has been limited in its ability to investigate trends over time or distinguish valid from invalid responders based on independent sources of information.

The purpose of the second manuscript's study was to provide new information on the survey response patterns and school adjustment of students who are identified as invalid responders. The study used a unique dataset in which each survey was given a code number that allowed student responses to remain confidential while being tracked over time and connected to independent sources of information. A sample of 382 middle

school students who participated in semi-annual school surveys were tracked across three years of middle school. Data were collected over five school years and the sample consisted of three cohorts of students who attended the school in sixth, seventh, and eighth grades. Students completed the School Climate Bullying Survey (SCBS; Cornell, 2011), as well as eight risk behavior items derived from the Youth Risk Behavior Surveillance System (California Healthy Kids Survey, 2010) in the fall and spring of each year. School records, including absences, discipline infractions and suspensions, grades, and standardized test scores were collected at the end of each school year.

A sizeable minority of students—slightly more than one-third (35.6%) of the study sample—generated at least one invalid survey during their three years of middle school. Latent class analysis suggested that there were only two reliable patterns of validity responding: those who never generated an invalid survey and those who occasionally generated an invalid survey. These findings suggest that invalid responding is not the habitual practice of a few students, nor an occasional practice of most students, but that a substantial group of students will occasionally complete a survey in an admittedly dishonest or careless manner. Furthermore, this study appears to be the first to examine response validity over repeated survey administrations to the same students. Hierarchical Linear Modeling (HLM) analyses showed that when students identified themselves as dishonest or careless, they were more likely to claim that they consumed alcohol, smoked cigarettes, used marijuana, carried weapons to school, and had thoughts of suicide than students who claimed to answer the survey in an honest and careful manner. These invalid responders also characterized their school climate as more supportive of aggressive attitudes (e.g., agreeing that "If you fight a lot, everyone will

look up to you") and less conducive to help-seeking (e.g., disagreeing that "There are adults at this school I could turn to if I had a personal problem"). These findings add further evidence of the need for more systematic research on the validity of adolescent self-report and support the use of validity screening items.

Third Manuscript

Although numerous studies have reported rates of bullying behavior using a single cross-sectional assessment (DeVoe & Bauer, 2010; Ttofi & Farrington, 2009; Wang, Ianotti, & Nansel, 2009), this approach cannot shed light on the course of bullying and how more persistently bullied students are affected. Because the majority of self-report surveys are anonymous, it is not possible to conduct longitudinal studies to determine whether the same students are bullied over time or whether new students are bullied at different grade levels. Furthermore, it is not possible to determine how more persistently bullied students are affected.

The third manuscript investigated the stability and cumulative impact of bully victimization across three years of middle school. Using the same sample as the second study, this study used peer- and self-report methods to assess victimization experiences in the fall and the spring of sixth, seventh, and eighth grades. Examination of the course of bullying demonstrated that there is more variation than stability in self-reported victimization. There were 49% of students who were never bullied, 39% who were sometimes bullied (i.e., bullied one or two years), and 12% who were persistently bullied throughout middle school (i.e., bullied each year of middle school). Peer-reports of victimization yielded substantially lower rates at each time point than self-reports.

Regression analyses revealed that peer- and self-reports of victimization predicted different types of outcomes. More peer-reports of victimization were associated with higher disciplinary infractions and lower grades and standardized test scores in eighth grade. In contrast, self-reports predicted more aggressive attitudes, less willingness to seek help from teachers or other staff at school, and more risk behaviors. These findings shed light on the need for more longitudinal research that can track student victimization experiences over time and support the view that both self- and peer-reports yield useful, non-redundant information.

Together, the three manuscripts making up this dissertation contribute to our understanding of the factors influencing self-report surveys and the current limitations of the measurement of bullying in schools. The first manuscript underscores the fragility of student self-report as a means of assessing the prevalence of bullying and the effects of intervention efforts. The second manuscript supports the use of validity screening items to improve data quality. Lastly, the third manuscript demonstrates the value of confidential surveys and how tracking students across multiple time points yields important information about the course of bullying and how more persistently bullied students are affected.

References

Boulton, M. J., Trueman, M., & Murray, L. (2008). Associations between peer victimization, fear of future victimization and disrupted concentration on class work among junior school pupils. *British Journal of Educational Psychology, 78*, 473–489. doi:10.1348/000709908X320471

Branson, C. E., & Cornell, D. G. (2009). A comparison of self and peer reports in the assessment of middle school bullying. *Journal of Applied School Psychology, 25*, 5-27. doi:10.1080/15377900802484133

Baly, M., & Cornell, D. (2011). Effects of an Educational Video on the Measurement of Bullying by Self-Report, *Journal of School Violence, 10*, 221-238. doi: 10.1080/15388220.2011.578275

Baly, M., Cornell, D., & Lovegrove, P. (In press). A longitudinal investigation of peer- and self-reports of bullying victimization across middle school. *Psychology in the Schools.*

California Healthy Kids Survey. (2010). *California Healthy Kids Survey, 2007–09. Statewide results: Main report*. San Francisco, CA: WestEd Health and Human Development Program for the California Department of Education.

Cornell, D.. (2011). *The School Climate Bullying Survey: Description and research summary*. Charlottesville, VA: University of Virginia.

Cornell, D. G., & Bandyopadhyay, S. (2010). The assessment of bullying. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective.* (pp. 265–276). New York, NY: Routledge.

Project Overview

Cornell, D. & Cole, J. (2011). Assessment of bullying. In S.R. Jimerson, A. B. Nickerson,

M. J. Mayer, & M.J. Furlong (Eds.), *The handbook of school violence and school

safety: International research and practice (2nd Edition)*. Mahwah, New Jersey:

Routledge.

Cornell, D., Klein, J., Konold, T., & Huang, F. (2011). Effects of validity screening items

on adolescent survey data. Psychological Assessment. Advance online

publication. doi: 10.1037/a0024824

Cornell, D. & Loper, A. (1998). Assessment of violence and other high-risk behaviors

with a school survey. *School Psychology Review*, *27*, 317-330.

Cornell, D., & Mehta, S. (2011). Counselor confirmation of middle school student self-

reports of bullying victimization. *Professional School Counseling*, *14*, 261-270.

DeVoe, J. F., & Bauer, L. (2010). *Student victimization in U.S. schools: Results from the

2007 School Crime Supplement to the National Crime Victimization Survey*

(NCES 2010-319). National Center for Education Statistics, Institute of Education

Sciences, U.S. Department of Education. Washington, DC.

Fan, X., Miller, B., Park, K., Winward, B., Christensen, M., Grotevant, H., & Tai, R.

(2006). An exploratory study about inaccuracy and invalidity in adolescent self-

report surveys. *Field Methods*, *18*(3), 223-244. doi:10.1177/152822X06289161

Furlong, M., Sharkey, J., Bates, M. P., & Smith, D. (2004). An examination of reliability,

data screening procedures, and extreme response patterns for the Youth Risk

Behavior Surveillance Survey. *Journal of School Violence, 3,* 109–130.

doi:10.1300/J202v03n02_07

Furlong, M. J., Sharkey, J. D., Felix, E. D., Tanigawa, D., & Green, J. G. (2010).

Bullying assessment: A call for increased precision of self-reporting procedures.

In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *The handbook of*

bullying in schools: An international perspective (pp. 329–346). New York, NY:

Routledge.

Hindelang, M. J., Hirschi, T., & Weis, J. G. (1979). Correlates of delinquency: The

illusion of discrepancy between self-report and official measures. *American*

Sociological Review, , 995-1014.

Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2010).

Monitoring the future national survey results on drug use, 1975-2009. Volume I:

Secondary school students (NIH Publication No. 10-7584). Bethesda, MD:

National Institute on Drug Abuse.

Lee, T., & Cornell, D. (2010). Concurrent validity of the Olweus Bully/Victim

Questionnaire. *Journal of School Violence, 9*, 56-73.

Lopez, C., & DuBois, D. L. (2005). Peer victimization and rejection: Investigation of an

integrative model of effects on emotional, behavioral, and academic adjustment in

early adolescence. *Journal of Clinical Child & Adolescent Psychology, 34*, 25–36.

doi:10.1207/s15374424jccp3401_3

Miller, B.C., Fan, X., Christensen, M., Grotevant, H.D., & van Dulmen, M. (2000).

Comparisons of adopted and nonadopted adolescents in a large, nationally

representative sample. *Child Development*, *71*(5), 1458-1473. doi:10.1111/1467-

8624.00239

Olweus, D. (2010). Understanding and researching bullying: Some critical issues. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 9–33). New York, NY: Routledge.

Robers, S., Zhang, J., & Truman, J. (2010). *Indicators of school crime and safety: 2010.* (NCES 2011-002/NCJ 230812). Washington, DC: National Center for Education Statistics, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.

Sharkey, J. D., Furlong, M. J., & Yetter G. (2006). An overview of measurement issues in school violence and school safety research. In S. R. Jimerson & M. J. Furlong (Eds.), *The handbook of school violence and school safety: From research to practice* (pp. 121-134). Mahwah, NJ: Lawrence Erlbaum Associates.

Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim questionnaire. *Aggressive Behavior, 29*, 239–268. doi:10.1002/ab.10047

Swearer, S. M., Espelage, D. L., Vaillancourt, T., & Hymel, S. (2010). What can be done about school bullying?: Linking research to educational practice. *Educational Researcher, 39*, 38–47. doi:10.3102/0013189X09357622

Ttofi, M., & Farrington, D. (2009). What works in preventing bullying: Effective elements of anti-bullying programmes. *Journal of Aggression, Conflict and Peace Research, 1*, 13–24.

Vreeman, R. C., & Carroll, A. E. (2007). A systematic review of school-based interventions to prevent bullying. *Archives of Pediatrics & Adolescent Medicine, 161*, 78–88. doi:10.1001/archpedi.161.1.78

Wang, J., Iannotti, R. J., & Nansel, T.R. (2009). School bullying among adolescents in

the United States: Physical, verbal, relational, and cyber. *Journal of Adolescent

Health, 45,* 368–375. doi:10.1016/j.jadohealth.2009.03.021

Abstracts

The first paper of this manuscript-style dissertation was published in the Summer 2011 issue of the *Journal of School Violence.* The second paper was submitted and provisionally accepted for publication. The third paper has been accepted for publication by *Psychology in the Schools* and is currently in press. The abstracts for all three papers are presented here:

Manuscript One

Effects of an Educational Video on the Measurement of Bullying by Self-Report

Abstract: This study of 1,283 middle school students examined the effect of an educational video designed to distinguish bullying from ordinary peer conflict. Randomly assigned classrooms of students either watched or did not watch a video prior to completing a self-report bullying survey. Compared to the control group, students who watched the video reported 32% less social bullying victimization and boys who watched the video reported 54% less physical bullying victimization and 68% less physical bullying of others. These results indicate that student self-reports could yield inflated estimates of the prevalence of bullying if students are not adequately educated about the distinction between bullying and other forms of peer conflict.

Manuscript Two

Invalid Survey Response Patterns among Middle School Students

Abstract: Student surveys are widely used to assess student risk behavior, bullying, and school climate in middle schools; however, because such surveys are usually conducted on an anonymous basis, little is known about the validity of student reports using external, independent criteria. This longitudinal study examined the response patterns of 382 middle school students who completed confidential (not anonymous) self-report surveys each fall and spring for three years of middle school (grades 6-8). Approximately 10% of students in each wave indicated on validity screening questions that they were either not telling the truth or paying attention (termed "invalid responders"). A repeated measures latent class analysis found that students could be classified into a large group (64%) that were never flagged by the validity questions and a smaller group (36%) that occasionally reported not telling the truth or not paying attention. Hierarchical linear modeling analyses found that invalid responding to validity questions was associated with higher self-reported rates of risk behavior and more negative perceptions of school climate. Based on independent criteria from school records, invalid responding students were more likely to be referred for disciplinary infractions than other students. This study provides new information about student survey validity and appears to be the first to identify characteristics of students who generate invalid response patterns.

Manuscript Three

A Longitudinal Investigation of Peer- and Self-Reports of Bullying Victimization

Across Middle School

Abstract: Cross-sectional studies indicate how many students are victims of bullying at a single time, but do not tell us whether the same students continue to be bullied or whether there is a cumulative impact of bullying over time. This study examined the longitudinal stability and the cumulative impact of victimization in a sample of 382 students assessed in the fall and the spring of grades 6, 7, and 8.Victimization assessed by both self- and peer-reports indicated substantial variability in who was bullied, with nearly 51% of students reporting bullying victimization during at least one of the six assessments. The cumulative impact of victimization over three years was demonstrated on grade 8 outcome measures of absences, disciplinary infractions, suspensions, grade point averages (GPA), standardized test scores, reports of youth risk behavior, and perceptions of school climate. This study provides new information about the cumulative impact of peer- and self-reported bullying across middle school.

Effects of an Educational Video on the Measurement of Bullying by Self-Report

Michael W. Baly

Dewey G. Cornell

Curry School of Education

University of Virginia

Address correspondence to Michael W. Baly, Programs in Clinical and School Psychology, Curry School of Education, University of Virginia, Charlottesville, Virginia, USA. E-mail: mwb5g@virginia.edu

Contact: (434) 294-8929

Abstract

This study of 1,283 middle school students examined the effect of an educational video designed to distinguish bullying from ordinary peer conflict. Randomly assigned classrooms of students either watched or did not watch a video prior to completing a self-report bullying survey. Compared to the control group, students who watched the video reported 32% less social bullying victimization and boys who watched the video reported 54% less physical bullying victimization and 68% less physical bullying of others. These results indicate that student self-reports could yield inflated estimates of the prevalence of bullying if students are not adequately educated about the distinction between bullying and other forms of peer conflict.

Keywords: bullying, educational video, middle school, peer conflict, self-report

Effects of an Educational Video on the Measurement of Bullying by Self-Report

Bullying is a serious problem in schools that can have negative effects on the social, emotional, and academic adjustment of children and adolescents (Swearer, Espelage, Vaillancourt, & Hymel, 2010). Victims of bullying can suffer from low self-esteem (Lopez & Dubois, 2005), anxiety (Troop-Gordon & Ladd, 2005), and other socioemotional adjustment problems (Hawker & Boulton, 2000; Rigby, 2001). They also exhibit higher rates of school avoidance, truancy, and academic difficulties (Boulton, Trueman, & Murray, 2008; Salmon, James, Cassidy & Javaloyes, 2000; Slee, 1994). Consequently, numerous studies have assessed the prevalence of bullying and the impact of intervention efforts (Wang, Ianotti, & Nansel, 2009; Ttofi & Farrington, 2009). Bullying occurs at all grade levels, but middle schools appear to have especially high rates (Goldbaum, Craig, Pepler, Connolly, 2007; Wang et al., 2009). For example, a national survey found that 36% to 43% of middle school students reported being bullied at school during the school year (DeVoe & Bauer, 2010). In this context, it is important to have accurate measures of bullying.

Many authorities have expressed concern that the measurement of bullying is limited by two interrelated problems: a lack of consensus on the complex definition of bullying and the reliance on student self-report (Cornell & Bandyopadhyay, 2010; Furlong, Sharkey, Felix, Tanigawa, & Green, 2010; Swearer, Espelage, Vaillancourt, & Hymel, 2010). Because student self-report is the most commonly used method to assess the prevalence of bullying and the effectiveness of intervention programs (Espelage & Swearer, 2003; Merrell, Gueldner, Ross, & Isava, 2008; Solberg & Olweus, 2003), the accuracy of self-report measures is critical. Self-report measures rely on students'

accurate understanding and application of the definition of bullying; however, bullying is

a complex concept that is difficult to distinguish from other forms of peer conflict

(Cornell & Bandyopadhyay, 2010).

Bullying is generally defined as a form of aggressive behavior that is intentional,

repetitive, and involves an imbalance of power between aggressor and victim (Olweus,

2010). These components are essential to distinguish bullying from other forms of peer

conflict, such as an argument between friends. Although students can easily grasp that

bullying is an intentional and repetitive behavior, it may be more difficult for them to

recognize the requirement for a power imbalance (Espelage & Swearer, 2003; Rigby,

2002). A power imbalance can be readily apparent if the aggressor is much larger or

stronger than the victim, or if the victim is outnumbered by a group, but in some cases the

aggressor's power is reflected in social status or popularity (Rodkin & Hodges, 2003).

Greif and Furlong (2006) point out that a victim can have less power because he or she is

different from others because of race, ethnicity, sexual orientation, physical

characteristics, or noticeable disabilities.

The imbalance of power requirement has the effect of narrowing the definition of

bullying to a subset of peer aggression, excluding fights between students of the same

size and strength, teasing that occurs between peers of equal status, or arguments between

friends where one does not have greater power or status over the other. In contrast, most

student surveys emphasize a broad conception of bullying that includes physical, verbal,

or social forms of bullying. The disparity between this broad, inclusive conception of

bullying and the narrow, exclusive criterion of a power imbalance makes the task of

distinguishing bullying from other forms of peer conflict especially challenging for

students. If students are not attentive to the power imbalance criterion, they could be over-inclusive in their reporting, resulting in inflated rates of bullying.

One strategy used in many surveys is to provide students with an explicit definition of bullying. These definitions often emphasize that bullying does not include conflicts between students of the same strength or power (Cornell, 2010; Nansel et al., 2001; Olweus, 1996). Vaillancourt et al. (2008) demonstrated that students were less likely to report being bullied when they were provided with a definition that distinguished bullying from other forms of peer conflict. Vaillancourt et al. (2008) also asked students to give their own definitions of bullying. Almost all (92%) of the students included some mention of negative behavior in their definition, but relatively few included the idea of power imbalance (26%). Another study (Kert, Codding, Tryon, & Shiyko, 2010) compared student responses to a self-report measure that made no reference to the word "bully" and its definition with variations that used the word "bully" and included a definition. As expected, students reported significantly less bullying behavior when provided with a definition and exposed to the word. These studies suggest the tendency of surveys to produce higher rates of bullying if students are not oriented toward the definition of bullying and how it can be distinguished from other forms of peer conflict.

A recent study examined the accuracy of student self-reports by administering a survey on a confidential, but not anonymous, basis so that students who identified themselves as victims of bullying could be interviewed (Cornell & Mehta, 2011). The survey presented students with a standard definition of bullying that clearly distinguished bullying from other forms of peer conflict and asked students how frequently they had

been bullied in the past 30 days. In this study, 43 middle school students reported they had been bullied one or more times weekly, but when school counselors interviewed these students, only 24 could be confirmed as victims of bullying. The remaining 19 students reported peer conflicts that did not involve a power imbalance (13 students), claimed to have marked the survey in error (four students), or reported bullying that had ceased more than 30 days ago. This study suggested that a definition of bullying alone was not sufficient to guide students in distinguishing between bullying and ordinary peer conflict. It may be necessary to develop additional methods of teaching students the complex concept of bullying.

Study Purpose

This study investigated whether an educational video might improve students' understanding of bullying and help them distinguish bullying from other forms of peer conflict. Videos may be an effective method of educating students about the definition of bullying and teaching them about the relatively abstract concept of power imbalance. Videos can offer concrete, real life examples of different forms of bullying in a context that may be easier for students to grasp and apply to their own experiences. More generally, video-based instruction is an effective way of improving motivation and attention (Choi & Johnson, 2005).

There is widespread use of educational videos as a component of antibullying programs (e.g., Baldry & Farrington, 2004; Boulton & Flemington, 1996; Olweus, 1991; 1993; Sharp & Smith, 1991). A recent review of literature found that use of videos was one of the few distinguishing features of effective programs (Ttofi & Farrington, 2009), a finding that is likely to spur their increased use. Nevertheless, our review of the literature

found no studies that examined the effects of an educational video on student self-report of bullying.

The purpose of the present study was to examine the impact of an educational video, *Bullying or not* (Cornell, 2008), that was designed to help students distinguish verbal, social, and physical forms of bullying from ordinary peer conflict among students of similar power. It was expected that students who watched the video would have a narrower and more specific understanding of bullying and therefore would be less likely than the control group to: (a) report being bullied and (b) admit bullying others. It is well established that bullying rates differ across grade and gender. For example, a national survey found that boys reported more involvement in physical and verbal bullying and girls reported more social bullying, and that sixth-grade students were more likely to be victimized than seventh- and eighth-grade students (Wang et al., 2009). Therefore, grade and gender were included as analytic variables.

Method

Participants

The sample included 661 boys (52%) and 606 girls (48%) from 3 middle schools (Grades 6, 7, and 8) in a public school district in central Virginia. There were 403 (32%) sixth graders, 450 (35%) seventh graders, and 415 (33%) eighth graders. The participants ranged from 11 to 15 years. There were 861 (67%) students who identified themselves as White, 161 (13%) as African American, 82 (7%) as Hispanic, 71 (6%) as Asian American, and 88 (7%) as another ethnicity.

Measures

School Climate Bullying Survey. The School Climate Bullying Survey (SCBS;

Cornell, 2010) is a self-report instrument used to assess the prevalence of bullying and measure related aspects of school climate. Validation studies demonstrated that the SCBS items used as self-reports of bullying corresponded with independent measures obtained from peer nominations and teacher nominations (Branson & Cornell, 2009; Cornell & Brockenbrough, 2004). In further support of its criterion-related validity, self-reports of victimization were also correlated with depression, negative perceptions of school, and lower academic performance, whereas self-reports of bullying others were correlated with aggressive attitudes, discipline referrals, and suspensions from school (Branson & Cornell, 2009). The SCBS also yields estimates of bullying comparable to the widely used Olweus Bullying Victimization Questionnaire (BVQ; Olweus, 1996). When a sample of 388 students were randomly assigned to complete either the SCBS or the BVQ, the two instruments produced very similar estimates of the prevalence of bullying others at least once a week (1.5% for the SCBS versus 1.0% for the Olweus BVQ) and being a victim of bullying at least once a week (3.6% for the SCBS versus 4.2% for the Olweus BVQ).

