The Impact of Validity Screening on Associations Between Self-Reports of Bullying Victimization and Student Outcomes

Yuane Jia[1], Timothy R. Konold[1], Dewey Cornell[1], and Francis Huang[2]

Abstract

Self-report surveys are widely used to measure adolescent risk behavior and academic adjustment, with results having an impact on national policy, assessment of school quality, and evaluation of school interventions. However, data obtained from self-reports can be distorted when adolescents intentionally provide inaccurate or careless responses. The current study illustrates the problem of invalid respondents in a sample (N = 52,012) from 323 high schools that responded to a statewide assessment of school climate. Two approaches for identifying invalid respondents were applied, and contrasts between the valid and invalid responses revealed differences in means, prevalence rates of student adjustment, and associations among reports of bullying victimization and student adjustment outcomes. The results lend additional support for the need to screen for invalid responders in adolescent samples.

Keywords

[1]University of Virginia, Charlottesville, VA, USA
[2]University of Missouri–Columbia, MO, USA

Corresponding Author:
Yuane Jia, Curry School of Education, University of Virginia, 417 Emmet Street South, PO Box 400265, Charlottesville, VA 22904-4265, USA.
Email: YJ2SU@Virginia.edu

A large body of research has found that bullying has a negative impact on student adjustment across academic and social-emotional domains (Juvonen & Graham, 2014; Swearer, Espelage, Vaillancourt, & Hymel, 2010). These findings are used to guide interventions, public policy, and research on bullying. However, most of the evidence linking bullying to student adjustment is based on anonymous self-reports that are not screened for response validity. Previous research has found that dishonest and careless responding can have unexpected effects on survey item intercorrelations and lead to erroneous conclusions about the relations between student experiences and student adjustment (Cornell, Klein, Konold, & Huang, 2012; Fan et al., 2006). The purpose of the present study was to examine how survey screening with various indices of validity affects the bullying–maladjustment relationship in a large, state-wide sample of high school students.

Bullying and Student Outcomes

Many studies have documented associations between bullying and academic achievement. A meta-analysis of 33 studies with 29,552 participants revealed a small but statistically significant association between peer victimization and academic achievement (Nakamoto & Schwartz, 2009). Although peer victimization is a somewhat broader concept than bullying, studies (Beran & Li, 2008; Schneider, O'Donnell, Stueve, & Coulter, 2012) have specifically linked the experience of being bullied with lower grades, school attachment, poor concentration, and absenteeism at the individual level (moderate correlations of .17-.43). Additionally, other studies have shown that school-level bullying was significantly associated with math and reading achievement (Konishi, Hymel, Zumbo, & Li, 2010) and was predictive of school passing rates for state-mandated testing (Lacey & Cornell, 2013). The association between bullying and academic achievement is also supported by longitudinal studies (e.g., Juvonen, Wang, & Espinoza, 2010).

Linkages between bullying and social-emotional problems are also well documented. Hawker and Boulton's (2000) meta-analysis characterized the mean effect size of associations between peer victimization and depression as moderate. Notably, correlations were .45 when both variables came from the same informant and .29 when the two variables came from different informants. Klomek, Sourander, and Gould (2010) reviewed 20 years of cross-sectional correlational research between bullying and suicidal behavior and concluded that student victims of bullying are more likely to have suicidal ideation and attempt suicide than nonvictims. Similarly, a meta-analysis by Van Geel, Vedder, and Tanilon (2014a) reported a substantial relationship between peer victimization and both suicide ideation (odds ratio effect size = 2.2) and suicide attempts (odds ratio effect size = 2.6). These are large effects that have strong policy implications for schools concerned about suicide among students who are victims of bullying.

Several studies have found that victims of bullying are more likely to report high-risk behaviors (e.g., weapon carrying, physical fight, substance use, and gang

membership) than nonvictims (Brockenbrough, Cornell, & Loper, 2002; Luk, Wang, & Simons-Morton, 2010; Smalley, Warren, & Barefoot, 2016; Van Geel, Vedder, & Tanilon, 2014b). In addition, Tharp-Taylor, Haviland, and D'Amico (2009) found that middle school students who reported either physical or verbal bullying were more likely to endorse use of alcohol, cigarettes, marijuana, and inhalants.