The middle school version of the SCBS presented students with the following definition of bullying: "Bullying is defined as the use of one's strength or popularity to injure, threaten, or embarrass another person. Bullying can be physical, verbal, or social. It is *not bullying* when two students of about the same strength argue or fight." Next, students responded to two sets of questions, one asking whether they have been bullied and the other asking whether they had bullied others. Response options were *never*, *once or twice*, *about once per week*, or *several times per week*. Students who marked either *about once per week* or *several times per week* were categorized as involved in bullying,

similar to the categorization recommended by Solberg and Olweus (2003).

This study is concerned with the survey questions about physical, verbal, and social bullying. The survey defined physical bullying as, "repeatedly hitting, kicking, or shoving someone weaker on purpose" and then presented two items: "I have been physically bullied or threatened with physical bullying" and "I have physically bullied or threatened to physically bully another student." Verbal bullying was defined as, "repeatedly teasing, putting down, or insulting someone on purpose." Items that measured verbal bullying were, "I have been verbally bullied" and "I have verbally bullied another student." Social bullying was defined as, "getting others repeatedly to ignore or leave someone out on purpose." Items that measured social bullying were, "I have been socially bullied" and "I have socially bullied another student."

Procedure

All three participating middle schools had implemented the Olweus Bullying Prevention Program (OBPP; Olweus, 1994) for at least one year. The OBPP aims to improve peer relations and promote a safe and positive school climate by fostering schoolwide awareness of bullying (Olweus, 1994). At least one staff member from each school became trained and certified in the OBPP. These staff helped create bullying prevention coordinating committees that developed schoolwide rules against bullying. The program included regular classroom discussions and other activities designed to engage students and raise awareness of bullying. The SCBS was administered each fall and spring to assess progress in reducing bullying.

Educational bullying video. Although students had been exposed to the concept of bullying through the Olweus Bullying Prevention Program, and used a survey that

included a standard definition of bullying, school authorities were still concerned that students were not using the term correctly. For example, counselors in one school conducted follow-up interviews with students who reported that they were victims of bullying and found that many of these students confused bullying with peer conflicts that did not involve a power imbalance (Cornell & Mehta, 2011). Therefore, school authorities decided to use a short educational video to remind students about the definition of bullying before completing the survey. This video was made by students in the school system as part of a school project, in collaboration with one of the authors (Cornell) of this study.

The 6.3 minute video was entitled *Bullying or Not?* (available online at www.youtube.com) because it was designed to help students distinguish bullying from other forms of peer conflict. In the opening scene of the video, two student commentators (boy and girl) reviewed the definition of bullying, emphasizing the power imbalance concept. Next, three pairs of scenes illustrated the difference between bullying and ordinary peer conflict that is not bullying. In each pair, the first scene demonstrated a clear instance of bullying and in the companion scene the same actors enacted a similar peer conflict that was not bullying. For example, two scenes illustrated the difference between verbal bullying and a verbal argument between two peers of comparable size and status. Similarly, two scenes distinguished social bullying from an argument between friends and two scenes distinguished physical bullying from a physical struggle between two boys of comparable size and strength. The student commentators explained the power imbalance present in each of the bullying scenes. At the end of the video, the student commentators emphasized the importance of preventing bullying and encouraged

students to answer survey questions correctly when asked about bullying.

     Survey administration. The school system routinely administered the survey as part of its bullying prevention program (not as a research study). In order to assess the impact of the video, the school administration determined that classes from three schools would be randomly assigned (by coin flip) to either a control (no video) or experimental (video) condition. The assignment of each condition within each school was stratified by grade level so that roughly half the classrooms within each grade were shown the video and half were not.

     Surveys were conducted in 73 homeroom classes during the same designated time period. Administrators were classroom teachers who introduced the survey according to a standard set of instructions that included reviewing the definition of bullying contained on the survey. As a result, students in both conditions were presented with the survey definition of bullying orally and in writing, but only classes assigned to the experimental condition were shown the educational video immediately before filling out the survey. Teachers were aware that their classroom had been selected to show the video, and could recognize that the video taught students about the distinction between bullying and other forms of peer conflict, but were not advised of any hypotheses about expected effects.

Results

     Students were identified on the survey as watching or not watching the video, but were not identified by classroom. Therefore, video effects were analyzed at the individual student level and not the classroom level. A 2 x 2 x 3 analysis of variance (ANOVA) measured the effects of video condition (watched or not watched), gender (boy or girl), and grade level (6th, 7th, and 8th for middle school) on student reports of

involvement in bullying. Separate analyses were conducted on reports of being bullied and bullying others.

Across all three middle schools, 3.6% of students reported being physically bullied, 13.1% reported being verbally bullied, and 6.1% reported being socially bullied (see Table 1). When asked about bullying others, 2.6% reported physically bullying, 3.0% reported verbally bullying, and 2.1% reported socially bullying others (see Table 2).

Reports of being bullied. For reports of being physically bullied, there were statistically significant main effects for gender, $F(1, 1247) = 16.83$ $p < .001$, $\eta^2 = .013$, and grade level, $F(2, 1247) = 6.34$, $p = .002$, $\eta^2 = .010$, but not video condition (see Table 3). However, reports of being physically bullied were found to be significantly influenced by the interaction between video condition and gender, $F(1, 1247) = 4.24$, $p = .040$, $\eta^2 = .003$. Post-hoc simple effect tests indicated that boys who did not watch the video reported being physically bullied significantly more than girls who did not watch the video, $F(1, 609) = 16.65$, $p < .001$, $\eta^2 = .027$. However, there was no significant difference between boys and girls who watched the video. About 7% of boys and 1.3% of girls who did not watch the video reported being physically bullied, compared to 3.2% of boys and 2.6% of girls in the video condition.

Analysis of student reports of being verbally bullied revealed statistically significant results for grade level, $F(2, 1246) = 5.33$, $p = .005$, $\eta^2 = .008$, but not gender or video. There were no significant interactions.

Students who watched the video reported being socially bullied less than students who did not watch the video, $F(1, 1247) = 4.99$, $p = .026$, $\eta^2 = .004$. Approximately 7.3% of students who did not watch the video reported that they had been socially

bullied, compared to 5.0% of students who watched the video. Girls reported being bullied more than boys, F (1, 1247) = 7.56, p = .006, η^2 = .006. However, there were no significant interactions between any factors.

Reports of bullying others. For student reports of physically bullying others, there were significant main effects for gender, F (1, 1247) = 8.08, p = .005, η^2 = .006, but not for video or grade level (see Table 4). The interaction between video and gender was also statistically significant, F (1, 1247) = 4.74, p = .030, η^2 = .004. Post-hoc simple effect tests indicated that, among those who did not watch the video, boys reported bullying others more than girls, F (1, 608) = 8.90, p = .003, η^2 = .014. However, there was no significant difference between boys and girls who watched the video. Boys who watched the video reported that they bullied significantly less than boys who did not watch the video, F (1, 657) = 6.49, p = .011, η^2 = .010. Only 1.8% of boys who watched the video reported physically bullying others compared with 5.7% of boys who did not watch the video. Girls who did or did not watch the video did not differ.

Analysis of student reports of verbally bullying others revealed no significant main effects. However, there was a statistically significant interaction among video condition, gender, and grade level, F (2, 1247) = 3.15, p = .043, η^2 = .005. Post-hoc simple effect tests indicated that eighth-grade boys who saw the video reported bullying others significantly less than eighth-grade boys who did not see the video, F (1, 228) = 7.17, p = .008, η^2 = .030. However, there was no comparable significant difference for eighth-grade girls. Analysis of student reports of socially bullying others revealed no significant main effects and no significant interactions among video condition, gender, or grade level.

Discussion

Two key findings of this study are that students who watched the video reported significantly less social bullying victimization than students who did not watch the video and that boys who watched the video reported less physical bullying – as both aggressor and victim – than boys who did not watch the video. One possible explanation for the effect found with social bullying is that this form of bullying is especially difficult to distinguish from the broader range of ordinary peer conflict and the video helped students make a more narrow distinction. With regard to physical bullying, boys typically report more of this form of bullying than girls, and as a result they may have been more affected by the video than were girls. Also, since the video depiction of physical bullying focused on two boys playing basketball, this segment may have been more salient for boys. Other explanations for these effects are possible, which might be determined by comparing videos with different examples of bullying using students of varied age, gender, and race.

Together, these results indicate that student surveys could yield inflated estimates of the prevalence of bullying if students are not adequately educated about the distinction between bullying and other forms of peer conflict. The effects were not found across all measures of bullying, but it is noteworthy that *any* effects were found since the schools had an ongoing bullying prevention program and used a survey that presented students with a standard definition of bullying. It would be useful to test for effects in a sample with less exposure to the definition of bullying as well as to consider the impact of videos on surveys that do not define bullying.

The effect sizes as measured by partial eta-squared values for these results were small, but it is difficult to judge practical significance using this metric. A more concrete

and practical indication of effect size is to consider the difference in reported victim rates

under the two conditions. The rate of student reports of social bullying dropped from

7.3% among students not watching the video to 5.0% among students in the video

condition. If these percentages are applied to the present sample of 1,283 students, the

use of the video would reduce the estimate of the prevalence of social bullying from

approximately 94 victims to 64 victims, a drop of 30 victims. Another way to consider

this effect is that it represents a reduction of 32% in the estimated rate of social bullying

victimization. This reduction is a larger effect than the reduction of 20-23% found in

most controlled bullying prevention studies (Ttofi & Farrington, 2009). Bullying

researchers and educators should be concerned about measuring a treatment effect using a

method — student self-report — that is so susceptible to change based on how students

are reminded about the concept of bullying.

     There were few significant video effects for student reports of bullying others.

One possible explanation for this result is that overall reports of bullying others were so

low that any effect of the video was attenuated. Another explanation may be that reports

of bullying others are largely influenced by student willingness to acknowledge and

admit engaging in a prohibited behavior with negative social connotations, so that

education regarding the concept of bullying had little effect.

     A number of bullying prevention programs use videos (Baldry & Farrington,

2004; Vreeman & Carroll, 2007), but there is little research on their effects. For

example, the Olweus Bullying Prevention Program (OBPP) includes a video comprised

of bullying scenarios that instructors can use to enhance lessons and promote classroom

discussion; however, we found no report of the impact of this video on the program's

effects. Boulton and Flemington (1996) examined the effects of an antibullying video on student attitudes toward bullying and their tendency to bully others over a two-week period. Watching this video, along with other elements of the program, did not decrease student reports of bullying others, nor did it cause students to have more negative attitudes towards bullying in general when they were tested two weeks later.

Study Limitations and Directions for Further Research

These results should be replicated in other samples across a wider range of students and in schools that do not have bullying prevention programs. The effect of an educational video should be compared to other ways of educating students about the definition of bullying. More generally, the video itself requires further study to understand its effects. As a brief video, no attempt was made to give multiple examples of each form of bullying or to systematically use both boys and girls across each form. Systematic study is needed to determine what elements of the video had an effect and whether a stronger effect could be obtained with a longer video or one with different messages about bullying. Since this video was acted, recorded, and edited by relatively inexperienced students, it may be useful to determine whether a more professional video could produce a stronger effect, or whether the apparent amateur qualities of the video made it more compelling for students who could relate to a product made by peers.

Another limitation is that the school survey procedure only identified whether students watched the video and not which classroom they attended. This meant that the data could not be analyzed for both individual and classroom effects using hierarchical linear modeling. It would also be informative to include measures of classroom adherence to the bullying prevention model and baseline measures of bullying levels to

examine their interaction with the effect of the video.

Finally, future research should explore the extent to which improved student understanding can lead to more accurate measurements of bullying prevalence. For example, one approach would be to test student understanding of bullying before and after watching the video by asking them to correctly identify bullying when presented with various scenarios or role-play situations.

Conclusion

In conclusion, this study demonstrated that student self-reports of bullying victimization appear to be somewhat malleable and susceptible to influence by an educational video. These results underscore the fragility of student self-report as a means of assessing the prevalence of bullying and the effects of intervention efforts. Schools should make a concerted effort to educate their students about the concept of bullying and make sure that all students are surveyed about bullying using standard procedures. Differences in how the students were taught and what lessons on bullying precede the survey might affect the prevalence rates obtained in a survey. Future studies should explore the best methods for educating students in order to have more assurance that they can distinguish bullying from other forms of peer conflict. As many others have noted (Cornell & Bandyopadhyay, 2010; Furlong et al., 2010; Swearer, Espelage, Vaillancourt, & Hymel, 2010), research on the validity and especially the accuracy of student self-reports is needed and comprehensive assessments of bullying should include multiple sources of information rather than rely solely on self-report.

References

Baldry, A. C., & Farrington, D. P. (2004). Evaluation of an intervention program for the

      reduction of bullying and victimization in schools. *Aggressive Behavior, 30*, 1–15.

      doi:10.1002/ab.20000

Boulton, M. J., & Flemington, I. (1996). The effects of a short video intervention on

      secondary school pupils' involvement in definitions of and attitudes towards

      bullying. *School Psychology International, 17*, 331–345.

      doi:10.1177/0143034396174003

Boulton, M. J., Trueman, M., & Murray, L. (2008). Associations between peer

      victimization, fear of future victimization and disrupted concentration on class

      work among junior school pupils. *British Journal of Educational Psychology, 78*,

      473–489. doi:10.1348/000709908X320471

Branson, C. E., & Cornell, D. G. (2009). A comparison of self and peer reports in the

      assessment of middle school bullying. *Journal of Applied School Psychology, 25*,

      5–27. doi:10.1080/15377900802484133

Choi, H. J., & Johnson, S. D. (2005). The effect of context-based video instruction on

      learning and motivation in online courses. *American Journal of Distance*

      Education, 19, 215–227. doi:10.1207/s15389286ajde1904_3

Cornell, D. G. (Producer). (2008) *Bullying or not* [Motion picture]. (Available from The

      Virginia Youth Violence Project, Ruffner Hall, 405 Emmet Street, Charlottesville,

      VA 22904-4261).

Cornell, D. G. (2010). *The School Climate Bullying Survey: Description and research*

      summary. Charlottesville, VA: University of Virginia.

Cornell, D. G., & Bandyopadhyay, S. (2010). The assessment of bullying. In S. R.

      Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in*

      schools: An international perspective. (pp. 265–276). New York, NY: Routledge.

Cornell, D. G., & Brockenbrough, K. (2004). Identification of bullies and victims: A

      comparison of methods. *Journal of School Violence, 3*, 63–87.

      doi:10.1300/J202v03n02_05

Cornell, D., & Mehta, S. (2011). Counselor confirmation of middle school student self-

      reports of bullying victimization. *Professional School Counseling*, *14*, 261-270.

DeVoe, J. F., & Bauer, L. (2010). *Student victimization in U.S. schools: Results from the*

      2007 School Crime Supplement to the National Crime Victimization Survey

      (NCES 2010-319). National Center for Education Statistics, Institute of Education

      Sciences, U.S. Department of Education. Washington, DC.

Espelage, D. L., & Swearer, S. M. (2003). Research on school bullying and victimization:

      What have we learned and where do we go from here? *School Psychology*

      Review, 32, 365–383.

Furlong, M. J., Sharkey, J. D., Felix, E. D., Tanigawa, D., & Green, J. G. (2010).

      Bullying assessment: A call for increased precision of self-reporting procedures.

      In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *The handbook of*

      bullying in schools: An international perspective (pp. 329–346). New York, NY:

      Routledge.

Goldbaum, S., Craig, W. M., Pepler, D., & Connolly, J. (2007). Developmental

    trajectories of victimization: Identifying risk and protective factors. In J. E. Zins,

    M. J. Elias, & C. A. Maher (Eds.), *Bullying, victimization, and peer harassment:*

    A handbook of prevention and intervention (pp. 143-160). New York, NY:

    Haworth Press.

Greif, J. L., & Furlong, M. J. (2006). The assessment of school bullying. *Journal of*

    School Violence, 5, 33–50. doi:10.1300/J202v05n03_04

Hawker, D. S. J., & Boulton, M. J. (2000). Twenty years' research on peer victimization

    and psychosocial maladjustment: A meta-analytic review of cross-sectional

    studies. *Journal of Child Psychology and Psychiatry, 41*, 441–455.

    doi:10.1111/1469-7610.00629

Kert, A. S., Codding, R. S., Tryon, G. S., & Shiyko, M. (2010). Impact of the word

    'bully' on the reported rate of bullying behavior. *Psychology in the Schools, 47*,

    193–204. doi:10.1002/pits.20464

Lopez, C., & DuBois, D. L. (2005). Peer victimization and rejection: Investigation of an

    integrative model of effects on emotional, behavioral, and academic adjustment in

    early adolescence. *Journal of Clinical Child & Adolescent Psychology, 34*, 25–36.

    doi:10.1207/s15374424jccp3401_3

Merrell, K. W., Gueldner, B. A., Ross, S. W., & Isava, D. M. (2008). How effective are

    school bullying intervention programs? A meta-analysis of intervention research.

    School Psychology Quarterly, 23, 26–42. doi:10.1037/1045-3830.23.1.26

Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P.

　　　(2001). Bullying behaviors among US youth: Prevalence and association with

　　　psychosocial adjustment. *JAMA: The Journal of the American Medical*

　　　Association, 285, 2094–2100.

Olweus, D. (1991). Bully/victim problems among schoolchildren: Basic facts and effects

　　　of a school based intervention program. In D. J. Pepler & K. H. Rubin (Eds.), *The*

　　　development and treatment of childhood aggression (pp. 411–448). Hillsdale, NJ:

　　　Erlbaum.

Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Malden, MA:

　　　Blackwell.

Olweus, D. (1994). Annotation: Bullying at school: Basic facts and effects of a school

　　　based intervention program. *Journal of Child Psychology and Psychiatry, 35*,

　　　1171–1190. doi:10.1111/j.1469-7610.1994.tb01229.x

Olweus, D. (1996). *The Revised Olweus Bully/Victim Questionnaire.* Mimeo. Bergen,

　　　Norway: Research Center for Health Promotion (HEMIL Center), University of

　　　Bergen.

Olweus, D. (2010). Understanding and researching bullying: Some critical issues. In S. R.

　　　Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in*

　　　schools: An international perspective (pp. 9–33). New York, NY: Routledge.

Rigby, K. (2001). Health consequences of bullying and its prevention in schools. In J.

　　　Juvonen, & S. Graham (Eds.), *Peer harassment in school: The plight of the*

　　　vulnerable and victimized (pp. 310−331). New York, NY: Guilford.

Rigby, K. (2002). *New perspectives on bullying.* London, England: Jessica Kingsley.

Rodkin, P. C., & Hodges, E. V. E. (2003). Bullies and victims in the peer ecology: Four questions for psychologists and school professionals. *School Psychology Review, 32,* 384–400.

Salmon, G., James, A., Cassidy, E. L., & Javaloyes, M. A. (2000). Bullying a review: Presentations to an adolescent psychiatric service and within a school for emotionally and behaviourally disturbed children. *Clinical Child Psychology and Psychiatry, 5*, 563–579. doi:10.1177/1359104500005004010

Sharp, S., & Smith, P. K. (1991). Bullying in UK schools: The DES Sheffield bullying project. *Early Child Development and Care, 77*, 47–55. doi:10.1080/0300443910770104

Slee, P. T. (1994). Situational and interpersonal correlates of anxiety associated with peer victimization. *Child Psychiatry and Human Development, 25*, 97–107. doi:10.1007/BF02253289

Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim questionnaire. *Aggressive Behavior, 29*, 239–268. doi:10.1002/ab.10047

Swearer, S. M., Espelage, D. L., Vaillancourt, T., & Hymel, S. (2010). What can be done about school bullying?: Linking research to educational practice. *Educational Researcher, 39*, 38–47. doi:10.3102/0013189X09357622

Ttofi, M., & Farrington, D. (2009). What works in preventing bullying: Effective elements of anti-bullying programmes. *Journal of Aggression, Conflict and Peace Research, 1*, 13–24.

Troop-gordon, W., & Ladd, G. W. (2005). Trajectories of peer victimization and

      perceptions of the self and schoolmates: Precursors to internalizing and

      externalizing problems. *Child Development*, *76*, 1072–1091. doi:10.1111/j.1467-

      8624.2005.00898.x

Vaillancourt, T., McDougall, P., Hymel, S., Krygsman, A., Miller, J., Stiver, K., et al.

      (2008). Bullying: Are researchers and children/youth talking about the same

      thing? *International Journal of Behavioral Development, 32*, 486–495.

      doi:10.1177/0165025408095553

Vreeman, R. C., & Carroll, A. E. (2007). A systematic review of school-based

      interventions to prevent bullying. *Archives of Pediatrics & Adolescent Medicine,*

      161, 78–88. doi:10.1001/archpedi.161.1.78

Wang, J., Iannotti, R. J., & Nansel, T.R. (2009). School bullying among adolescents in

      the United States: Physical, verbal, relational, and cyber. *Journal of Adolescent*

      Health, 45, 368–375. doi:10.1016/j.jadohealth.2009.03.021

Table 1
Reports of Being Physically, Verbally, and Socially Bullied (one or more times per week)

Video	Gender	Grade	Being Physically Bullied				Being Verbally Bullied				Being Socially Bullied			
			Mean	*SD*	*N*	%	*Mean*	*SD*	*N*	%	*Mean*	*SD*	*N*	%
Watched	Boy	6	1.35	.66	100	6.0	1.60	.94	100	16.0	1.24	.49	101	3.0
		7	1.25	.57	117	3.4	1.49	.87	118	14.4	1.28	.68	118	7.6
		8	1.16	.39	125	0.8	1.34	.73	125	7.2	1.17	.50	125	2.4
		Total	1.25	.55	342	3.2	1.47	.85	343	12.2	1.23	.57	344	4.4
	Girl	6	1.26	.57	105	2.9	1.75	.89	105	16.2	1.31	.64	105	5.7
		7	1.17	.51	109	1.8	1.58	.87	109	12.8	1.30	.71	109	5.5
		8	1.14	.43	94	3.2	1.47	.70	93	9.7	1.31	.72	94	6.4
		Total	1.19	.51	308	2.6	1.61	.84	307	13.0	1.31	.69	308	5.8
	Total	6	1.30	.62	205	4.3	1.68	.92	205	16.1	1.28	.57	206	4.4
		7	1.21	.54	226	2.7	1.53	.87	227	14.0	1.29	.69	227	6.6
		8	1.15	.41	219	1.8	1.40	.72	218	8.3	1.23	.61	219	4.1
		Total	1.22	.53	650	2.9	1.53	.84	650	12.6	1.27	.63	652	5.0
Not Watched	Boy	6	1.36	.69	101	7.9	1.61	.99	101	18.8	1.38	.71	101	8.9
		7	1.37	.75	107	8.4	1.59	.92	107	13.1	1.22	.60	106	3.7
		8	1.20	.61	106	4.7	1.48	.82	105	11.5	1.27	.63	106	7.5
		Total	1.31	.69	314	6.9	1.56	.91	313	14.3	1.29	.65	313	6.7
	Girl	6	1.15	.44	95	1.1	1.56	.85	95	10.5	1.38	.77	95	7.4
		7	1.13	.39	112	1.8	1.71	.83	112	17.0	1.48	.83	111	10.8
		8	1.08	.31	88	1.1	1.49	.82	88	9.1	1.40	.74	88	5.6
		Total	1.12	.39	295	1.3	1.60	.84	295	12.6	1.42	.78	294	8.2

Table 1 continued

Video	Gender	Grade	Being Physically Bullied				Being Verbally Bullied				Being Socially Bullied			
			Mean	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*
Not Watched	Total	6	1.26	.59	196	4.6	1.59	.92	196	14.8	1.38	.74	196	8.2
		7	1.25	.60	219	5.0	1.65	.88	219	15.0	1.35	.74	217	7.3
		8	1.14	.50	194	3.0	1.48	.82	193	10.9	1.33	.68	194	6.7
		Total	1.22	.57	609	4.2	1.58	.88	608	13.3	1.35	.72	607	7.3
Total	Boy	6	1.35	.67	201	7.0	1.61	.96	201	17.5	1.31	.61	202	6.0
		7	1.31	.66	224	5.8	1.54	.89	225	13.8	1.25	.64	224	5.8
		8	1.18	.50	231	3.0	1.40	.78	230	9.5	1.22	.56	231	4.7
		Total	1.28	.62	656	5.2	1.51	.88	656	13.4	1.26	.61	657	5.5
	Girl	6	1.21	.51	200	2.0	1.66	.87	200	13.5	1.35	.71	200	6.5
		7	1.15	.45	221	1.8	1.65	.85	221	15.0	1.39	.78	220	8.2
		8	1.11	.38	182	2.2	1.48	.76	181	9.4	1.35	.73	182	6.0
		Total	1.16	.45	603	2.0	1.60	.83	602	12.8	1.36	.74	602	6.9
	Total	6	1.28	.60	401	4.4	1.63	.92	401	15.5	1.33	.66	402	6.2
		7	1.23	.57	445	3.8	1.59	.87	446	14.5	1.32	.72	444	6.9
		8	1.15	.45	413	2.6	1.44	.77	411	9.7	1.28	.64	413	5.4
		Total	1.22	.55	1259	3.6	1.55	.86	1258	13.1	1.31	.67	1259	6.1

Table 2
Reports of Physically, Verbally, and Socially Bullying Others (One or more times per week)

Video	Gender	Grade	Physically Bullying Others				Verbally Bullying Others				Socially Bullying Others			
			Mean	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*
Watched	Boy	6	1.11	.37	100	2.0	1.28	.47	100	1.0	1.18	.39	101	0.0
		7	1.14	.46	118	2.5	1.31	.58	118	4.2	1.22	.53	118	3.3
		8	1.12	.35	125	0.8	1.21	.43	125	0.8	1.11	.41	125	3.2
		Total	1.13	.40	343	1.8	1.27	.50	343	2.0	1.17	.45	344	2.3
	Girl	6	1.04	.19	105	0.0	1.18	.39	105	0.0	1.16	.44	105	1.0
		7	1.16	.43	109	2.8	1.28	.49	109	1.8	1.15	.38	109	0.9
		8	1.13	.45	94	2.2	1.26	.51	94	3.2	1.18	.39	94	0.0
		Total	1.11	.38	308	1.6	1.24	.46	308	1.6	1.16	.40	308	0.6
	Total	6	1.07	.30	205	1.0	1.23	.43	205	0.5	1.17	.41	206	0.5
		7	1.15	.45	227	2.6	1.30	.54	227	3.5	1.19	.46	227	2.6
		8	1.12	.39	219	1.4	1.23	.46	219	1.8	1.14	.40	219	1.8
		Total	1.12	.39	651	1.7	1.25	.48	651	2.0	1.17	.43	652	1.7
Not Watched	Boy	6	1.26	.64	101	5.0	1.23	.53	101	3.0	1.19	.44	101	2.0
		7	1.17	.50	107	3.7	1.28	.58	107	2.8	1.11	.32	106	0.0
		8	1.27	.78	106	8.5	1.43	.80	105	9.6	1.23	.64	106	3.8
		Total	1.23	.65	314	5.7	1.31	.65	313	5.0	1.18	.49	313	1.9
	Girl	6	1.08	.38	95	1.1	1.26	.55	95	3.2	1.24	.68	95	5.3
		7	1.12	.40	111	0.9	1.34	.65	112	4.5	1.21	.43	112	0.9
		8	1.10	.34	88	1.1	1.26	.44	88	0.0	1.19	.54	88	4.5
		Total	1.10	.37	294	1.0	1.29	.56	295	2.8	1.21	.55	295	2.4

Table 2 continued

Video	Gender	Grade	Physically Bullying Others				Verbally Bullying Others				Socially Bullying Others			
			Mean	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*
Not Watched	Total	6	1.17	.54	196	3.0	1.24	.54	196	3.0	1.21	.57	196	3.5
		7	1.14	.45	218	2.3	1.31	.62	219	3.7	1.16	.38	218	0.5
		8	1.20	.62	194	5.2	1.35	.66	193	5.2	1.21	.60	194	4.1
		Total	1.17	.54	608	3.3	1.30	.61	608	4.0	1.19	.52	608	2.5
Total	Boy	6	1.18	.53	201	3.5	1.25	.50	201	2.0	1.18	.41	202	1.0
		7	1.16	.48	225	3.1	1.30	.58	225	3.5	1.17	.44	224	1.7
		8	1.19	.59	231	4.7	1.31	.63	230	4.8	1.16	.53	231	3.4
		Total	1.18	.54	657	3.8	1.29	.58	656	3.5	1.17	.47	657	2.2
	Girl	6	1.06	.30	200	0.5	1.22	.47	200	1.5	1.20	.57	200	3.0
		7	1.14	.42	220	1.9	1.31	.58	221	3.2	1.18	.41	221	0.9
		8	1.12	.40	182	1.6	1.26	.48	182	1.6	1.19	.47	182	2.1
		Total	1.10	.38	602	1.3	1.26	.51	603	2.2	1.19	.48	603	2.0
	Total	6	1.12	.43	401	2.0	1.24	.49	401	1.7	1.19	.50	402	1.9
		7	1.15	.45	445	2.5	1.30	.58	446	3.5	1.17	.42	445	1.5
		8	1.16	.51	413	3.3	1.29	.57	412	3.4	1.17	.50	413	2.9
		Total	1.14	.47	1259	2.6	1.28	.55	1259	3.0	1.18	.47	1260	2.1

Table 3

Analysis of Variance for Reports of Being Physically, Verbally, and Socially Bullied

Source	Reports of Being Physically Bullied				Reports of Being Verbally Bullied				Reports of Being Socially Bullied			
	df	*F*	η^2	*p*	*df*	*F*	η^2	*p*	*df*	*F*	η^2	*p*
Video (V)	1	0.04	.000	.835	1	0.48	.000	.490	1	4.99	.004	.026
Gender (G)	1	16.85	.013	.000	1	2.40	.002	.122	1	7.56	.006	.006
Grade (Gr)	2	6.37	.010	.002	2	5.33	.008	.005	2	0.40	.001	.668
V x G	1	4.24	.003	.040	1	0.97	.001	.325	1	0.42	.000	.519
V x Gr	2	0.81	.001	.446	2	1.69	.003	.185	2	0.15	.000	.860
G x Gr	2	0.82	.001	.440	2	0.12	.000	.884	2	0.72	.001	.485
V x G x Gr	2	0.12	.000	.835	2	0.57	.001	.566	2	1.63	.003	.197

Table 4

Analysis of Variance for Reports of Physically, Verbally, and Socially Bullying Others

Source	Reports of Physically Bullying Others				Reports of Verbally Bullying Others				Reports of Socially Bullying Others			
	df	*F*	η^2	*p*	*df*	*F*	η^2	*p*	*df*	*F*	η^2	*p*
Video (V)	1	0.82	.003	.052	1	2.41	.002	.121	1	1.10	.001	.295
Gender (G)	1	1.74	.006	.005	1	0.77	.001	.379	1	0.33	.000	.567
Grade (Gr)	2	0.12	.001	.573	2	1.59	.003	.205	2	0.22	.000	.804
V x G	1	1.02	.004	.030	1	0.01	.000	.926	1	0.70	.001	.405
V x Gr	2	0.30	.002	.246	2	1.11	.002	.331	2	1.03	.002	.359
G x Gr	2	0.29	.002	.266	2	0.45	.001	.639	2	0.01	.000	.987
V x G x Gr	2	0.09	.001	.654	2	3.15	.005	.043	2	2.16	.003	.115

Invalid Survey Response Patterns among Middle School Students

Dewey Cornell

Peter Lovegrove

Michael Baly

Curry School of Education, University of Virginia

Author Note

Abstract

Student surveys are widely used to assess student risk behavior, bullying, and school climate in middle schools; however, because such surveys are usually conducted on an anonymous basis, little is known about the validity of student reports using external, independent criteria. This longitudinal study examined the response patterns of 382 middle school students who completed confidential (not anonymous) self-report surveys each fall and spring for three years of middle school (grades 6-8). Approximately 10% of students in each wave indicated on validity screening questions that they were either not telling the truth or paying attention (termed "invalid responders"). A repeated measures latent class analysis found that students could be classified into a large group (64%) that were never flagged by the validity questions and a smaller group (36%) that occasionally reported not telling the truth or not paying attention. Hierarchical linear modeling analyses found that invalid responding to validity questions was associated with higher self-reported rates of risk behavior and more negative perceptions of school climate. Based on independent criteria from school records, invalid responding students were more likely to be referred for disciplinary infractions than other students. This study provides new information about student survey validity and appears to be the first to identify characteristics of students who generate invalid response patterns.

Keywords: student self-report, survey validity screening, risk behavior, bullying, school climate

Invalid Survey Response Patterns among Middle School Students

Adolescent self-report surveys are widely used to assess the prevalence rates for bullying, drug use, fighting, and other forms of risk behavior, as well as to assess the effectiveness of school-based prevention and intervention programs (Johnston, O'Malley, Bachman, & Schulenberg, 2010; Robers, Zhang, & Truman, 2010; Sharkey, Furlong, & Yetter, 2006). Although there have been studies of the potential for under-reporting of drug use and delinquent behavior (e.g., Hindelang, Hirschi, & Weiss, 1979), there has been little attention to the problem of over-reporting (Fan et al., 2006; Sharkey et. al, 2006). Immature and rebellious adolescents may be tempted to offer inflated reports of risky behaviors or they may not take a survey seriously and mark it haphazardly, producing an elevation in otherwise low base rate behaviors. A recent study by Cornell, Klein, Konold, and Huang (2012) found that validity screening questions could identify students who gave seemingly inflated reports of their risky behavior and negative views of their school climate. The purpose of the present study was to provide new information on the survey response patterns and school adjustment of students identified as invalid responders.

There are several documented examples of survey findings distorted by adolescents who gave invalid responses. For example, Miller, Fan, Christensen, Grotevant, and van Dulmen (2000) published a study claiming that adopted adolescents had higher rates of drug use, fighting, and other risky behavior in comparison to non-adoptees on the Add Health survey. However, subsequent in-home interviews found that about 19% of these adolescents falsely claimed to be adopted (Fan et al., 2002). When the data were reanalyzed, the group differences diminished or disappeared. Fan and

colleagues (2006) later identified other evidence of adolescent over-reporting on the Add Health survey, including false claims of having been born outside the United States and having a permanent physical disability.

The uncritical acceptance of findings from adolescent surveys can have widespread impact (Cornell, 2006). An egregious example is based on the 1987 National Adolescent Student Health Survey (American School Health Association, 1989), which asked 8th and 10th grade students whether they had brought a handgun to school. Because approximately 2.6% of boys reported that they brought a gun to school, survey results were extrapolated to the sensational conclusion that 135,000 guns are brought to school every day in the United States. This alarming statistic was reported first in U.S. News & World Report (Witkin, 1991) and found its way into many other news reports and policy statements by professional organizations (Cornell, 2006) such as the National School Boards Association (1993) and American Sociological Association (Levine & Rosich, 1996). A 2012 Google search of "135,000 guns" generated about 1400 results, including reports by the National Crime Prevention Council and the New York Times.

Use of Validity Screening Items

One simple strategy for validity screening is to ask students whether they are telling the truth on the survey. However, the interpretation of this question is not simple. When asked a question such as "I am telling the truth on this survey" the two most straightforward outcomes are that the student is telling the truth on the survey and answers the question affirmatively, or that the student is not telling the truth on the survey and admits this dishonesty by answering the question negatively. However, even students who admit dishonesty may be answering some or even most of the questions honestly.

Two additional possibilities are that a student not telling the truth on the survey might falsely report that he or she is telling the truth and that a student who is telling the truth on other survey questions might falsely claim that he or she is not telling the truth. Although it may not be possible to accurately determine whether or not a student is telling the truth, a validity question is nonetheless valuable if it can detect a sufficient number of dishonest students that the quality of the survey data is improved. In this article, the terms "invalid responder" and "valid responder" are used as shorthand for students who responded negatively or affirmatively, respectively, to validity screening questions (i.e. not telling the truth or telling the truth).

In a survey of 10,909 middle and high school students, Cornell and Loper (1998) employed two validity questions: "I am reading this survey carefully" and "I am telling the truth on this survey." They found that approximately 8% of students gave a negative response to one or both of these items. There was no attempt in this study to demonstrate that the data from these presumably invalid responders was inaccurate beyond the presence of suspiciously high rates of typically low base rate behaviors. The students classified as invalid responders tended to endorse fighting, carrying a gun, and using drugs at rates three to five times higher than other students.

Rosenblatt and Furlong (1997) used an improbable question ("I took ten field trips in the previous month") and looked for inconsistent responding to a pair of similarly worded questions. Students who failed either validity check reported dramatically higher rates of violence in their schools and seemed biased toward portraying their school in a negative light.

Cornell and colleagues (2012) presented results from two studies of validity

screening. The first study showed that students labeled as invalid responders (because they answered negatively to questions asking whether they were telling the truth and paying attention in answering questions) produced elevated rates of risk behaviors compared to valid responders. The second study compared the perceptions of school climate and safety conditions of valid and invalid responders drawn from a sample of 284 public high schools. Invalid responders gave more negative reports of their school climate and safety conditions than valid responders. Moreover, the responses of valid responders, compared to invalid responders, tended to be more highly correlated with teacher perceptions of the school. This study also found that invalid responders were more likely to be male than female and more likely to be minority students than white students.

Present study

The purpose of this study was to gain new insight into students who mark their surveys in an invalid manner and further examine the value of validity screening described by Cornell and colleagues (2012). More specifically, the study investigated three questions: (1) How frequently and consistently do students respond negatively to validity screening questions? (2) How do student reports of risk behavior and perceptions of their schools vary with their survey validity status? (3) How do students who generate survey responses screened as valid compare to students who generate survey responses screened as invalid in their school adjustment?

To investigate these three questions, a cohort of 382 middle school students who participated in semi-annual school surveys were tracked for three years and their responses were linked to information on their school attendance, disciplinary record, and

academic performance at the end of 8th grade. This unique sample permitted us to ask three previously unexamined questions, beginning with a determination whether students are consistent across surveys in their tendency to respond negatively to validity screening questions. Many schools administer surveys on an annual or semi-annual basis, and it would be useful to know whether there is a small group of students who consistently respond negatively to validity screening questions or whether this group changes over time.

The second question concerned how student reports of risk behavior and perceptions of their schools varied with their answers to validity questions. There are two approaches to this question. One is to compare students with survey responses screened as valid and students with survey responses screened as invalid. A second approach, unique to this sample, is that we conducted within-student comparisons of students who produced both survey responses screened as valid and survey responses screened as invalid surveys during their three years of middle school. If these students report higher rates of risk behavior and more negative perceptions of their school when their screening items are invalid, it would provide additional evidence that the validity items indicate an intentional effort to give exaggerated answers.

The final question concerned differences in school adjustment between valid responders (students who generate survey responses screened as valid) compared to invalid responders (students who generate survey responses screened as invalid) in their school adjustment. If the invalid responders are inclined to be rebellious and less compliant with school authority, it can be hypothesized that they would have poorer

school attendance, more disciplinary problems, and lower academic performance than students who consistently pass validity screening.

Methods

Participants

The School Climate Bullying Survey (SCBS; Cornell, 2011) was administered in the fall and spring semester for five years to students attending a single middle school in a suburban community in central Virginia. The school enrollment ranged from 491 to 522 (mean 502) during these five years. The percentage of students eligible for a free or reduced price meal in the school ranged from 29% to 42% (mean 36%).

The initial sample consisted of 495 students who entered the sixth grade during the first three years of the study period. There were 113 of these students who transferred away from the school before completing the 8th grade, leaving a study sample of 382 students who attended the school for 6th, 7th, and 8th grades. Preliminary analyses compared the study sample with the students lost to transfer. The two groups did not differ in gender or survey response validity (valid/invalid), but non-white students (29%) were more likely to transfer than white students (19%), $\chi^2 (1) = 6.5$, p = .011. We next compared the study sample with the students lost to transfer on bullying victimization and the three school climate scales from their sixth grade fall survey using a series of ANOVAs controlling for race (white/non-white). There were no statistically significant group differences for Bullying Victimization, Prevalence of Bullying and Teasing, or Willingness to Seek Help, but students who transferred reported higher Aggressive Attitudes than students in the study sample, $F (91, 403) = 50.3$, $p = .035$. These analyses suggest that there may be some differences in race and Aggressive Attitudes between our

study sample and students who were lost to the study sample because they transferred

from the middle school before completing the 8th grade.

The final sample of 195 boys and 187 females included 239 (62.6%) students who

identified themselves as White, 59 (15.4%) as Black, 42 (11.0%) as Hispanic, 15 (3.9%)

as Asian, and 27 (7.1%) as Other. Information on the socioeconomic status of individual

students was not available.

Measures

Students completed the SCBS (Cornell, 2011), a 45-item self-report instrument

that collects demographic information, reports of involvement in bullying, and three

school climate scales. The SCBS was selected for use because it contains measures of

bullying and school climate that have generated scores with favorable psychometric

properties in samples with similar demographic characteristics, as elaborated below

(Bandyopadhyay, Cornell, & Konold, 2009; Branson & Cornell, 2009).

To assess bullying, students were presented with a standard definition of bullying,

"Bullying is defined as the use of one's strength or status to injure, threaten, or humiliate

another person. Bullying can be physical, verbal, or social. It is *not* bullying when two

students of about the same strength argue or fight" (Cornell, 2011). Bully victimization

was assessed with the item, "I have been bullied at school in the past month" and

bullying others was assessed with the item, "I have bullied others at school in the past

month". There were four response options for both questions (*never, once or twice, about

once a week,* or *several times per week*). These two items have been found to correspond

with independent measures obtained from peer nominations and teacher nominations

(Branson & Cornell, 2009; Cornell & Brockenbrough, 2004). The self-report of being

bullied was also correlated with depression, negative perceptions of school, and lower academic performance, whereas the self-report of bullying others was correlated with aggressive attitudes, discipline referrals, and suspensions from school (Branson & Cornell, 2009). In a previous study of middle school students, the SCBS produced estimates of the prevalence of bullying victimization and bullying others that were similar to the Olweus Bullying Victimization Questionnaire (Cornell, 2011).

The 20-item school climate portion of the SCBS consisted of a seven-item Aggressive Attitudes scale, a four-item Prevalence of Bullying and Teasing scale, and a nine-item Willingness to Seek Help scale. Each scale included four response options: "*strongly disagree*," "*disagree*," "*agree*," and "*strongly agree*." The Aggressive Attitudes scale included items such as "If you fight a lot, everyone will look up to you" and "Bullying is sometimes fun to do." The Prevalence of Teasing and Bullying scale included items such as "Bullying is a problem at this school" and "Students here often get teased about their clothing or physical appearance." The Willingness to Seek Help scale included sample items such as "There are adults at this school I could turn to if I had a personal problem" and "If I tell a teacher that someone is bullying me, the teacher will do something to help." Items on this scale were reverse coded to align with the direction of the other SCBS items.

Previous exploratory and confirmatory factor analyses in a middle school sample (n = 2,111) supported the factor structure of these three school climate scales (Bandyopadhyay et al., 2009). Further analyses with a sample of 7,318 ninth grade students found that these scales were predictive of teacher reports of bullying and teasing, teacher reports of student help-seeking behaviors, teacher reports of gang-related

violence, and school records of suspensions and expulsions (Bandyopadhyay et al.,

2009). Another study supported the factor structure of these scales in a sample of 3,687

high school students (Klein, Cornell, & Konold, 2012).

For the current sample, measures of internal consistency (Cronbach's alphas) for

scores from the three scales were calculated for each of six waves: Aggressive Attitudes

scale (α ranged from .74 to .89 across six waves), Prevalence of Bullying and Teasing

scale (α ranged from .67 to .77 across six waves), and Willingness to Seek Help scale (α

ranged from .75 to .88 across six waves).

The school climate survey included two of the three screening validity items used

by Cornell et al. (2011): (1) "I am telling the truth on this survey," and (2) "I am not

paying attention to how I answer this survey." These items both had four Likert-type

answer choices ranging from "*Strong disagree*" to "*Strongly agree*." The items were then

dichotomized into those students who either disagreed or agreed. Across the six waves of

data collection, the two items had a concordance of 90 to 94 percent agreement (phi

coefficients ranged from .11 to .47). Students who endorsed either not telling the truth or

not paying attention were classified as "invalid responders" and compared with "valid

responders" who affirmatively answered both validity screening items.

In study years four and five, the SCBS was augmented with eight risk behavior

items derived from the Youth Risk Behavior Surveillance System (Eaton et al., 2010).

Because these items were not administered at the outset of the study, they were

completed only by the second and third cohorts (n = 251) of students. These items are

used nationwide to assess the prevalence of student risk behavior and include questions

about smoking cigarettes, drinking alcohol, using marijuana, carrying a weapon on school

property, and getting into a physical fight on school property. The items also covered not going to school because of feeling unsafe, feeling so sad or hopeless that you stopped doing some usual activities, and seriously considering suicide. The timeframe for the items were the past 30 days with answer choices ranging from either "*0 days*" to "*20-30 days*" or "*0 times*" to "*6 or more times*."

Measures of school adjustment were obtained from school records for 8th grade. Adjustment at the end of 8th grade was used because it represented student status at the conclusion of middle school. School adjustment was measured with school attendance, disciplinary infractions and suspensions, GPA, and standardized test scores. These items were selected because they provide an assessment of school adjustment that was independent of student self-report. Academic achievement was measured using the Virginia Standards of Learning (SOL) exams, a series of state-mandated tests given to all students for designated subjects (Virginia Department of Education, 2010). Each exam contains 35-50 items. The exams are scored on a scale of 0-600 with 400 representing the minimum level of proficiency. (Proficiency levels are based on percentage answered correctly, which vary somewhat across exams and generally increase each year). In middle school, students completed exams in Mathematics, Reading, Science, Social Studies, and Writing. In addition, student grade point averages (GPAs) were calculated by averaging year-end grades in Language Arts, Mathematics, Science, and Social Studies using the traditional American four-point grading scale (4 = A, 3 = B, 2 = C, 1 = D, 0 = F). School records also indicated the number of absences for the year, the number of times a student was referred for a disciplinary violation, and the number of times a student received an out-of-school suspension.

Procedure

The survey was administered in classrooms during regularly scheduled advisory periods. Teachers supervised the administration using a standard set of instructions. For students who were absent on the survey day, an additional make-up session was arranged. The survey was administered in paper and pencil format for the first three years and online during the final two years. A Spanish translation of the survey was available.

An important context for this study is that it was conducted in a school that had an ongoing Olweus Bullying Prevention Program (OBPP; Olweus & Limber, 2010). The OBPP is designed to reduce bullying through interventions at the school-wide, classroom, and individual levels. As part of the program, the school adopted school-wide rules against bullying and teachers reinforced rules and worked to increase student knowledge and empathy regarding bullying. Because the survey was administered as a routine part of the school's bullying prevention program, active parental consent was not required. The school principal sent letters to all parents offering them the option to withhold permission for their children to participate in the survey, but no parents chose to do so. Survey data were sent to the researchers using code numbers rather than names to identify surveys. A single staff member at the school served as the code master who had the key linking code numbers to student names. Data were provided to the researchers in archival form.

Data Analysis Plan

The data for this study were obtained from three complete cohorts of middle school students that entered middle school in three consecutive years. Three years of data (six waves of surveys) were obtained from each cohort over a five-year period. Our

analyses addressed three main research questions: (1) How frequently and consistently do students respond negatively to validity screening questions? (2) How do student reports of risk behavior and perceptions of their schools vary with their survey validity status? (3) How do students who generate survey responses screened as valid compare to students who generate survey responses screened as invalid in their school adjustment? To address the first question, we began with descriptive statistics to determine how many students were identified as invalid responders and whether the same students tended to generate invalid surveys over the six waves of survey administration. Then we constructed a repeated measures latent class analysis (RMLCA) to determine the number of groups with distinct patterns of invalid status over time, and examined whether these groups differed in gender and minority status using a conditional RMLCA. Latent class analysis is a procedure to identify latent, unobservable groups within a population based on responses to an array of observed variables. Latent class analysis is similar to cluster analysis, but identifies classes based on the conditional probabilities of membership for each case. In this study, repeated measures latent class analysis was used to examine the data for complex patterns of invalid responding to surveys over time.

The best number of groups for the latent class model was decided using an array of statistics: the Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), adjusted BIC; Log-likelihood, model entropy, and estimated group membership probabilities (Nylund, Asparouhov, & Muthén, 2007; Collins & Lanza, 2009). See the results section for explanation how these statistics were used in the model enumeration process.

The second question was examined with hierarchical linear modeling (HLM) using a two-level regression for each of the self-report measures of risk behavior and school climate as outcomes. In these analyses, observations from multiple time points were nested within each individual. The two dummy variables for race and gender were entered at the between-level, where measures were time-invariant. There were two series of HLM analyses. The first series of HLM analyses used all 382 students and investigated whether valid survey responding generated lower reports of bullying victimization and risk behavior, and more positive reports of school climate, than invalid survey responding, at each time point within respondent. The second series of HLM analyses was confined to the 136 students who provided an invalid response on at least one survey and investigated whether students reported higher levels of bullying victimization and risk behavior, and more negative views of school climate (Aggressive Attitudes, Prevalence of Bullying and Teasing, and Willingness to Seek Help) when their survey responding was flagged as invalid in comparison to when their survey responding was not flagged as invalid. Again, invalid response status for each time point was estimated within each student. This set of analyses is informative because it would indicate that validity screening was sensitive to changes within students in their attitudes toward completing the survey in an honest and careful manner.

The third question compared the 8th grade school adjustment (achievement test scores, attendance, and disciplinary violations, and school suspensions) of invalid and valid responders. Invalid responders in these analyses were the 136 students who were flagged as producing invalid responses to at least one survey and valid responders were 246 students that were never flagged as producing invalid responses. Invalidity items

from all six time points were used to determine whether or not a student was classified an invalid or valid responder. The analysis was conducted using regressions where validity status (ever invalid versus never invalid) was the primary independent measure. Race and gender were included as control variables because of previous findings that invalid responders were more likely to be male than female and more likely to be minority students than white students (Cornell et al., 2012). The explanatory and control variables were entered simultaneously into each model. Five explanatory models were run in total, each employing a regression technique appropriate to the distribution of the outcome(s). In accordance with the distributions of the measures, absences were analyzed in Model 1 using a negative binomial regression, while suspensions and disciplinary infractions were dummy-coded on an absence/presence basis and analyzed using logistic regression in Models 2 and 3. Model 4's single outcome was GPA and employed Ordinary Least Squares regression. In Model 5, the standardized test scores were run as a set of multiple intercorrelated outcomes in order to reduce the possibility of type-I error due to the high correlations between scores. This final model included separate estimates of the explanatory and control variables for each of the test scores, as well as estimates of the intercorrelations between each of the test scores. To provide a sense of the overall fit to the data of the relations of 'ever invalid' to all the standardized test outcome, a MANOVA analysis was run in SPSS. This model had an N of about 80, due to missing data in the standardized test scores. This model found that the F statistics for Wilks' Lambda, which evaluates the overall fit of a covariate's relation to multiple outcome, had a p-value of .906 (value was .98, $F = .310$ (5 degrees of freedom) indicating poor overall fit. However, due to the many differences between this MANOVA model and the

regression model we conducted in Mplus (for example, the final model used maximum likelihood to incorporate missing data, and used a different estimator, and the N for the two models were very different), we chose not to reject our final model based on the MANOVA fit statistics.

Univariate statistics were calculated using SPSS (version 20), while all other analyses were conducted using Mplus (version 6.11; Muthén & Muthén, 1998-2010). All analyses used a maximum likelihood estimator with robust standard errors (MLR), which meant that all analyses included as many cases as possible, and accounted for non-normality of outcomes. Missing data needed to be accounted for in all the analyses, though estimates were relatively similar when using maximum likelihood and listwise deletion, which is a sign that data were not "missing not at random" (MNAR; Enders, 2011).

One self-reported variable was omitted from all analyses because maximum likelihood models would not converge: the bullying perpetration scale. Transformations of this measure were attempted – cases with relative extremely large values were recoded to smaller values to reduce the skewness of the measure, and the measure was dichotomized to indicate prevalence of any bullying perpetration – but were unsuccessful. Auxiliary variables were also employed in the maximum likelihood estimation. A Bayesian estimator was used, and though a main effects model (comparing invalid to valid responses without the regions of significance) converged, the Bayesian posterior parameter trace plots and autocorrelation plots showed that the model estimates were not trustworthy. SPSS plots of the residuals for a linear model containing the variables that

would be used in the nested models showed an extreme non-normality in the residuals for the bullying perpetration measure.

Results

Across six survey administrations, the percentage of students classified as invalid responders ranged from 8% to 12% (mean 9.9%; see Table 1). In the sample of 382 students, there were 136 students (36%) classified as invalid responders for at least one of the six survey administrations. As indicated in Table 2, most of the students (87 of 136, 64%) who provided an invalid response did so on just one survey, whereas only one student generated an invalid survey on all six waves.

Latent Class Analysis. Across a series of latent class models with the number of specified groups increasing each time, the BIC was lowest for two groups. A log-likelihood statistic, the BIC penalizes model fit for sample size and the number of parameters, and a lower value indicates better model fit (Nylund et al., 2007). The BIC was 1400.29 for two groups, while it was 1438.87, 1433.59, 1467.22, and 1501.00 for models containing one, three, four, and five groups, respectively. Comparing the three-group and two-group models, the three-group model contained a group with a very low proportional membership (2.6% of the sample, or 10 members), and one of the groups, the smallest group, had an average probability of group membership below .80. The two-group had estimated group membership probabilities of above .80 for both groups. The stronger degree of group separation indicated by the group membership probabilities was reflected in the model entropy estimates for each model—entropy is an overall estimate of the posterior probabilities of each individual belonging to a particular class, and higher scores (ranging from 0 to 1) indicate more successful classification of individuals (Clark

& Muthén, 2009), While the three-class model had an entropy of .68, the two-class model entropy was higher, at .76.

The conditional item responses for each class, as shown in Figure 1, reveal that one class of students had a low probability of being flagged as invalid at every time point, while the other class had a higher probability of being flagged as invalid at every time point. Thus, the classes were termed 'higher likelihood' and 'lower likelihood.' The measurement error at each time point for the higher likelihood class—if it were expected that students in the class would be flagged as invalid at each year—was quite high: only 37% of members in the higher likelihood class were actually flagged as invalid at wave 1, while this percentage was 38%, 48%, 40%, 33%, and 49% for waves 2 through 6, respectively. In other words, the higher likelihood class only had odds of between 33% and 49% of being flagged as invalid at each time point. The model estimated that 13.3% of students and 87% of students were in the higher likelihood and lower likelihood classes, respectively.

The conditional model results (not tabulated) indicated that boys were almost 30% more likely to be in the higher likelihood class than girls (OR = 1.28; p = .008), and students with minority status were nearly twice as likely (OR = 1.75; p < .001) to be members of the higher likelihood class.

Comparisons of Invalid and Valid Response Patterns

Table 3 provides estimates from four separate HLM equations, with the results from the within-person measures (time and invalid status) above the between-level variables (race and male). The estimates for "invalid" (student reported that he or she did not answer truthfully or did not pay attention in answering survey questions) in the table

represent the increase or decrease in the outcome measure (victimization and three school

climate scales) associated with a survey's invalid status, net of controls for survey wave,

race, and gender. Table 4 provides similar information for eight separate models, each

with a different problem behavior variable as the outcome. The estimates for 'invalid' in

Table 4 represent the increase or decrease in the problem behavior outcome associated

with a student's invalid status, net of controls. Tables 5 and 6 present results of the same

type and format as Tables 3 and 4, respectively, but these tables present results only for

those students who ever provided invalid survey respondents.

Table 3 presents results from the HLM analyses comparing surveys with

responses flagged as invalid versus surveys not flagged for invalid responses from 392

students over six waves on the victimization and school climate measures. The

significance value for the estimates of the invalid and time variables in the table was

adjusted from p = .05 to p = .04, in accordance with the Benjamini-Hochberg method

(Benjamini & Hochberg, 1995). As hypothesized, surveys flagged for invalid responding

had higher levels of aggressive attitudes (Cohen's d = .39; Cohen, 1988) and lower levels

of help-seeking (d = -.61; both $p < .04$) than unflagged surveys. The association between

validity status and victimization fell short of statistical significance ($d = .13$; $p < .10$).

There was not a statistically significant main effect for the measure of the prevalence of

bullying and teasing.

Table 4 presents the HLM analysis comparing valid and invalid survey

responding from 392 students over four waves on the risk behavior items. The

Benjamini-Hochberg adjusted significance value for the invalid survey and time variables

in the table was p = .02. Reports of alcohol use ($d = .78$), cigarette use ($d = .77$) , and

marijuana use (d = .67), were all significantly higher on surveys flagged for invalid responding versus unflagged surveys (p < .02 for all), net of the effects of gender, time, and race. Although initially significant at the .05 level, after applying the Benjamini-Hochberg adjustment, the adjusted p values for weapons carrying (d = .69) and thoughts of suicide (d - .45) were not significant (p < .10). Surveys flagged for invalid responding were not significantly different from unflagged surveys in reports of fighting, feelings of sadness, and school avoidance.

Table 5 shows results from a second series of HLM analyses confined to 136 students who provided at least one survey flagged for invalid responding, using reports of victimization and school climate as outcomes. The Benjamini-Hochberg adjusted significance value for the invalid survey and time variables in the table was p = .02. After controlling for effects of race, gender, and time, students reported higher aggressive attitudes (d = .31) and lower help-seeking (d = -.44), when their surveys were flagged for invalid responding in comparison to their own surveys that were not flagged. Although initially significant at the .05 level, after applying the Benjamini-Hochberg adjustment, the adjusted p value for victimization (d =.18) was not significant (p < .10).There were no statistically significant differences within students in their reports of the prevalence of bullying and teasing when their survey responses were flagged as invalid versus not flagged.

Table 6 shows results from a similar analysis, using the risk behavior items as outcomes. The Benjamini-Hochberg adjusted significance value for the invalid survey and time variables in the table was p = .003.Students reported higher rates of alcohol (d = .61) and cigarette (d = .53) use when their survey responses were flagged as invalid in

comparison to when their survey responses were not flagged. After applying the Benjamini-Hochberg adjustment, the adjusted p values for suicidal thoughts ($d = .34$), marijuana use ($d = .50$), and weapons carrying ($d = .55$) was not significant ($p < .10$). There were no statistically significant differences within students for measures of fighting, feelings of sadness, and school avoidance when their survey responses were flagged versus not flagged.

School Adjustment of Valid and Invalid Responders

Table 7 presents the results from regression equations that controlled for race and gender and used a dummy variable indicating that a respondent was ever flagged as invalid (1) versus never invalid (2) as the primary independent variable. In these regression, being male was significantly (.05) and positively associated with Science test scores, and significantly lower GPA. Nonwhite status was associated with significantly lower GPA, higher probability of having a disciplinary infraction, and significantly lower test scores across all tests. As could be expected from the prior analyses, being male and nonwhite was consistently associated with invalid responding. The Benjamini-Hochberg adjusted significance value was .006. As presented in Table 7, students who had produced at least one survey flagged for invalid responses were more than twice as likely as other students ($OR = 2.16$; p = .003) to have one or more disciplinary infractions, which was the only significant relation using the adjusted threshold. Although initially significant at the .05 level, after applying the Benjamini-Hochberg adjustment, the adjusted p value was $< .10$ for the following results: Invalid responders were 66% more likely to be suspended, had, on average a .10 lower overall GPA than other students (d = -.04), and had lower scores on Math (d = -.06), Reading (d = -.05), and Social Studies (d =

-.05) tests. Invalid responding students scored approximately 31, 23, and 24 points lower on achievement tests than other students.

Discussion

These results provide the first examination of the characteristics of students who reported that they answered untruthfully or carelessly on a self-report survey. The study examined six waves of surveys as well as independent measures of school adjustment for each student. Based on our screening criteria, 35.6% of the study sample generated at least one survey flagged for invalid responding during their three middle school years. However, these students typically generated only one or two flagged surveys out of the six they completed. Notably, only nine students (2%) marked the validity items negatively on more than half of their surveys. As a result, on average, only about 10% of surveys were flagged for invalid responding for each wave.

We identified only two statistically reliable patterns of validity responding: those who never generated a flagged survey and those who occasionally generated a flagged survey. The latent class analysis suggested there were just two groups rather than additional subgroups, such as students who repeatedly generated invalid survey responses versus those who did so only once. These findings suggest that invalid responding is not the habitual practice of a few students, nor an occasional practice of most students, but that a substantial group of students will occasionally report that they completed the survey in a dishonest or careless manner. Students were 30 percent more likely to produce an invalid survey response if they were boys and twice as likely to produce invalid survey responses if they were minority (non-white) students. These group differences seem consistent with general trends for boys and minority students to engage

in more misbehavior at school (Skiba, Trachok, Chung, Baker, & Hughes, 2012). In a larger sample, it would be useful to investigate differences within racial/ethnic subgroups and relate them to school attitudes and experiences.

Although invalid responding is not common, it is associated with elevated reports of high risk behavior and more negative perceptions of school in a previous study (Cornell et al., 2012) and in the present study. The HLM analyses in the present study showed that when students identified themselves as dishonest or careless, they were more likely to claim that they consumed alcohol, smoked cigarettes, and used marijuana. The effect sizes for these results according to Cohen (1988) were medium-sized (range $d =$ .67 to .78). These invalid responders also characterized their school climate as more supportive of aggressive attitudes (e.g., agreeing that "If you fight a lot, everyone will look up to you") and less conducive to help-seeking (e.g., disagreeing that "There are adults at this school I could turn to if I had a personal problem"). The effect sizes were small for aggressive attitudes (.39) and medium for help seeking (-.61). These students did not, however, produce elevations in reports of bullying victimization ($p <$.10). They also did not report higher rates of fighting, feeling sad, or avoiding school, and they did not describe more pervasive bullying and teasing in their school.

This study appears to be the first to examine response validity over repeated survey administrations to the same students. There were 136 students who generated both valid and invalid survey responses at different times. These within-student comparisons revealed that student reports of risk behavior and perceptions of school climate did vary in accord with their response to validity items. When students reported that they were not being honest or not answering carefully, they tended to endorse higher rates of risk

behavior and more negative perceptions of their school climate than when they reported being honest and answering carefully. The results of these within-student comparisons were consistent with the more general between-student comparisons of valid and invalid survey responding.

The consistent differences between surveys flagged or not flagged for invalid responses support the inference that students are intentionally inflating their reports of risk behavior and projecting a negative view of school conditions, although of course it is not possible to confirm this conclusion without direct knowledge of students' actual risk behavior and perceptions of their school. Some students might be answering in a dishonest manner and others may be answering in a careless manner that raises the rate of otherwise low-frequency behaviors such as carrying weapons to school.

Only about one-third of students generated invalid survey responses at least once during middle school, but this behavior appears to be a meaningful marker of the adjustment of these students in school. The 8th grade school adjustment measures indicated that the invalid responding students were more likely than other students to incur disciplinary infractions. There is weaker evidence ($p < .10$ after Benjamini-Hochberg adjustment) that they are more likely to be suspended from school, make slightly lower grades, and perform lower on state-mandated achievement tests.

There are probably multiple reasons for students to generate invalid survey responses. One possibility is that invalid responding reflects a rebellious or defiant state of mind in which certain students decide not to answer all items honestly or carefully. Although the inclination to dismiss the survey might be an occasional state of mind experienced by many students, the willingness to openly do so might be indicative of a

more pronounced and enduring rebelliousness or disaffection toward school. This quality might be a characteristic that also affects their attitude toward school authorities and their commitment to doing their school work, hence leading to their higher level of school misbehavior and lower academic performance. Study of the personality characteristics and mental states associated with invalid responding would be helpful in investigating this theory.

Of course, some students might give invalid responses because they have reading problems or are confused, in a hurry, or for some other reason inattentive to the task. These problems might be persistent enough to contribute to discipline problems and lower academic performance, too. In all of these cases, however, there is reason to doubt the credibility of their survey results and it seems preferable to estimate prevalence rates and school climate scores based on surveys that have passed validity screening. Certainly this is an area for further research.

Limitations and Directions for Further Study

An important limitation to this study is that these validity questions only identify students who were willing to acknowledge their dishonesty or carelessness in answering questions. There could be additional students who answer the validity questions appropriately but then answer other questions inaccurately. These validity questions offer only an incremental improvement in the quality of survey data and other means of improving survey validity screening might be considered, such as the use of nonsense items (e.g., "I have not seen a car in ten years") or the comparison of pairs of items with similar or opposite meaning .

The two validity questions address dishonesty and carelessness. With additional items it might be possible to distinguish two constructs and determine whether they have differential correlates. However, adding more validity items would lengthen the survey and bring more attention to the validity issue. In our experience, students react negatively to implausible items and too much emphasis on survey validity might be self-defeating because it could provoke the sort of rebellious response that surveyors want to avoid.

It should be recognized that validity screening does not assume that all students who negatively endorsed validity items were not telling the truth on all items. Some invalid responders might be accurately reporting on many items and so removal of participants classified as invalid responders could mean the removal of valid as well as invalid answers. The key issue is whether the removal of invalid respondents has a net positive impact on data quality. The study by Cornell and colleagues (2012) and the present findings support the contention that there is a net improvement in data quality by removal of invalid respondents, although of course further studies are needed. The most valuable evidence would come from studies using confidential survey administration and independent criteria to verify student reports.

The anonymity of survey research presumably makes it possible for adolescents to answer sensitive questions honestly, but this anonymity prevents researchers from verifying the accuracy of their claims. However, the Add Health survey (Carolina Population Center, n.d.) demonstrates that confidential surveys are viable and productive sources of information. Furthermore, there is evidence that confidential surveys are quite comparable to anonymous surveys. In a study by Chan, Myron, and Crawshaw (2005), students were randomly assigned to take a bullying survey anonymously or to write their

names on the survey. There were no statistically significant group differences in rates of peer aggression and victimization experiences. O'Malley, Johnston, Bachman, and Schulenberg (2000) found little or no group difference between anonymous and confidential adolescent reporting of drug use and illegal behaviors (i.e., stealing and weapon carrying) on the *Monitoring the Future* survey.

There is a need for more research on the conditions that affect participant attitudes toward taking a survey. How do survey explanations and instructions affect student responses? It seems likely that students are more cooperative when they receive more complete explanations of the purpose of the survey and information about its value. Teacher attitudes and engagement with students also may influence student cooperation and the validity of their answers. With the increasing use of online survey administration, there is considerable work needed to understand the effects of survey format and conditions of administration.

In conclusion, adolescent self-report surveys are a staple of psychological research because of their convenience and efficiency, but evidence for the accuracy of adolescent self-reports using external, independent criteria is limited. Our results indicate that a small proportion of adolescents will report they are not answering questions truthfully or carefully, and will produce results that are systematically different from those of other adolescents. Furthermore, the adolescents flagged for invalid responding produce survey results that differ consistently from their own results when they report that they are telling the truth and answering carefully. These findings provide further evidence of the need for more systematic research on the validity of adolescent self-report and support the use of validity screening items to improve data quality.

References

American School Health Association. (1989). *National Adolescent Student Health Survey*

    (NASHS). A report on the health of America's youth. Oakland, CA: Third Party

    Publishing Company.

Bandyopadhyay, S., Cornell, D. G., & Konold, T. R. (2009). Internal and external

    validity of three school climate scales from the School Climate Bullying Survey.

    School Psychology Review, *38*, 338–355.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical

    and powerful approach to multiple testing. *Journal of the Royal Statistical*

    Society. Series B (Methodological),57, 289-300. doi:10.2307/2346101

Branson, C., & Cornell, D. (2009). A comparison of self and peer reports in the

    assessment of middle school bullying. *Journal of Applied School Psychology, 25*,

    5–27. doi:10.1080/15377900802484133

Carolina Population Center (n.d.). *Add Health*. Retrieved from

    http://www.cpc.unc.edu/projects/addhealth

Chan, H. F. J., Myron, R., Crawshaw, M. (2005). The efficacy of non-anonymous

    measures of bullying. *School Psychology International, 26,* 443–458.

Clark, S., & Muthén, B. (2009). Relating latent class analysis results to variables not

    included in the analysis. Retrieved from

    https://www.statmodel.com/download/relatinglca.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (*2nd ed.*). New

    Jersey: Lawrence Erlbaum.

Collins, L. M., & Lanza, S. T. (2009). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Cornell, D. G. (2011). *The School Climate Bullying Survey: Description and research summary*. Unpublished report, University of Virginia, Charlottesville, Virginia.

Cornell, D. G. (2006). *School violence: Fears versus facts.* Hillsdale, New Jersey: Erlbaum.

Cornell, D.G., & Brockenbrough, K. (2004). Identification of bullies and victims: A comparison of methods. *Journal of School Violence, 3*, 63–87. doi:10.1300/J202v03n02_05

Cornell, D.G., Klein, J., Konold, T., & Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychological Assessment. 24*, 21–35.

Cornell, D. G., & Loper, A. B. (1998). Assessment of violence and other high-risk behaviors with a school survey. *School Psychology Review*, *27*, 317–330.

Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Ross, J., Hawkins, J., ... & Wechsler, H. (2010). Youth risk behavior surveillance-United States, 2009. *MMWR Surveill Summ*, *59*, 1–142.

Enders, C. K. 2010. Applied missing data analysis. New York, NY: Guilford.

Fan, X., Miller, B., Park, K., Winward, B., Christensen, M., Grotevant, H., & Tai, R. (2006). An exploratory study about inaccuracy and invalidity in adolescent self-report surveys. *Field Methods*, *18*, 223–244. doi:10.1177/152822X06289161

Fan, X., Miller, B., Christensen, M., Bayley, B., Park, K., Grotevant, H., & van Dulmen, M., & Dunbar, N. (2002). Questionnaire and interview inconsistencies exaggerated differences between adopted and non-adopted adolescents in a national sample. *Adoption Quarterly*, *6*, 7–27. doi:10.1300/J145v06n02_02

Hindelang, M.J., Hirschi, T., & Weis, J.G. (1979). Correlates of delinquency: the illusion of discrepancy between self-report and official measures. *American Sociological Review, 44,* 995–1014.

Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2010). *Monitoring the future national survey results on drug use, 1975-2009. Volume I: Secondary school students* (NIH Publication No. 10-7584). Bethesda, MD: National Institute on Drug Abuse.

Klein, J., Cornell, D., Konold, T. (2012). Relationships between school climate and student risk behaviors. *School Psychology Quarterly*, 27, 154-169. doi: 10.1037/a0029350

Levine, F. J., & Rosich, K. J. (1996). *Social causes of violence: Crafting a science agenda.* Washington, DC: American Sociological Association.

Miller, B. C., Fan, X., Christensen, M., Grotevant, H. D., & van Dulmen, M. (2000). Comparisons of adopted and nonadopted adolescents in a large, nationally representative sample. *Child Development*, *71*, 1458–1473. doi:10.1111/1467-8624.00239

Muthén, L., & Muthén, B. (1998-2010). Mplus user's guide (6th ed.). Los Angeles, CA: Muthén & Muthén.

National School Boards Association (1993). *Violence in the schools: How America's school boards are safeguarding your children*. Alexandria, VA: Author.

Nylund, K. L., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 535–569.

Olweus, D., & Limber, S. P. (2010). The Olweus Bullying Prevention Program: Implementation and evaluation over two decades. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *The Handbook of School Bullying: An International Perspective* (pp. 377-402). New York: Routledge.

O'Malley, P.M., Johnston, L.D., Bachman, J.G., & Schulenberg, J.E. (2000). A comparison of confidential versus anonymous survey procedures: Effects on reporting of drug use and related attitudes and beliefs in a national study of students. *Journal of Drug Issues*, *30*, 35–54.

Robers, S., Zhang, J., & Truman, J. (2010). *Indicators of school crime and safety: 2010*. (NCES 2011-002/NCJ 230812). Washington, DC: National Center for Education Statistics, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.

Rosenblatt, J. A., & Furlong, M. J. (1997). Assessing the reliability and validity of student self-reports of campus violence. *Journal of Youth and Adolescence*, *26*, 187–202. doi:10.1023/A:1024552531672

Sharkey, J. D., Furlong, M. J., & Yetter G. (2006). An overview of measurement issues in school violence and school safety research. In S. R. Jimerson & M. J. Furlong (Eds.), *The handbook of school violence and school safety: From research to practice* (pp. 121–134). Mahwah, NJ: Lawrence Erlbaum Associates.

Skiba, R. J., Trachok, M., Chung, C., Baker, T., & Hughes, R. (2012, April). Parsing disciplinary disproportionality: Contributions of behavior, student, and school characteristics to suspension and expulsion. Paper presented at the Annual Meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.

Virginia Department of Education (VDOE) (2010). *Virginia Standards of Learning and Common Core Standards.* Retrieved from http://www.doe.virginia.gov/testing/common_core/index.shtml

Witkin, G. (1991, April 8). Kids who kill; disputes once settled with fists are now settled with guns. *U.S. News & World Report*. *110,* 26–32.

Table 1.

Percentage of Invalid Surveys at each Survey Wave

Invalid at Each Wave	N	% of Total N = 382
Wave 1	42	11.41%
Wave 2	35	9.54%
Wave 3	33	8.94%
Wave 4	36	9.86%
Wave 5	28	8.12%
Wave 6	42	11.70%

Table 2.

Number of Waves with Invalid Surveys

Survey Validity	N	% Total (N=382)	% Invalid (N = 136)
Never Invalid	246	64.40%	--
Invalid at least one wave	136	--	
1 Wave Invalid	87	22.77%	63.97%
2 Waves Invalid	30	7.85%	22.06%
3 Waves Invalid	10	2.62%	7.35%
4 Waves Invalid	7	1.83%	5.15%
5 Waves Invalid	1	.26%	.74%
6 Waves Invalid	1	.26%	.74%

Table 3

HLM Comparisons of Invalid and Valid Surveys on Victimization and School Climate Measures

	Victimization				Help-Seeking		
	Estimate	S.E.	*p*		Estimate	S.E.	*p*
Within				Within			
Invalid	.30+	.17	.08	Invalid	-3.23*	.52	<.01
Time	-.09*	.03	<.01	Time	-.51	.06	<.01
Between				Between			
Male	.83[a]	.18	<.01	Male	-.75[a]	.36	.04
Non-White	-.65[a]	.20	<.01	Non-White	.08	.41	.84
	Aggressive Attitudes				Teasing and Bullying		
	Estimate	S.E.	*p*		Estimate	S.E.	*p*
Within				Within			
Invalid	1.62*	.39	<.01	Invalid	-.18	.23	.44
Time	.35*	.05	<.01	Time	.12*	.03	<.01
Between				Between			
Male	1.47[a]	.30	<.01	Male	-.63[a]	.25	<.01
Non-White	1.05[a]	.30	<.01	Non-White	.70[a]	.26	<.01

Note. N = 382 respondents, 6 waves, 2092 observations.
* *p* < .04. + *p* < .10. Significance level was adjusted from *p* = .05 to *p* = .04 using the Benjamini and Hochberg (1995) method for multiple domain comparisons
[a] *p* < .05; the *p* value for control variables was not adjusted.

Table 4

HLM Comparisons of Invalid and Valid Surveys on Risk Behavior Items

	Alcohol				Cigarettes		
	Estimate	S.E.	*p*		Estimate	S.E.	*p*
Within				Within			
Invalid	.43*	.16	.01	Invalid	.42*	.17	.01
Time	.07*	.02	<.01	Time	.05*	.02	<.01
Between							
Male	.05	.06	.45	Male	-.04	.09	.67
Non-				Non-			
White	.02	.08	.76	White	-.04	.12	.71
	Marijuana				Fighting		
	Estimate	S.E.	*P*		Estimate	S.E.	*p*
Within				Within			
Invalid	.37*	.16	.02	Invalid	.20	.14	.17
Time	.05*	.02	.01	Time	-.01	.02	.80
Between							
Male	.02	.15	.90	Male	.13[a]	.06	.05
Non-				Non-			
White	-.02	.20	.91	White	.07	.07	.32
	Weapon-Carrying				Feelings of Sadness		
	Estimate	S.E.	*P*		Estimate	S.E.	*p*
Within				Within			
Invalid	.26[+]	.13	.04	Invalid	.23	.16	.14
Time	.02	.01	.12	Time	.02	.03	.65
Between							
Male	.05	.09	.61	Male	-.24[a]	.09	.01
Non-				Non-			
White	-.01	.10	.93	White	-.11	.09	.24
	Suicide				Skip School Because Unsafe		
	Estimate	S.E.	*P*		Estimate	S.E.	*p*
Within				Within			
Invalid	.28[+]	.14	.05	Invalid	.14	.09	.13
Time	-.02	.02	.39	Time	-.02	.02	.18
Between							
Male	-.16[a]	.08	.04	Male	-.05	.05	.27
Non-				Non-			
White	-.10	.07	.15	White	-.03	.05	.55

Note. N = 382 respondents, 4 waves, 1528 observations.
* $p < .02$. + $p < .10$. Significance level was adjusted from $p = .05$ to $p = .02$ using the Benjamini and Hochberg (1995) method for multiple domain comparisons.
[a] $p < .05$; the *p* value for control variables was not adjusted.

Table 5

HLM Within-Students Comparisons of Invalid and Valid Surveys on Victimization

and School Climate Measures

	Victimization				Help-Seeking		
	Estimate	S.E.	*p*		Estimate	S.E.	*p*
Within				Within			
Invalid	.38[+]	.17	.03	Invalid	-2.70*	.51	<.01
Time	-.06	.04	.15	Time	-.69*	.12	<.01
Between							
Male	-.74[a]	.22	<.01	Male	-1.57[+]	.93	.09
Non-White	-.88	.22	<.01	Non-White	1.24[+]	.73	.09
	Aggressive Attitudes				Teasing and Bullying		
	Estimate	S.E.	*p*		Estimate	S.E.	*p*
Within				Within			
Invalid	1.40*	.39	<.01	Invalid	-.21	.25	.41
Time	.34*	.09	<.01	Time	.02	.06	.78
Between							
Male	1.52[a]	.47	<.01	Male	-.40	.34	.25
Non-White	.59	.42	.16	Non-White	.49	.33	.14

Note. N = 136 respondents, 6 waves, 816 observations.
* $p < .02$. + $p < .10$. Significance level was adjusted from $p = .05$ to $p = .02$ using the Benjamini and Hochberg (1995) method for multiple domain comparisons.
[a] $p < .05$; the *p* value for control variables was not adjusted.

Table 6

HLM Within-Students Comparisons of Invalid and Valid Surveys on Risk Behavior Items

	Alcohol				Cigarettes		
	Estimate	S.E.	*p*		Estimate	S.E.	*p*
Within				Within			
Invalid	.45*	.15	<.01	Invalid	.40*	.15	.01
Time	.09[+]	.04	.03	Time	.10[+]	.04	.02
Between							
Male	.12[a]	.04	<.01	Male	-.02	.09	.81
Non-				Non-			
White	-.02	.07	.75	White	-.20	.14	.16
	Marijuana				Fighting		
	Estimate	S.E.	*P*		Estimate	S.E.	*p*
Within				Within			
Invalid	.36[+]	.14	.01	Invalid	.13	.14	.34
Time	.10[+]	.04	.02	Time	.00	.05	.96
Between							
Male	.08	.08	.33	Male	.10	.10	.32
Non-				Non-			
White	-.19	.17	.25	White	-.09	.11	.44
	Weapon-Carrying				Feelings of Sadness		
	Estimate	S.E.	*P*		Estimate	S.E.	*p*
Within				Within			
Invalid	.27[+]	.12	.03	Invalid	.22	.16	.17
Time	.02	.03	.51	Time	.07	.06	.22
Between							
Male	.05	.04	.22	Male	-.46[a]	.18	.01
Non-				Non-			
White	-.08	.12	.46	White	-.38[a]	.19	.04
	Suicide				Skip School Because Unsafe		
	Estimate	S.E.	*p*		Estimate	S.E.	*p*
Within				Within			
Invalid	.25+	.13	.06	Invalid	.06	.12	.61
Time	-.03	.03	.40	Time	-.07	.04	.13
Between							
Male	-.13	.16	.39	Male	-.20	.13	.14
Non-				Non-			
White	-.24[a]	.12	.05	White	-.17	.16	.29

Note. N = 136 respondents, 4 waves, 544 observations.

* $p < .003$. + $p < .10$. Significance level was adjusted from $p = .05$ to $p = .003$ using the Benjamini and Hochberg (1995) method for multiple domain comparisons.

[a] $p < .05$; the *p* value for control variables was not adjusted.

Table 7

School Adjustment of Invalid and Valid Responders

Regression Method Adjustment Measure	Ever invalid B (SE)	*p*
Negative Binomial Regression		
School absences	-.16 (.09)	.074+
Logistic Regression		
1 or more disciplinary infractions	.77 (0.25)	.003*
1 or more school suspensions	.51(0.36)	.149
Ordinary Least Squares Regression		
GPA	-.13 (0.06)	.039[+]
Ordinary Least Squares Regression		
Math test score[1]	-31.25 (14.01)	.026[+]
Reading test score[1]	-23.23 (11.10)	.036+
Science test score[1]	-11.37 (7.60)	.135
Social studies test score[1]	-24.45 (11.51)	.034[+]
Writing test score[1]	-5.91 (4.43)	.182

Note. N for all models = 382; All effects net of male and non-white status
*$p < .006$. + $p < .10$. Significance level was adjusted from $p = .05$ to $p = .006$ using the Benjamini and Hochberg (1995) method for multiple domain comparisons.
[1] Test scores were run as outcomes simultaneously, and all effects are net of the intercorrelations between scores.

Figure 1. Conditional Item Response Probabilities (1 = invalid). N = 382
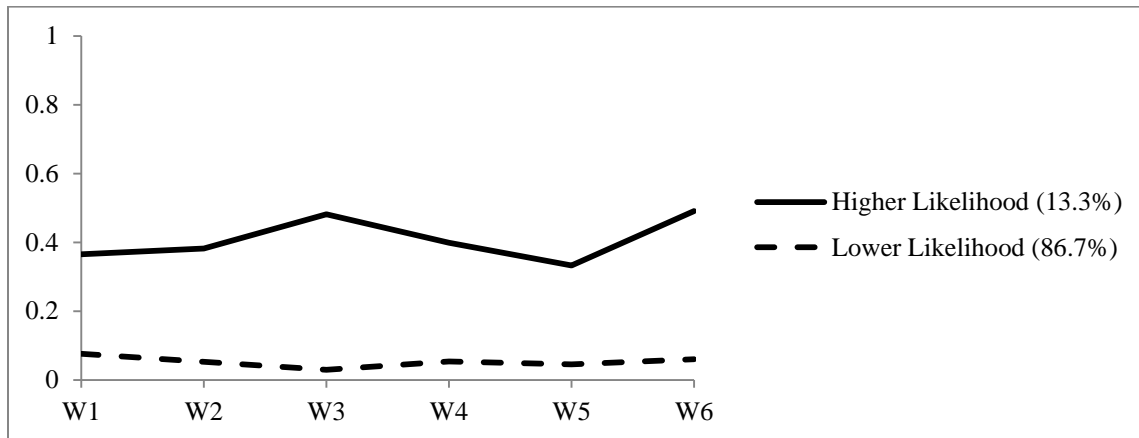
A Longitudinal Investigation of Peer- and Self-Reports of Bullying Victimization

Across Middle School

Michael W. Baly

Dewey G. Cornell

Peter Lovegrove

Curry School of Education,

University of Virginia, Charlottesville, Virginia, USA

Abstract

Cross-sectional studies indicate how many students are victims of bullying at a single time, but do not tell us whether the same students continue to be bullied or whether there is a cumulative impact of bullying over time. This study examined the longitudinal stability and the cumulative impact of victimization in a sample of 382 students assessed in the fall and the spring of grades 6, 7, and 8.Victimization assessed by both self- and peer-reports indicated substantial variability in who was bullied, with nearly 51% of students reporting bullying victimization during at least one of the six assessments. The cumulative impact of victimization over three years was demonstrated on grade 8 outcome measures of absences, disciplinary infractions, suspensions, grade point averages (GPA), standardized test scores, reports of youth risk behavior, and perceptions of school climate. This study provides new information about the cumulative impact of peer- and self-reported bullying across middle school.

Keywords: middle school student, self-report, peer-report, bullying, victimization, risk behavior, school climate, academic achievement, longitudinal

A Longitudinal Investigation of Peer- and Self-Reports of Bullying Victimization

Across Middle School

Anonymous self-report surveys are the primary instruments used for the measurement of bullying in schools. They are the basis for widely cited national prevalence rates (Nansel et al., 2001; Wang, Ianotti, & Nansel, 2009) and are used to measure the effectiveness of intervention programs (DeVoe & Bauer, 2010; Ttofi & Farrington, 2009; Wang, Ianotti, & Nansel, 2009).

Despite the ubiquity of anonymous self-report surveys in bullying research, they have some important limitations. Because the surveys are anonymous, it is not possible to conduct longitudinal studies to determine whether the same students are bullied over time or whether new students are bullied at different grade levels. Furthermore, it is not possible to determine how more persistently bullied students are affected.

Several studies have used longitudinal data to investigate trajectories of bullying victimization. Nylund and colleagues (2007) investigated the severity of peer victimization during three years of middle school in a sample of about 2000 students. Notably, they used anonymous surveys and could not track students over time, although the same cohort was tested repeatedly. Using latent class analyses (LCA), the authors identified three victim classes at each time point that were based on degree of victimization, rather than type of bullying experienced. The most victimized students tended to feel less safe at school and more depressed than other students. However it is impossible to tell whether the same students continue to be bullied or there are changes in victimization patterns.

Pellegrini and Bartini (2000) used a confidential survey that allowed them to track

survey responses and examine the course of victimization from the last year of elementary school (5th grade) through the first year of middle school (6th grade). Following a sample of 138 students across three time points (one in grade 5 and two in grade 6), the authors used a combination of self- and peer-reports to determine victim status. They reported that bullying victimization was relatively stable from the spring of grade 5 through the spring of grade 6. No information was reported about the concordance of victimization experiences across time points or the distinguishing features of individuals with more chronic victimization.

Other studies using longitudinal data have investigated the stability of victimization and the impact of timing and duration on various outcomes. Juvonen, Adrienne, and Graham (2000) examined the stability of confidential self-reports of victimization over a one year time period and related changes in psychological adjustment and school functioning. Concurrent perceptions of victimization, as opposed to early or chronic perceptions of victimization, predicted loneliness and self-worth. Changes in perceptions of victimization (and subsequent self-worth and loneliness) across the one-year period predicted grade point average, absenteeism, and teacher-rated social adjustment.

Rueger, Malecki, and Demaray (2011) administered confidential questionnaires to 694 7th and 8th graders in the fall and the spring during one academic year. The onset of victimization, as well as concurrent victimization at either time point, was associated with maladjustment in anxiety, depression, self-esteem, poor school attitude, GPA, and attendance. Rosen and colleagues collected data on the social victimization experiences of 153 students from 4th to 7th grade. They found that persistently victimized children had

continuously elevated levels of internalizing problems. Those who were intermittently bullied did not demonstrate chronic internalizing issues.

In summary, several studies have investigated the stability of victimization and links of certain trajectories to academic, emotional and behavioral adjustment problems. However, there is little information on the cumulative impact of bullying on student adjustment.

Peer Report of Bullying

Several authorities have recommended the use of peer-reports in combination with self-report assessments (Cornell & Cole, 2011; Cullerton-Sen & Crick, 2005; Graham, Bellmore, & Mize, 2006). The present study was designed to use both self- and peer- report measures to examine the prevalence and stability of bullying victimization across middle school, and to investigate the cumulative impact of victimization. Conventional peer-report methods involve asking students to identify classmates who have been bullied or who match a descriptive statement (Pakaslahti & Keltikangas-Jarvinen, 2000; Ladd & Kochenderfer-Ladd, 2002; Nabuzoka, 2003; Branson & Cornell, 2009).

Peer report has advantages as well as potential shortcomings when compared to self-report measures. One important advantage of peer reports is that they permit the identification of victims for follow-up study. Two Korean studies tracked students across grades 7 and 8 to show that certain demographic characteristics predicted victimization (Kim, Boyce, & Koh, 2009) and that symptoms of emotional and behavioral maladjustment were better understood as a consequence of being bullied rather than a contributing factor (Kim, Leventhal, Koh, Hubbard, & Boyce, 2006). Another advantage

is that peer report data is aggregated from multiple sources, which will likely increase reliability. Some children may have inaccurate perceptions about a classmate's involvement in bullying, but the combined judgment of a group of students should be more accurate. Also, peer-reports provide researchers and school personnel with the names of students so that reports can be verified. A disadvantage is that students might judge their peers by reputation, reporting them as victims even after bullying has stopped (Fox & Boulton, 2005).

Branson and Cornell (2009) compared peer and self-report methods, finding modest correspondence in reports of bullying victimization ($r = .32$) in a sample of 355 middle school students. Despite only modest agreement between self- and peer reports, both provided unique predictive value for school maladjustment, such as discipline referrals, school suspensions, and aggressive attitudes, supporting the view that both self and peer reports yield useful, non-redundant information (Juvonen, Nishina, & Graham, 2001; Ladd & Kochenderfer-Ladd, 2002; Swearer, Espalage, Vaillancourt, & Hymel, 2010).

Victimization and Adjustment

Many studies have linked victimization to increased emotional, behavioral, social, and academic problems (Boulton, Trueman, & Murray, 2008; Hawker & Boulton, 2000; Lopez & Dubois, 2005; Rigby, 2001; Ttofi, Farrington, Lösel, & Loeber, 2011; Vieno, Gini, & Santinello, 2011). However, a limited number of studies have examined the cumulative impact of repeated or prolonged victimization experiences (Juvonen, Wang, & Espinoza, 2011; Rueger, Malecki, & Demaray, 2011; Scholte, Engels, Overbeek, de Kemp, & Haselager, 2007).

School maladjustment. Juvonen and colleagues (2011) administered confidential surveys to about 2,300 6th graders and conducted multilevel models (MLMs) to examine the association between peer victimization and lower school performance across three years of middle school. Through rating scales for both self- and peer-report methods, they found that being a victim of bullying was consistently related to lower grade point averages and teacher-rated academic engagement during middle school (Juvonen, Wang, & Espinoza, 2011). Notably, this study demonstrates that the link between victimization and poor academic performance is largely due to individual differences in bullying experiences. Specifically, it suggests that high levels of victimization are consistently related to academic disengagement and poor grades throughout middle school. However, this study did not examine the relationship between persistent victimization and compromised academic achievement.

A meta-analysis of 33 studies by Nakamoto and Schwartz (2010) found a small but statistically significant negative association between victimization and concurrent academic achievement. Notably, the victimization informant (i.e., self, peer, or teacher reports) moderated the strength of the association, with studies that used peer-reports and multiple informants having larger effect sizes than those that used self-reports.

A recent review of research reported that bullying victimization is an important risk factor for childhood and adolescent psychopathology (Arseneault, Bowes, & Shakoor, 2010); however, the authors note that the majority of studies are either cross-sectional or have a retrospective design. One prospective, longitudinal study compared self-reports and parent-reports of victimization. Parent-reports were strong and significant predictors of depression, while self-reports were weak, but significant predictors of

depression at a later age. (Farrington, Loeber, Stallings, & Ttofi, 2011). This study did

not examine how more persistent bullying over time might affect emotional adjustment.

There is relatively little work on the cumulative impact of bullying victimization.

A three-year longitudinal investigation by Scholte and colleagues (2007) examined social

adjustment as a function of stability of victimization experiences. The authors categorized

participants into one of four categories based on peer-reports during both childhood and

adolescence: childhood only victims, adolescence-only victims, stable victims, and

noninvolved students. There were no apparent differences between the stable victims and

the adolescence-only victims, while the childhood only victim group was better socially

adjusted than both the adolescence only victim group and the stable victim group.

School Climate. School climate has been defined as the quality of interactions

among adults and students at school (Emmons, 1993). Research suggests that bullying

victimization is linked to more negative views of school. Victims of bullying feel less

connected to classmates and staff (O'Brennan & Furlong, 2009). One longitudinal study

found that students identified as victims during sixth grade reported having more negative

perceived school climate than nonvictims. However, there is a need for further research

on how the stability and cumulative impact of victimization is associated with school

climate, such as prevalence of teasing and bullying, aggressive attitudes, and willingness

to seek help.

Youth Risk Behaviors. Cross-sectional research on the association between

victimization and risk behaviors, such as smoking and alcohol and drug use, has

produced mixed findings. Several earlier studies found that victimization was not

associated with smoking and was linked to lower alcohol use (Morris, Zhang, & Bondy,

2006; Nansel et al., 2001). However, a more recent study found that victims are at

increased risk for both smoking and drinking behavior (Vieno, Gini, & Santinello, 2010).

Relational victimization, but not physical victimization, has been associated with

increased drug use (Sullivan, Farrell, & Kliewer, 2006). Furthermore, research indicates

that self-reported risk behavior is prone to exaggeration (Klein, Cornell, and Konold, in

press). Bullying victimization experiences have a range of severity, and some of the

mixed findings in the literature might be explained by studies that can examine the

chronicity of bullying.

Victims have also been found to have more internalizing difficulties. Depression

appears to be common among victims (Seals & Young, 2003). In one study, students who

reported being bullied "2 or 3 times" a month or more endorsed significantly higher

levels of depression compared to their peers (Solberg & Olweus, 2003). Not surprisingly,

victimization is also associated with increased thoughts of suicide (Kim, Koh, &

Leventhal, 2005).

Present Study

The current study examined the longitudinal stability and cumulative impact of

bullying victimization through both peer- and self-report measures. The study also tested

the cumulative impact of bullying on persistent victims relative to those who have been

bullied less frequently. Specifically, we hypothesized that students who were bullied

more persistently over three years of middle school would report higher rates of risk

behaviors in grade 8, including smoking, drinking, using drugs, fighting, weapon

carrying, missing school due to feeling unsafe, feeling sad, and considering suicide.

We also examined the cumulative impact of victimization and school adjustment, which included grade 8 independent records of student discipline infractions, suspensions, absences, grades, and standardized test scores. Although cross sectional studies on student adjustment are mixed, we hypothesized that the most persistently bullied students would have lower academic achievement and more compromised behavioral adjustment.

There were several challenges to studying the stability of bullying over the course of middle school. We needed a sample of students who completed repeated measures over a three-year period. Students transferred in and out of school from year to year; so, in order to obtain a sufficiently large sample, we aggregated data across three cohorts who entered school in grade 6 and remained through grade 8. Student surveys were confidential rather than anonymous in order track students across multiple time points and link survey responses to independent sources of school data, including grades, standardized test scores, discipline infractions, suspensions, and absences. The sample was limited to students who completed a survey in the fall and spring of each year so that we had six time points for each student.

Method

Participants

Data were collected from a public middle school in central Virginia from 2006 to 2011. The sample consisted of three cohorts of students who entered the sixth grade in years 2006, 2007, and 2008, and who remained enrolled through all three years of middle school. The initial sample consisted of 495 students who entered grade 6 during the first three years of the study period. There were 113 of these students who transferred away

from the school before completing grade 8, leaving a study sample of 382 students who attended the school for grades 6, 7, and 8.

Preliminary analyses compared the study sample with the students lost to transfer on fall of sixth grade surveys. There were significantly more minority compared to white students who transferred out of the school ($p = .011$). This is not surprising because student mobility tends to be higher among minority and lower income students (Xu, Hannaway, & D"Souza, 2009). The two groups did not differ in gender (boy/girl) or survey validity (valid/invalid). A multivariate analysis of variance (MANOVA) comparing the two groups on peer- and self-reported bullying victimization and the three school climate scales from their sixth grade fall survey was not significant.

Among the students who attended all three years, there were 292 students who completed six surveys (fall and spring of each grade), and 90 students who for various reasons (primarily absence from school due to illness) did not complete all six surveys. A second round of preliminary analyses compared these two groups on sixth grade fall surveys. The two groups did not differ in race, gender, or survey validity. A multivariate analysis of variance (MANOVA) comparing the two groups on peer- and self-reported bullying victimization and the three school climate scales was significant ($p = .003$, $\eta^2 = .052$).

The final sample of 292 students included 144 (49.3%) boys and 148 (50.7%) girls. There were 179 (61.3%) students who identified themselves as White, 50 (17.1%) as Black, 32 (11.0%) as Hispanic, 9 (3.1%) as Asian, and 22 (7.5%) as another ethnicity.

Measures

School Climate Bullying Survey. The School Climate Bullying Survey (SCBS;

BULLYING VICTIMIZATION ACROSS MIDDLE SCHOOL   108

Cornell, 2011) is a self-report instrument used to assess the prevalence of bullying and

related aspects of school climate. Validation studies demonstrated that the SCBS items

used as self-reports of bullying corresponded with independent measures obtained from

peer-reports and teacher nominations (Branson & Cornell, 2009; Cornell &

Brockenbrough, 2004). For example, Branson and Cornell (2009) conducted an ROC

(receiver operating characteristic) analysis which found that self-reported victimization

predicted peer reports of victimization with a sensitivity of .72 and specificity of .61, with

an AUC (Area Under Curve, a measure of effect size) of .71. In further support of its

criterion-related validity, self-reports of victimization on the SCBS were correlated with

depression (r = .30), negative perceptions of school (-.20), and lower academic

performance (-.11), whereas self-reports of bullying others were correlated with

aggressive attitudes (.38), discipline referrals (.28), and suspensions from school (.28)

(Branson & Cornell, 2009). The SCBS also yields estimates of bullying comparable to

the widely used Olweus Bullying Victimization Questionnaire (BVQ; Olweus, 1996;

Cornell, 2011).

　　　　The middle school version of the SCBS presented students with the following

definition of bullying: "Bullying is defined as the use of one's strength or popularity to

injure, threaten, or embarrass another person. Bullying can be physical, verbal, or social.

It is *not bullying* when two students of about the same strength argue or fight."

　　　　Students responded to four self-report items about bullying victimization. The

first is a general item about bullying: "By this definition, I have been bullied at school in

the past month." Items about physical, verbal, and social bullying victimization followed.

Physical bullying was defined as, "repeatedly hitting, kicking, or shoving someone

weaker on purpose;" verbal bullying was defined as, "repeatedly teasing, putting down, or insulting someone on purpose;" and social bullying was defined as, "getting others repeatedly to ignore or leave someone out on purpose." Response options for each item were *never*, *once or twice*, *about once per week*, or *several times per week*.

The four bullying victimization items were dichotomized at each time point to distinguish victims and non-victims. Students who marked either *about once per week* or *several times per week* were categorized as involved in bullying, consistent with the recommendation by Solberg and Olweus (2003).

In addition to the self-report bullying victimization items, students were asked to nominate their classmates whom they have seen being bullied. At the end of the survey, students were asked, "Who is being bullied?" and then presented with the item: "Help us stop bullying. Write the first and last name of students who have been bullied at school in the past month." Studies have demonstrated the validity of peer-report methods to identify victims of bullying (Branson & Cornell, 2009; Cornell & Brockenbrough, 2004). The peer-report item (total number of peer-reports) was dichotomized at each time point to distinguish victims from non-victims. Students who were nominated 3 or more times were categorized as victims. Previous research suggests that students with 1-2 nominations were unlikely to be considered victims of bullying when interviewed by school counselors adhering to the definition of bullying used in the survey (Phillips & Cornell, 2012).

The dichotomous self- and peer-report items provided a method for examining the stability of bullying victimization. Student victim status (victim or non-victim) was assessed in the fall and spring of grades 6, 7, and 8. Any student who was categorized as

a victim in either the fall or spring survey, or both, was categorized as a victim for that year. The internal consistency of the 4-item self-report scale for each of the six time points was measured. Cronbach's alpha was .75 for the fall of grade 6, .80 for the spring of grade 6, .81 for the fall of grade 7, .81 for the spring of grade 7, .77 for the fall of grade 8, and .86 for the spring of grade 8.

Total victimization was calculated for peer- and self-reports to measure the cumulative amount of victimization that students experienced during middle school. Peer- and self-reports were summed across the six time points to create cumulative scores.

School Climate Scales. The 20-item school climate portion of the SCBS includes a four-item Prevalence of Teasing and Bullying scale, a seven-item Aggressive Attitudes scale, and a nine-item Willingness to Seek Help scale (see Table 1). The Aggressive Attitudes scale included items such as "If you fight a lot, everyone will look up to you" and "Bullying is sometimes fun to do." The Prevalence of Teasing and Bullying scale included items such as "Bullying is a problem at this school" and "Students here often get teased about their clothing or physical appearance." The Willingness to Seek Help scale included sample items such as "There are adults at this school I could turn to if I had a personal problem" and "If I tell a teacher that someone is bullying me, the teacher will do something to help."

Bandyopadhyay, Cornell, and Konold (2009) established a reasonable fit for the 20-items with their corresponding scales through exploratory and confirmatory factor analysis in a middle school sample (n = 2,111). Additional analyses with a sample of 7,318 ninth grade students found that these scales were predictive of teacher reports of

bullying and teasing, teacher reports of student help-seeking behaviors, teacher reports of

gang-related violence, and school records of suspensions and expulsions

(Bandyopadhyay et al., 2009). Response options for each of these items were *strongly*

disagree, disagree, agree, and *strongly agree.*

　　　　Youth Risk Behaviors. The Youth Risk Behavior Survey (YRBS) is one of the

most widely used instruments to monitor student risk behavior (Eaton et al., 2008). In

order to meet federal grant reporting requirements for a project that started one year after

the present study began, the SCBS was augmented with eight risk behavior items derived

from the Youth Risk Behavior Surveillance System (Centers for Disease Control and

Prevention, 2010). These items are used nationwide to assess the prevalence of student

risk behavior. Specifically, they measure how often students smoke cigarettes, drink

alcohol, use marijuana, carry weapons on school property, stay home from school

because they felt unsafe, fight on school property, feel so sad or hopeless, and consider

attempting suicide (see Table 1). The items had answer choices ranging from either *0*

days to *20-30 days* or *0 times* to *6 or more times* (Centers for Disease Control and

Prevention, 2010). Because these items were not administered at the outset of the study,

they were completed only by the second and third cohorts (n = 251) of students.

　　　　Academic Achievement, Absences, and Discipline Infractions. Academic

achievement was measured using the Virginia Standards of Learning (SOL) exams, a

series of state-mandated tests given to all students for designated subjects (Virginia

Department of Education, 2010). In middle school, students completed exams in

Mathematics, Reading, Science, Social Studies, and Writing. In addition, student grade

point averages (GPAs) were calculated by averaging year-end grades in Language Arts,

Mathematics, Science, and Social Studies using the traditional American four-point grading scale (4 = A, 3 = B, 2 = C, 1 = D, 0 = F). Office staff recorded absences for each day a student missed school as well as a record of each time a student was referred for a disciplinary infraction or given a suspension.

Procedure

An important context for this study is that it was conducted in a school that had implemented the Olweus Bullying Prevention Program (OBPP; Olweus, 1994) since the fall of 2003. The OBPP aims to improve peer relations and promote a safe and positive school climate by fostering schoolwide awareness of bullying (Olweus, 1994). The program included schoolwide rules against bullying, regular classroom discussions, and other activities designed to engage students and raise awareness of bullying.

The SCBS was administered each fall and spring for a total of six time points to assess progress in reducing bullying during middle school. A Spanish translation of the survey was prepared by the central school administration, reviewed by the researchers, and made available to any students who requested it. Students who chose to use the Spanish version were not identified. School staff estimated that about a dozen students used the Spanish version.

During the first three years of the study, teachers administered surveys in a paper and pencil format in each classroom. However, the school system ceased using its scanner system and switched to using online survey administration. Consequently, in the last two years, surveys were administered online in a classroom containing enough computers for each student in the class. Students continued to complete the surveys in classroom groups under teacher supervision. Surveys were given a code number,

allowing student responses to remain confidential while being tracked. A single school staff member served as the code master who could link code numbers and student names. Data were provided to the researchers in archival form.

Screening of Invalid Data

Previous studies have recommended the use of validity screening for adolescent self-report surveys (Cornell, Klein, Konold, & Huang, 2011; Cornell & Loper, 1998; Rosenblatt & Furlong, 1997). The SCBS included two validity screening items: "I am telling the truth on this survey," and "I am not paying attention to how I answer this survey." Response options for both items were *strongly disagree, disagree, agree,* and *strongly agree.* The two items were dichotomized into those students who either disagree/strongly disagree or agree/strongly agree. Students who responded inappropriately to one or both of the validity items on the spring grade 8 survey were considered invalid responders. The grade 8 spring survey was chosen because it provided outcome data for the longitudinal analyses. In order to investigate the impact of validity screening, regression analyses were conducted with and without validity screening. Results without validity screening are presented here, followed by a summary of changes in results after validity screening.

<div align="center">

Results

</div>

Stability of Bullying Victimization

Self-Report. The majority of students (68.0%, $n = 208$) reported no victimization at each time point (fall or spring) during grade 6, 55.6% (170) reported no victimization during grades 6 and 7, and 49.4% (151) reported no victimization during grades 6, 7, and 8 (see Chart 1). In contrast, only about one-third (32.0%, $n = 98$) of students reported

being a victim during grade 6, 18% (55) reported being a victim during grades 6 and 7, and 11.8% (36) reported being a victim during grades 6, 7, and 8. The concordance (bullied or not bullied) across years was 72.3% (Kappa = .369) from grade 6 to grade 7, 76.0% (Kappa = .400) from grade 7 to grade 8, and 72.4% (Kappa = .312) from grade 6 to grade 8.

Peer-Report. Using the criterion of 3 or more peer-reports, the majority of students (87.8%, $n = 337$) were not identified as victims of bullying during grade 6, 81.3% (312) grades 6 and 7, or 78.9% (303) grades 6, 7, and 8 (see Chart 2). In contrast, far fewer students (12.2%, $n = 47$) were nominated as victims by their peers during grade 6, 4.9% (19) grades 6 and 7, or 2.1% (8) grades 6, 7, and 8. The concordance (bullied or not bullied) across years was 86.2% (Kappa = .339) from grade 6 to grade 7, 90.1% (Kappa = .434) from grade 7 to grade 8, and 85.7% (Kappa = .210) from grade 6 to grade 8.

Correlations among repeated measures of the total number of peer-reports of bullying victimization for the fall and spring of grades 6, 7, and 8 (see Table 2) showed moderate associations between consecutive time points. There were significant, yet moderate, correlations in self-reported victimization between the fall and spring of grade 6, $r = .443$, $p < .001$; spring of grade 6 and fall of grade 7, $r = .276$, $p < .001$; fall and spring of grade 7, $r = .380$, $p < .001$; spring of grade 7 and fall of grade 8, $r = .450$, $p < .001$; and fall and spring of grade 8, $r = .402$, $p < .001$.

Self-Report vs. Peer-report Concordance. The concordance (bullied or not bullied) between self-report and peer-report was 79.7% (Kappa = .229) in the fall of grade 6, 79.3% (Kappa = .184) in the spring of grade 6, 80.2% (Kappa = .216) in the fall

of grade 7, 80.8% (Kappa = .153) in the spring of grade 7, 87.6% (Kappa = .133) in the

fall of grade 8, and 83.3 (Kappa = .135) in the spring of grade 8.

Predictive Accuracy of Self-Report and Peer-Report

Multiple regressions (see Table 3) were conducted to determine the joint

predictive accuracy of peer-and self-report measures on a number of student outcomes in

grade 8. The four-item self-report scales in the fall and spring of grades 6, 7, and 8 were

summed to create a cumulative self-report measure. The cumulative peer-report measure

was the total number of peer-reports that students received across the six time points

during middle school. Correlations between the two scales ranged from .153 to .246 over

the six waves of surveys.

Race (white and non-white) and gender (boy and girl) were entered as control

variables in the first block of each regression. Cumulative peer- and self-reports were

added in the second block. The outcome measures included grade 8 school-based

behavior (absences, disciplinary infractions, and suspensions), academic achievement

(grade point averages and standardized test scores), school climate measures (Aggressive

Attitudes, Willingness to Seek Help, and Prevalence of Teasing and Bullying), and youth

risk behaviors (smoking cigarettes, drinking alcohol, using marijuana, weapon carrying,

missing school due to feeling unsafe, fighting in school, feeling sad, and considering

suicide).

School-Based Behavior. Students who had more cumulative peer-reported

victimization, but not self-reported victimization, were more likely to have grade 8

disciplinary infractions ($\beta = .124$, $p < .05$) (see Table 4). Neither peer- nor self-reports

were predictive of grade 8 suspensions or absences.

Academic Achievement. Peer-reported victimization, but not self-reported victimization, was associated with lower grade 8 combined (β = -.110, $p < .05$) and Math (β = -.134, $p < .05$) GPA. After controlling for race and gender, peer- and self-reports accounted for 2.3% of the variance in Math GPA, $\Delta R^2 = .023$, $F(2, 276) = 3.351$, $p < .05$. Neither peer- nor self-reports were predictive of Language Arts GPA, Science GPA, or Social Studies GPA.

Peer-reported victimization, but not self-reported victimization, was associated with lower Math (β = -.263, $p < .001$) and Reading (β = -.164, $p < .01$) SOL scores. Peer- and self-reports accounted for 5.7% of the variance in Math SOL scores, $\Delta R^2 = .057$, $F(2, 193) = 5.976$, $p < .01$, and 2.2% of the variance in Reading SOL scores, $\Delta R^2 = .022$, $F(2, 277) = 3.484$, $p < .05$. Neither peer- nor self-reports were predictive of Science or Social Studies SOL scores.

School Climate Scales. Self-reported victimization was associated with more aggressive attitudes (β = .219, $p < .001$), lower willingness to seek help (β = -.167, $p < .01$), and higher prevalence of teasing and bullying (β = .506, $p < .001$). Self- and peer-reports accounted for 4.5% of the variance in aggressive attitudes, $\Delta R^2 = .045$, $F(2, 271) = 7.010$ $p < .001$; 2.7% in willingness to seek help, $\Delta R^2 = .027$, $F(2, 273) = 4.076$ $p < .05$; and 23.1% in prevalence of teasing and bullying, $\Delta R^2 = .231$, $F(2, 275) = 42.198$, $p < .001$. Unexpectedly, peer-reported victimization was associated with lower aggressive attitudes and more willingness to seek help. Peer-reports were not associated with prevalence of teasing and bullying.

Youth Risk Behaviors. Self-reported victimization, but not peer-reported victimization, was associated with more weapon carrying (β = .148, $p < .05$), physical

fighting (β = .238, p < .01), feeling sad and hopeless (β = .368, p < .001), and considering

suicide (β = .232, p < .01) (see Table 4). Peer- and self-reports accounted for 7.2% of the

variance in physical fighting, ΔR^2 = .072, F(2, 164) = 6.510, p < .01; 13.1% in feeling

sad and hopeless, ΔR^2 = .131, F(2, 164) = 12.842, p < .001; and 4.5% in considering

suicide, ΔR^2 = .045, F(2, 164) = 4.006, p < .05. Both self- (β = .207, p < .01) and peer-

reports (β = .151, p < .05) were associated with missing school more often due to feeling

unsafe. Peer- and self-reports accounted for 9.0% of the variance in missing school due to

feeling unsafe, ΔR^2 = .090, F(2, 164) = 8.411, p < .001. Neither peer- nor self-reports

were associated with smoking cigarettes, drinking alcohol, or using marijuana in grade 8.

Predictive Accuracy After Validity Screening

Regressions predicting the School Climate scales and youth risk behaviors were

rerun after screening for validity in the spring of grade 8. Regressions generated the same

pattern of significant and non-significant findings for all three School Climate scales. All

but two youth risk behaviors had the same significant and non-significant findings. The

two differences in results were" (1) cumulative self-reported victimization no longer

predicted weapon carrying; and (2) peer- and self-reports no longer predicted considering

suicide.

Alternative Self-Report and Peer-report Measures

One additional question concerned whether using the dichotomous cut-off scores

for the self- and peer-report measures would produce different results for the regression

analyses. The alternative peer- and self-report measures were the sum of these

dichotomous scores across each of the six time points. Regression analyses with the

alternative measures produced a similar pattern of results and have been organized in an appendix that is available upon request.

Discussion

This study used peer- and self-report measures to investigate the course and cumulative impact of bullying victimization over three years of middle school. Our findings indicate the need for more longitudinal studies that can track students and their victimization experiences over time. Overall, there was a high degree of variance in reports of bullying across the six waves of surveys. The majority of bullying victimization is transient; however, almost 12% of students were bullied each year of middle school. Regression analyses demonstrated that more victimization during middle school predicted lower academic achievement, more negative perceptions of school climate, and higher levels of risk behaviors. There were some differences between grade 8 correlates of peer- and self-reports in examining the cumulative impact of victimization during middle school, supporting the use of both self- and peer-reports in bullying research.

Stability of Victimization

Self-reports. This study shows that self-reported bullying victimization decreases during middle school. While about a third of students reported being bullied in sixth and seventh grade, there is less victimization in eighth grade. Cross sectional studies have suggested decreases in bullying over the course of middle school; however, these studies cannot report decreases in the same cohort of students by tracking them over time. Instead they compare sixth, seventh, and eighth grade victimization rates at the same time point (Nansel et al., 2001; Wang, Iannotti, & Nansel, 2009). Because we assessed

bullying twice at each grade level, the rates of victimization reported in this study are higher than cross sectional studies that only measure bullying once per year. In a large, nationally representative sample, Nansel and colleagues found that 13%, 11%, and 8% of students reported being bullied in sixth, seventh, and eighth grades respectively. In contrast, 32%, 30%, and 25% of students in our study reported being bullied in sixth, seventh, and eighth grades respectively.

Victimization experiences varied widely. There were 49% who were never bullied, 39% who were sometimes bullied (i.e., bullied one or two years), and 12% who were persistently bullied throughout middle school (i.e., bullied each year of middle school). Most studies of prevalence rates do not differentiate between students who are victims just once and those who are bullied repeatedly. Studies examining the stability of bullying show that students who are persistently bullied demonstrate consistently elevated levels of internalizing problems (Rosen et al., 2009) and compromised academic outcomes (Rueger, Malecki, & Demaray, 2011) compared to those who are more intermittently bullied or never bullied at all. Despite the transience of bullying in the current sample, a sizeable minority of students reported being bullied all three years of middle school. These persistent victims represent an important group who may require more support from school personnel, underlining the need for self-report assessments to use confidential information that can track students over time.

Peer-reports. Peer-reports yielded substantially smaller rates of victimization than self-reports. There was approximately 61% less peer-reported victimization than self-reported victimization in sixth grade, 62% in seventh grade, and 68% in eighth grade. This finding is consistent with the limited number of studies that have directly compared

self- and peer-reports of bullying (Branson & Cornell, 2009; Pellegrini & Bartini, 2000). Furthermore, according to their classmates, the large majority of students (78.9%) were never bullied.

What might account for the much higher rates of self- over peer-reported victimization? Some researchers have pointed to the possibility that self-report measures are prone to inflation (Baly & Cornell, 2011; Kert, Codding, Tryon, & Shiyko, 2010). Students must incorporate the power imbalance criteria in order to accurately report their bullying experiences. If they do not apply this concept, they may include other forms of peer conflict that are not bullying, such as teasing amongst friends or arguments between two students of similar size and status. Peer-report measures are less hindered by this problem because multiple raters are used to identify victims. Only students who received three or more nominations were categorized as victims in the current study, because previous research has suggested that those with one or two nominations were unlikely to be considered victims of bullying when interviewed by school counselors (Phillips & Cornell, 2012). If this is the case, then it makes sense that self-report surveys would yield higher estimates because peer-reports contain a level of protections against students who are overly inclusive. However, it is also possible that students are less willing to tell teachers about victims of bullying. Additionally, students may still feel bullied even if their classmates do not see them as victims.

There was also limited correspondence between self- and peer-report measures at each time point (range: .15 to .25), which is consistent with previous studies that have compared these two methods (Branson & Cornell, 2009; Pellegrini & Bartini, 2000). In fact, there was very little overlap between the two measures. A range of only 1.4% to

5.1% (see Chart 3) of students were identified as victims through both self- and peer-reported victimization at each time point.

Although the combined judgment of a group of students should be more accurate, it is possible that peer-report measures fail to identify certain indices of victimization or types of victims. There could be bullying experiences that bystanders feel less obligated to report, such as when they are the only person to notice the incident or if they are worried about retribution from the bullies. In these cases, only the victims themselves are likely to report these events. In addition, peer reports may rely too heavily on stereotypes of victims as lower status, passive participants in bullying. This perception might bias students against identifying classmates who are in more popular circles, portray confidence, or behave more aggressively. Consequently, more research is needed on the comparative advantages and disadvantages of self versus peer reports of victimization and why students might be identified by one method but not the other.

Cumulative Impact of Victimization

Regression analyses revealed that peer- and self-reports of victimization predict different types of outcomes. Cumulative peer-reported victimization over three years in middle school was associated with more disciplinary infractions. This finding indicates that the students identified by peers as persistently bullied may be more inclined to be disruptive and break school rules, possibly as a maladaptive form of coping with the stress of repeated victimization. We can only speculate as to how repeated victimization might influence misbehavior in school as the data does not provide information about the specific behaviors that prompted these infractions. Despite the increased likelihood of discipline infractions, more peer-reported victimization did not predict more suspensions,

suggesting that these students' misbehavior was less severe. Also, predicting suspensions is more difficult because of its low base rate. Absences were also not predicted by peer-reported victimization.

Students who had more cumulative peer-reported victimization fared worse academically in eighth grade. Specifically, persistently victimized students had lower combined and math GPAs. Additionally, peer-reports were associated with lower Math and Language Arts standardized test scores. Previous studies have reported direct associations between victimization and academic achievement (Branson & Cornell, 2009; Juvonen, Wang, & Espinoza, 2011); however, the current findings are noteworthy because they account for victimization experiences that occur throughout middle school – instead of at just one time point – and suggest that there is a cumulative impact of those experiences on student grades. Furthermore, standardized test scores are a primary method of assessing academic achievement and development. Our study demonstrates that cumulative bullying is associated with lower standardized test scores at the end of middle school, which contributes to cross-sectional evidence that schools should provide early and continuous support for victims.

Curiously, cumulative self- and peer-reports were predictive of school climate scales in different directions. Students who self-reported more victimization were more likely to have aggressive attitudes and were less willing to seek help from teachers or other staff at school. In contrast, peer-reports predicted less aggressive attitudes and more help seeking. One possible explanation for these contradictory results is that the two types of reporting methods may be identifying somewhat different types of victims. Peers may fail to notice when their more aggressive classmates are bullied because their

BULLYING VICTIMIZATION ACROSS MIDDLE SCHOOL   123

attitudes are not typically associated with victims. These students are more prone to resolve conflicts with fighting and associate violence with higher status. So, it is not surprising that they would also be less inclined to seek help from adults in school.

It is possible that self-report is a better method for identifying bully-victims, which may explain why those who consistently report being bullied throughout middle school are more likely to hold aggressive attitudes and less likely to seek help. On the other hand, peer-reports may be a better method for identifying students who fit the stereotype of a victim. Victims who are less willing to fight back against bullies and are more apt to seek help from adults are probably recognized as needing more support, which provides classmates with the motivation to write their names down on a survey.

Cumulative self-reports predicted more risk behaviors than peer-reports, adding further support for the theory that self- and peer-reports identify different types of victims. Students with more victimization throughout middle school measured with either method were more likely to report missing school because they felt unsafe. However, students with higher levels of self-reported victimization were more likely to get into physical fights, feel sad and hopeless, and consider suicide by the end of middle school. These risk behavior findings indicate that cumulative self-reports capture victims who are more likely to develop depression and suicidal tendencies.

There were some notable gender and race differences in our findings for 8th grade outcomes. Boys scored higher on the state-mandated science test, whereas girls attained slightly higher grades in language arts. As might be expected, boys were more likely to express aggressive attitudes and less likely to report willingness to seek help from school personnel than were girls. Boys were also more likely to report getting in fights at school

and drinking alcohol, but less likely to report feeling sad or considering suicide than were girls. These gender differences might be reflected in different ways that boys and girls respond to bullying.

White students recorded fewer disciplinary infractions and school suspensions than non-white students. Also, white students obtained consistently higher grades and higher scores on state-mandated achievement tests than non-white students. These findings are consistent with other studies of racial disparities in both school discipline and academic performance, and it has been argued that school suspension practices have a damaging effect on academic performance without any positive impact on student behavior (Losen & Skiba, 2010). Notably, the minority students reported less willingness to seek help from school personnel than the white students, but there were no race differences in reports of risk behaviors (e.g., smoking, drinking alcohol, using marijuana, carrying a weapon, fighting in schools). In order to examine these findings further it would be important to distinguish outcomes associated with race from those due to student socio-economic status, but this information was not available for this study.

Limitations and Areas for Further Study

This study is limited to a sample of three cohorts from a single middle school in central Virginia. Results may differ across schools as well as student populations. The sample was also limited because a large portion (N = 113) of students transferred out of the school prior to completing eighth grade. Student mobility is a common concern in schools and tends to be higher among lower income and minority students, and is associated with lower academic performance (Xu, Hannaway, & d'Souza, 2009). Another source of attrition was students who remained in the school but for various reasons

(primarily absence from school) did not complete all six surveys. Those students who were lost to the sample might be different in their levels of bullying victimization, discipline infractions and suspensions, academic achievement, and risk behavior.

This sample was drawn from a school with an established bullying prevention program, which could have affected levels of self- and peer-reported bullying. Because of the bullying prevention program, school staff were perhaps more sensitive to the needs of victims and could have been more likely to engage in efforts to stop bullying when they learned about it or saw it happening. Nevertheless, a notable group of students (about 12%) reported being bullied each year of middle school. In schools without bullying prevention programs, the number of persistently bullied students might be larger, and the cumulative impact of bullying in schools without such programs might be even more severe. However, as bullying prevention programs become increasingly prevalent, these findings become more relevant.

One direction for future study is to consider how 8th grade students manage the transition to 9th grade after attending a middle school with an active bullying prevention program. Students in the 9th grade might be at increased risk for bullying since they become the youngest students in their new high school. However, bullying prevention programs teach students to take an active role supporting students who are teased or picked on and to seek help from adults when they are not able to resolve a bullying situation. Perhaps these 8th grade students will be better prepared for 9th grade experiences of bullying. Overall, there is need for research on the longitudinal impact of bullying prevention efforts as students progress through elementary, middle, and high schools.

Another limitation is that our results rely on student report data without verification of the accuracy of student perceptions. There is evidence students might use an overly broad conception of bullying (Baly & Cornell, 2011) and that some students might exaggerate their involvement in risk behaviors (Cornell et al., 2012). Also, results from self-report data could be confounded by shared method variance, which can create artificial correlations between scales or constructs because of the consistency in how the reporter answered.

It should also be noted that we only measured bullying at two time points during the year and students are asked to report if they have been bullied within the last thirty days. Rates of victimization could be even higher if some students happened not to be bullied within these time periods. Future studies should track the victimization of students in a larger sample across multiple schools.

Conclusion

Our study demonstrates how confidential surveys that can track student responses over time and yield valuable information that cross-sectional research cannot provide. About half of students are bullied at one time point over the course of middle school, but the large majority of victims are not bullied across multiple years.

These findings also support the view that both self and peer reports yield useful, non-redundant information. Cumulative self-reports were predictive of other self-reported items. Namely, more self-reported victimization was associated with more aggressive attitudes, perceptions that teasing and bullying was more prevalent, and lower willingness to seek help. Self-reports were also associated with higher levels of feeling sad and unsafe, thinking about suicide, carrying weapons, and fighting. Cumulative peer-reports

were predictive of independent measures of school adjustment and predicted more

discipline infractions and lower academic achievement.

References

Arseneault, L., Bowes, L., & Shakoor, S. (2010). Bullying victimization in youths and
mental health problems:'Much ado about nothing'. *Psychological Medicine, 40*,
717–729.

Bandyopadhyay, S., Cornell, D. G., & Konold, T. R. (2009). Internal and external
validity of three school climate scales from the School Climate Bullying Survey.
School Psychology Review, 38, 338-355.

Baly, M., & Cornell, G. (2011). Effects of an educational video on the measurement of
bullying by self-report. *Journal of School Violence, 10*, 221-238.

Boulton, M. J., Trueman, M., & Murray, L. (2008). Associations between peer
victimization, fear of future victimization and disrupted concentration on class
work among junior school pupils. *British Journal of Educational Psychology, 78*,
473–489. doi:10.1348/000709908X320471

Branson, C. E., & Cornell, D. G. (2009). A comparison of self and peer reports in the
assessment of middle school bullying. *Journal of Applied School Psychology, 25*,
5-27. doi:10.1080/15377900802484133

Centers for Disease Control and Prevention. [Youth Risk Behavior Surveillance —
United States, 2009]. Surveillance Summaries, [June 4, 2010]. MMWR 2010; 59
(No. SS-5).

Chan, H. F. John (2002) 'The *School Life Survey* – A New Instrument for Measuring
Bullying and Victimization', unpublished doctoral dissertation, University of
Hull, UK.

Chan, H. F. J. (2006). Systemic patterns in bullying and victimization. *School Psychology International, 27*, 352-369.

Cornell, D. (2011). *The School Climate Bullying Survey: Description and research summary*. Charlottesville, VA: University of Virginia.

Cornell, D., & Brockenbrough, K. (2004). Identification of bullies and victims: A comparison of methods. *Journal of School Violence, 3*, 63–87. doi:10.1300/J202v03n02_05

Cornell, D. & Cole, J. (2011). Assessment of bullying. In S.R. Jimerson, A. B. Nickerson, M. J. Mayer, & M.J. Furlong (Eds.), *The handbook of school violence and school safety: International research and practice (2nd Edition)*. Mahwah, New Jersey: Routledge.

Cornell, D., Klein, J., Konold, T., & Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychological Assessment, 24*, 21-35. doi:10.1037/a0024824

Cornell, D., & Loper, A. (1998). Assessment of violence and other high-risk behaviors with a school survey. *School Psychology Review, 27*, 317-330.

Cullerton-Sen, C., & Crick, N. R. (2005). Understanding the effects of physical and relational victimization: The utility of multiple perspectives in predicting social-emotional adjustment. *School Psychology Review, 34,* 147–160.

DeVoe, J. F., & Bauer, L. (2010). Student victimization in US schools: Results from the 2007 school crime supplement to the national crime victimization survey. NCES 2010-319. *National Center for Education Statistics,* 75.

Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Ross, J., Hawkins, J., Harris, W. A.,
    Lowry, R., McManus, T., Chyen, D., Lim, C., Brener, N. D., & Wechsler, H
    (2008). Youth Risk Behavior Surveillance --- United States, 2007. *MMWR
    Surveillance Summaries*. *57*(SS04): 1-131.

Emmons, C. L. (1993). School development in an inner city: An analysis of factors
    selected form Comer's program using latent variable structural equations
    modeling. *Dissertation Abstracts International, 54*, 1287A.

Farrington, D. P., Loeber, R., Stallings, R., & Ttofi, M. M. (2011). Bullying perpetration
    and victimization as predictors of delinquency and depression in the pittsburgh
    youth study. *Journal of Aggression, Conflict and Peace Research, 3*, 74-81.

Fox, C. L., & Boulton, M. J. (2005). The social skills problems of victims of bullying:
    Self, peer and teacher perceptions. *British Journal of Educational Psychology*, *75*,
    313–328.

Graham, S., Bellmore, A. D., & Mize, J. (2006). Peer victimization, aggression, and their
    co-occurrence in middle school: Pathways to adjustment problems. *Journal of
    Abnormal Child Psychology, 34,* 363–378.

Hawker, D. S. J., & Boulton, M. J. (2000). Twenty years' research on peer victimization
    and psychosocial maladjustment: A meta-analytic review of cross-sectional
    studies. *Journal of Child Psychology and Psychiatry, 41*, 441–455.
    doi:10.1111/1469-7610.00629

Juvonen, J., Nishina, A., & Graham, S. (2000). Peer harassment, psychological
    adjustment, and school functioning in early adolescence. *Journal of Educational
    Psychology, 92*, 349-359.

Juvonen, J., Nishina, A., & Graham, S. (2001). Self-views versus peer perceptions of
     victim status among early ado- lescents. In J. Juvonen & S. Graham (Eds.), *Peer*
     harassment in school: A plight of the vulnerable and victimized (pp. 105–124).
     New York: Guilford.

Juvonen, J., Wang, Y., & Espinoza, G. (2011). Bullying experiences and compromised
     academic performance across middle school grades. *The Journal of Early*
     Adolescence, 31, 152.

Kert, A. S., Codding, R. S., Tryon, G. S., & Shiyko, M. (2010). Impact of the word
     'bully' on the reported rate of bullying behavior. *Psychology in the Schools, 47,*
     193–204. doi:10.1002/pits.20464

Kim, Y. S., Boyce, W. T., Koh, Y. J., & Leventhal, B. L. (2009). Time trends,
     trajectories, and demographic predictors of bullying: A prospective study in
     Korean adolescents. *Journal of Adolescent Health, 45*, 360-367.

Kim, Y. S., Koh, Y. J., & Leventhal, B. (2005). School bullying and suicidal risk in
     Korean middle school students. *Pediatrics, 115*, 357-363.

Kim, Y. S., Leventhal, B. L., Koh, Y. J., Hubbard, A., Boyce, W. T. (2006). School
     bullying and youth violence: Causes or consequences of psychopathologic
     behavior. *Archives of General Psychiatry, 63*, 1035-1041. doi:
     10.1001/archpsyc.63.9.1035

Klein, J., Cornell, D., Konold, T. (in press) Relationships between bullying school
     climate and student risk behaviors. *School Psychology Quarterly*.

Ladd, G., & Kochenderfer-Ladd, B. (2002). Identifying victims of peer aggression from

    early to middle childhood: Analysis of Cross-informant data from concordance,

    estimation of relational adjustment, prevalence of victimization, and

    characteristics of identified victims. *Psychological Assessment, 14*, 74-96. doi:

    10.1037/1040-3590.14.1.74

Lee, T., & Cornell, D. (2010). Concurrent validity of the Olweus bully/victim

    questionnaire. *Journal of School Violence, 9*, 56–73.

    doi:10.1080/15388220903185613

Lopez, C., & DuBois, D. L. (2005). Peer victimization and rejection: Investigation of an

    integrative model of effects on emotional, behavioral, and academic adjustment in

    early adolescence. *Journal of Clinical Child & Adolescent Psychology, 34*, 25–36.

    doi:10.1207/s15374424jccp3401_3

Losen, D., & Skiba, R. (2010). *Suspended education: Urban middle schools in crisis*.

    Montgomery, AL: Southern Poverty Law Center. http://www.splcenter.org/get-

    informed/publications/suspended-education

Morris, E. B., Zhang, B., & Bondy, S. J. (2006). Bullying and smoking: Examining the

    relationships in ontario adolescents. *Journal of School Health, 76*, 465-470.

Nabuzoka, D. (2003). Teacher ratings and peer nominations of bullying and other

    behaviour of children with and without learning difficulties. *Educational*

    Psychology, 23, 307-321. doi: 10.1080/0144341032000060147

Nakamoto, J., & Schwartz, D. (2010). Is peer victimization associated with academic

    achievement? A meta-analytic review. *Social Development, 19*, 221-242.

    doi:10.1111/j.1467-9507.2009.00539.x

Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment. *JAMA: The Journal of the American Medical Association, 285*, 2094–2100.

Nylund, K., Bellmore, A., Nishina, A., & Graham, S. (2007). Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? *Child Development, 78*, 1706-1722.

Olweus, D. (1994). Bullying at school: Long-term outcomes for the victims and an effective school-based intervention program. In L. R. Huesmann, & L. R. Huesmann (Eds.), *Aggressive behavior: Current perspectives.* (pp. 97-130). New York, NY US: Plenum Press.

Olweus, D. (1996). The revised Olweus bully/victim questionnaire. Bergen, Norway: Research Center for Health Promotion (HEMIL), University of Bergen.

Phillips, V., & Cornell, D. (2012). Identifying victims of bullying: Use of counselor interviews to confirm peer nominations. *Professional School Counseling, 15*, 123-131. doi:10.5330/PSC.n.2012-15.123

Rigby, K. (2001). Health consequences of bullying and its prevention in schools. In J. Juvonen, & S. Graham (Eds.), *Peer harassment in school: The plight of the vulnerable and victimized* (pp. 310−331). New York, NY: Guilford.

Rosen, L. H., Underwood, M. K., Beron, K. J., Gentsch, J. K., Wharton, M. E., & Rahdar, A. (2009). Persistent versus periodic experiences of social victimization: Predictors of adjustment. *Journal of Abnormal Child Psychology, 37*, 693-704.

Rosenblatt, J. A., & Furlong, M. J. (1997). Assessing the reliability and validity of

    student self-reports of campus violence. *Journal of Youth and Adolescence, 26*,

    187-202.

Rueger, S. Y., Malecki, C. K., & Demaray, M. K. (2011). Stability of peer victimization

    in early adolescence: Effects of timing and duration. *Journal of School*

    Psychology, 49, 443-464. doi:10.1016/j.jsp.2011.04.005

Pakaslahti, L., & Keltikangas-Jarvinen, L. (2000). Comparisons of peer, teacher, and self-

    assessments on adolescent direct and indirect aggression. *Educational*

    Psychology, 20, 177-190. doi: 10.1080/713663710.

Pellegrini, A. D., & Bartini, M. (2000). A longitudinal study of bullying, victimization,

    and peer affiliation during the transition from primary school to middle school.

    American Educational Research Journal, 37, 699.

Scholte, R. H. J., Engels, R. C. M. E., Overbeek, G., de Kemp, R. A. T., & Haselager, G.

    J. T. (2007). Stability in bullying and victimization and its association with social

    adjustment in childhood and adolescence. *Journal of Abnormal Child Psychology,*

    35, 217-228.

Seals, D., & Young, J. (2003). Bullying and Victimization: Prevalence and relationship to

    gender, grade level, ethnicity, self-esteem, and depression. *Adolescence, 38*, 735-

    747

Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the

    Olweus Bully/Victim questionnaire. *Aggressive Behavior, 29*, 239-268.

Sullivan, T. N., Farrell, A. D., & Kliewer, W. (2006). Peer victimization in early

    adolescence: Association between physical and relational victimization and drug

    use, aggression, and delinquent behaviors among urban middle school students.

    Development and Psychopathology, 18, 119.

Swearer, S. M., Espelage, D. L., Vaillancourt, T., & Hymel, S. (2010). What can be done

    about school bullying?: Link- ing research to educational practice. *Educational

    Researcher, 39,* 38–47. doi:10.3102/0013189X09357622

Thunfors, P., & Cornell, D. (2008). The popularity of middle school bullies. *Journal of

    School Violence, 7*, 65-82. doi: 10.1300/J202v07n01_05

Ttofi, M., & Farrington, D. (2009). What works in preventing bullying: Effective

    elements of anti-bullying programmes. *Journal of Aggression, Conflict and Peace

    Research, 1*, 13–24.

Ttofi, M. M., Farrington, D. P., Lösel, F., & Loeber, R. (2011). Do the victims of school

    bullies tend to become depressed later in life? A systematic review and meta-

    analysis of longitudinal studies. *Journal of Aggression, Conflict and Peace

    Research, 3*, 63-73.

Vieno, A., Gini, G., & Santinello, M. (2011). Different forms of bullying and their

    association to smoking and drinking behavior in Italian adolescents. *Journal of

    School Health, 81*, 393-399.

Wang, J., Iannotti, R. J., & Nansel, T.R. (2009). School bullying among adolescents in

    the United States: Physical, verbal, relational, and cyber. *Journal of Adolescent

    Health, 45,* 368–375. doi:10.1016/j.jadohealth.2009.03.021

Xu, Z., Hannaway, J., & D"Souza, S. (2009). *Student transience in North Carolina: The*
      effect of school mobility on student outcomes using longitudinal data. Calder
      Working Paper No. 22. National Center for Analysis of Longitudinal Data in
      Education Research. http://www.urban.org/publications/1001256.htm

Figure 1
Stability of Self-Reported Bullying Victimization Across Three Years of Middle School

Grade 6 (Column %)	Grade 7 (Column %)	Grade 8 (Column %)
Not Bullied 208 (68.0%)	Not Bullied 170 (55.6%)	Not Bullied 151 (49.4%)
		Newly Bullied 19 (6.2%)
	Newly Bullied 38 (12.4%)	Not Bullied 26 (8.5%)
		Still Bullied 12 (3.9%)
Bullied 98 (32.0 %)	Not Bullied 43 (14.0%)	Not Bullied 35 (11.4%)
		Again Bullied 8 (2.6%)
	Still Bullied 55 (18.0%)	Not Bullied 19 (6.2%)
		Still Bullied 36 (11.8%)

Figure 2

Stability of Peer-Reported Bullying Victimization Across Three Years of Middle School

Grade 6 (Column %)	Grade 7 (Column %)	Grade 8 (Column %)
Not Bullied 337 (87.8%)	Not Bullied 312 (81.3%)	Not Bullied 303 (78.9%)
		Newly Bullied 9 (2.3%)
	Newly Bullied 25 (6.5%)	Not Bullied 15 (3.9%)
		Still Bullied 10 (2.6%)
Bullied 47 (12.2%)	Not Bullied 28 (7.3%)	Not Bullied 25 (6.5%)
		Again Bullied 3 (0.8%)
	Still Bullied 19 (4.9%)	Not Bullied 11 (2.9%)
		Still Bullied 8 (2.1%)

Figure 3

Overlap Between Peer- and Self-Reports of Victimization
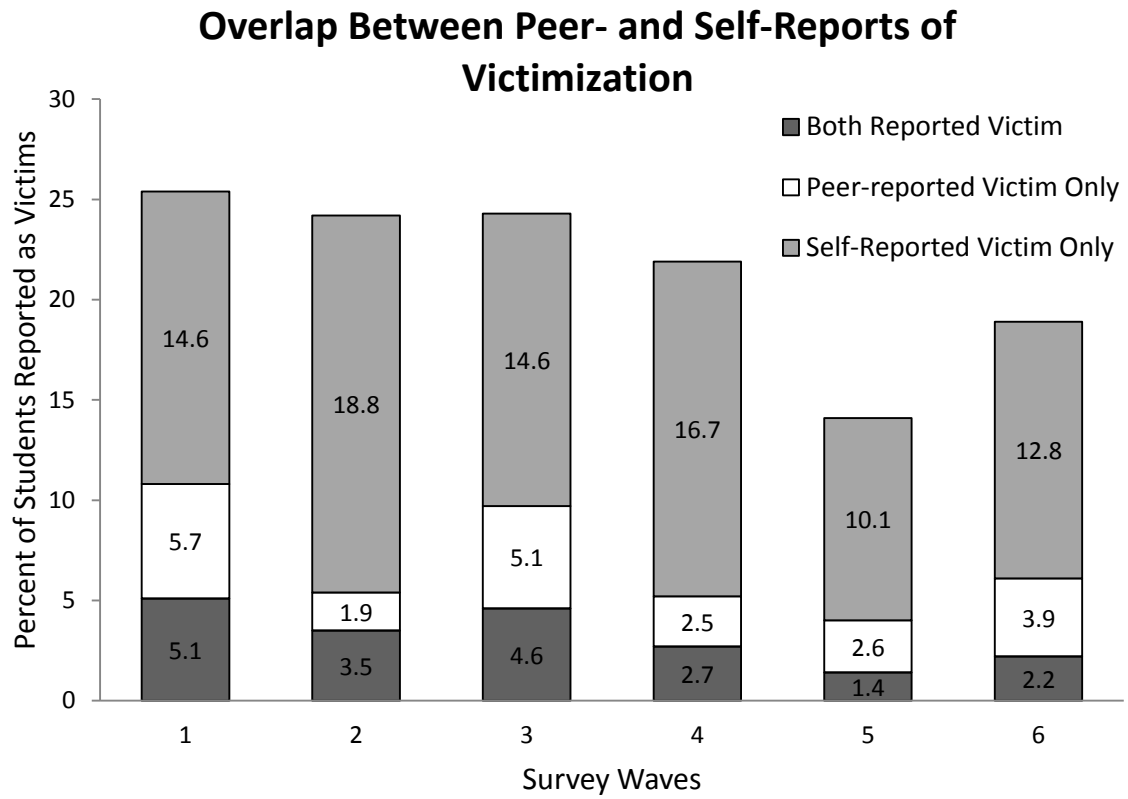
Table 1

School Climate Scales and Youth Risk Behaviors

Aggressive Attitudes

1. If someone threatens you, it is okay to hit that person.
2. It feels good when I hit someone.
3. Bullying is sometimes fun to do.
4. Sometimes you only have two choices—get punched or punch the other person first.
5. Students who are bullied or teased mostly deserve it.
6. If you fight a lot, everyone will look up to you.
7. If you are afraid to fight, you won't have many friends.

Willingness to Seek Help

1. If I tell a teacher that someone is bullying me, the teacher will do something to help.
2. Students tell teachers when other students are being bullied.
3. Teachers here make it clear to students that bullying is not tolerated.
4. There are adults at this school I could turn to if I had a personal problem.
5. If another student was bullying me, I would tell one of the teachers or staff at school.
6. Students here try to stop bullying when they see it happening.
7. The teachers at this school are genuinely concerned about me.
8. If another student talked about killing someone, I would tell one of the teachers or staff at school.
9. If another student brought a gun to school, I would tell one of the teachers or staff at school.

Prevalence of Teasing and Bullying

1. Bullying is a problem at this school.
2. Students here often get teased about their clothing or physical appearance.
3. There is a lot of teasing about sexual topics at this school.
4. Students here often get put down because of their race or ethnicity.

Youth Risk Behaviors

During the past 30 days…
1. On how many days did you smoke cigarettes?
2. How many days did you have at least one drink of alcohol?
3. How many times did you use marijuana?
4. On how many days did you carry a weapon such as a gun, knife, or club on school property?
5. On how many days did you not go to school because you felt you would be unsafe at school or on your way to or from school?

During the past 12 months…
1. How many times were you in a physical fight on school property?
2. Did you ever feel so sad or hopeless almost every day for two weeks or more in a row that you stopped doing some usual activities?
3. Did you ever seriously consider attempting suicide?

Table 2

Correlations among Repeated Measures of Self-Reports and Peer-reports of Bullying Victimization for Fall and Spring of Grades 6, 7, and 8.

		Fall 6		Spring 6		Fall 7		Spring 7		Fall 8		Spring 8	
		Self	Peer	Self	Peer	Self	Peer	Self	Peer	Self	Peer	Self	Peer
		r (p)	r (p)	r (p)	r (p)	r (p)	r (p)	r (p)	r (p)	r (p)	r (p)	r (p)	r (p)
Fall 6	Self	-	.243 (.000)	.265 (.000)	.231 (.000)	.183 (.000)	.066 (.203)	.120 (.024)	.069 (.184)	.175 (.001)	.036 (.492)	.205 (.000)	.126 (.016)
	Peer	.243 (.000)	-	.231 (.000)	.443 (.000)	.066 (.203)	.329 (.000)	.069 (.184)	.276 (.000)	.036 (.492)	.195 (.000)	.126 (.016)	.237 (.000)
Spring 6	Self	.265 (.000)	.212 (.000)	-	.246 (.000)	.456 (.000)	.175 (.001)	.269 (.000)	.272 (.000)	.241 (.000)	.154 (.003)	.200 (.000)	.140 (.007)
	Peer	.212 (.000)	.443 (.000)	.246 (.000)	-	.175 (.001)	.276 (.000)	.272 (.000)	.209 (.000)	.154 (.003)	.129 (.011)	.140 (.007)	.229 (.000)
Fall 7	Self	.183 (.000)	.184 (.000)	.456 (.000)	.187 (.000)	-	.233 (.000)	.394 (.000)	.228 (.000)	.377 (.000)	.178 (.001)	.286 (.000)	.130 (.012)
	Peer	.184 (.000)	.329 (.000)	.187 (.000)	.276 (.000)	.233 (.000)	-	.228 (.000)	.380 (.000)	.178 (.001)	.350 (.000)	.130 (.012)	.258 (.000)
Spring 7	Self	.120 (.024)	.166 (.001)	.259 (.000)	.206 (.000)	.394 (.000)	.057 (.279)	-	.196 (.000)	.328 (.000)	.154 (.003)	.246 (.000)	.137 (.009)
	Peer	.166 (.001)	.276 (.000)	.206 (.000)	.209 (.000)	.057 (.279)	.380 (.000)	.196 (.000)	-	.154 (.003)	.450 (.000)	.137 (.009)	.398 (.000)
Fall 8	Self	.175 (.001)	.173 (.001)	.241 (.000)	.145 (.007)	.377 (.000)	.190 (.000)	.328 (.000)	.200 (.000)	-	.155 (.004)	.238 (.000)	.143 (.008)
	Peer	.173 (.001)	.195 (.000)	.145 (.007)	.129 (.011)	.190 (.000)	.350 (.000)	.200 (.000)	.450 (.000)	.155 (.004)	-	.143 (.008)	.402 (.000)
Spring 8	Self	.205 (.000)	.140 (.008)	.200 (.000)	.118 (.025)	.286 (.000)	.136 (.010)	.246 (.000)	.127 (.016)	.238 (.000)	.085 (.105)	-	.153 (.004)
	Peer	.140 (.008)	.237 (.000)	.118 (.025)	.229 (.000)	.136 (.010)	.258 (.000)	.127 (.016)	.398 (.000)	.085 (.105)	.402 (.000)	.153 (.004)	-

Table 3

Results of Multiple Regression Analyses of Peer- and Self-Report Measures on Grade 8 Independent School Outcomes.

Variable	Absences β	R^2	ΔR^2	Disciplinary Infractions β	R^2	ΔR^2	Suspensions β	R^2	ΔR^2
Step 1		.005	.005		.062	.062***		.067	.067***
Race	-.064			.241***			.255***		
Gender	.035			-.102			.019		
Step 2		.010	.005		.076	.014		.072	.005
Race	-.062			.235***			.251***		
Gender	.038			-.085			.030		
Peer-Reports	.039			.124*			.074		
Self-Reports	.044			-.087			-.051		

Variable	Overall GPA β	R^2	ΔR^2	Language Arts GPA β	R^2	ΔR^2	Math GPA β	R^2	ΔR^2
Step 1		.138	.138***		.062	.062***		.041	.041**
Race	-.373***			-.202***			-.197**		
Gender	.090			.175**			.075		
Step 2		.148	.010		.067	.005		.063	.023*
Race	-.370***			-.199***			-.198**		
Gender	.076			.165**			.061		
Peer-Reports	-.110*			-.075			-.134*		
Self-Reports	.047			.045			-.034		

Variable	Science GPA β	R^2	ΔR^2	Social Studies GPA β	R^2	ΔR^2	Math SOL Scores β	R^2	ΔR^2
Step 1		.175	.175***		.149	.149***		.019	.019
Race	-.412***			-.389***			-.136		
Gender	-.038			.103			-.002		
Step 2		.180	.005		.151	.002		.076	.057**
Race	-.407***			-.387***			-.141*		
Gender	-.046			.096			-.043		
Peer-Reports	-.066			-.044			-.263***		
Self-Reports	.057			.045			.061		

*p < .05   **p < .01   ***p < .001

Table 3 continued

Variable	Reading SOL Scores			Science SOL Scores			Social Studies SOL Scores		
	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2
Step 1		.098	.098***		.086	.086***		.052	.052
Race	-.315***			-.222***			-.219*		
Gender	.025			-.166**			-.076		
Step 2		.120	.022*		.089	.004		.057	.005
Race	-.310***			-.217***			-.217*		
Gender	.003			-.170**			-.089		
Peer-Reports	-.164**			-.026			-.065		
Self-Reports	.059			.066			.067		

Variable	Aggressive Attitudes Scale			Willingness to Seek Help Scale			Prevalence of Teasing and Bullying Scale		
	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2
Step 1		.080	.080***		.060	.060***		.015	.015
Race	.094			-.135*			.037		
Gender	-.278***			.222***			.112		
Step 2		.125	.045**		.087	.027*		.246	.231***
Race	.109			-.146**			.068		
Gender	-.303***			.243***			.090		
Peer-Reports	-.167**			.135*			-.068		
Self-Reports	.219***			-.167**			.506***		

Variable	Smoking Cigarettes			Drinking Alcohol			Using Marijuana		
	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2
Step 1		.006	.006		.039	.039*		.014	.014
Race	.078			.159*			.089		
Gender	-.044			-.164*			-.104		
Step 2		.009	.003		.041	.002		.013	.002
Race	.082			.156			.092		
Gender	-.051			-.159*			-.110		
Peer-Reports	-.046			.040			-.048		
Self-Reports	.048			-.045			.039		

*p < .05   **p < .01   ***p < .001

Table 3 continued

Variable	Carrying a Weapon			Feeling Unsafe			Fighting in School		
	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2
Step 1		.008	.008		.027	.027		.026	.026
Race	.034			-.122			.088		
Gender	-.092			.145			-.158*		
Step 2		.028	.020		.118	.090***		.098	.072**
Race	.048			-.089			.121		
Gender	-.108			.150			-.162*		
Peer-Reports	-.112			.151*			.062		
Self-Reports	.148*			.207**			.238**		

Variable	Feeling Sad			Considering Suicide		
	β	R^2	ΔR^2	β	R^2	ΔR^2
Step 1		.034	.034		.043	.043*
Race	-.062			-.107		
Gender	.190*			.207**		
Step 2		.165	.131***		.088	.045*
Race	-.015			-.080		
Gender	.164*			.191*		
Peer-Reports	-.105			-.069		
Self-Reports	.396***			.232**		

*p < .05   **p < .01   ***p < .001

Table 4

Results of Multiple Regression Analyses of Peer and Self-Report Measures on Grade 8 School Climate and Youth Risk Behavior Outcomes After Screening for Validity.

Variable	Aggressive Attitudes Scale			Willingness to Seek Help Scale			Prevalence of Teasing and Bullying Scale		
	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2
Step 1		.061	.061***		.028	.028*		.029	.029*
Race	.090			-.093			.081		
Gender	-.242***			.151*			.142*		
Step 2		.103	.041**		.078	.050**		.268	.238***
Race	.097			-.097			.085		
Gender	-.264***			.171**			.130*		
Peer-Reports	-.173**			.154*			-.066		
Self-Reports	.201**			-.241***			.512***		

Variable	Smoking Cigarettes			Drinking Alcohol			Using Marijuana		
	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2
Step 1		.007	.007		.006	.006		.005	.005
Race	.071			.069			.070		
Gender	.033			-.050			-.044		
Step 2		.019	.011		.012	.007		.018	.012
Race	.077			.077			.076		
Gender	.026			-.052			-.052		
Peer-Reports	-.073			-.010			-.082		
Self-Reports	.115			.086			.119		

Variable	Carrying a Weapon			Feeling Unsafe			Fighting in School		
	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2
Step 1		.021	.021		.016	.016		.044	.044*
Race	.081			-.094			.162		
Gender	-.140			.113			-.178*		
Step 2		.030	.010		.105	.089***		.137	.093***
Race	.082			-.068			.190*		
Gender	-.149			.126			-.168*		
Peer-Reports	-.101			.170*			.126		
Self-Reports	.083			.184*			.231**		

*p < .05   **p < .01   ***p < .001

Table 4 continued

Variable	Feeling Sad			Considering Suicide		
	β	R^2	ΔR^2	β	R^2	ΔR^2
Step 1		.024	.024		.031	.031
Race	-.028			-.075		
Gender	.158			.179*		
Step 2		.134	.110***		.065	.034
Race	.000			-.060		
Gender	.149			.172*		
Peer-Reports	-.078			-.063		
Self-Reports	.359***			.204*		

*$p < .05$ **$p < .01$ ***$p < .001$