Self-Report Surveys

One potential limitation of research examining associations between bullying and student adjustment outcomes is the near exclusive reliance on anonymous self-reports that are often unverified by independent sources. There are many understandable reasons to use anonymous self-reports. For example, teachers may be unaware that a student is being bullied, and the student may require the protection of anonymity in order to reveal his or her victimization status (Solberg & Olweus, 2003). At the same time, anonymity may have the effect of reducing accountability and decreasing respondent motivation to answer questions carefully and precisely (Lelkes, Krosnick, Marx, Judd, & Park, 2012). For example, Lelkes et al. (2012) found that ensuring complete anonymity in self-reports resulted in less accurate and honest responses among college students.

The use of self-administered surveys in school settings with peer groups may increase the likelihood of students engaging in invalid responding (Fan et al., 2006). Spirrison, Gordy, and Henley (1996), for example, employed a validity scale to demonstrate that students were more likely to provide inconsistent/invalid responses during in-class administrations than during after-class administrations of their survey. In addition to situational influences on response validity, the truthfulness and accuracy of adolescent self-reports of risk behavior have also been found to be a function of cognitive factors. For instance, adolescents have been found to intentionally under- and over-report difficult-to-recall and sensitive risk behaviors due to social desirability beliefs (Brener, Billy, & Grady, 2003).

Even a small proportion of invalid responders can compromise study findings. For example, the National Study of Adolescent Health (Add Health) self-report survey results revealed that adoption was correlated with smoking, drinking, skipping school, fighting, lying to parents, and other problematic behavior (Miller et al., 2000). However, when researchers later checked in-home interviews, they found that about 19% of the adolescents who claimed to be adopted on the school survey were in fact not (Fan et al., 2002). Group differences diminished or disappeared when data were reanalyzed following screening for invalid respondents. This study demonstrated that even a relatively low rate of overreporting could produce statistically significant group differences and false findings. Another study of the Add Health Survey identified students who made inaccurate claims about their nationality and disability status. These so-called ''jokester'' responders also reported significantly higher rates of risk behaviors (e.g., drinking, skipping school, and fighting) and

lower rates of positive outcomes (e.g., positive school feelings, self-esteem, and school grades) when contrasted with truthful responders (Fan et al., 2006).

Similar invalid responder effects have been reported elsewhere. For example, Robinson and Espelage (2011) used a screening procedure to exclude adolescents who consistently provided unusual or infrequent responses. Contrasts between straight- and transgender-identified adolescents that were statistically significant on a variety of outcomes (e.g., suicide attempts, victimization, and school belongingness) were erased when invalid respondents were removed from the sample.

Cornell et al. (2012) used validity screening items to identify middle school students who answered carelessly or admitted they were not being truthful in their survey responses. Removal of the invalid responders resulted in significantly lower prevalence rates of risk behavior, and the identification of a different factor structure among the items. Moreover, in comparison to invalid responders, valid responders had more positive perceptions of their schools and showed higher associations with teacher perceptions of school climate. A longitudinal study over 3 years of middle school using confidential (not anonymous) surveys found that invalid responders reported higher rates of risk behavior and more negative perceptions of their schools (Cornell, Lovegrove, & Baly, 2014). They were identified from school records as having higher disciplinary infractions.

Methods for Detecting Invalid Responders

A number of methods have been described in the literature for detecting potentially invalid responders. The use of screening items/scales (e.g., I am telling the truth on this survey) is one of the most widely used methods for detecting invalid respondents in psychological assessments. For example, the Minnesota Multiphasic Personality Inventory–2 Restructured Form (MMPI-2-RF) includes several validity scales sensitive to content-based and non–content-based invalid responding behaviors, and procedures that can be used to identify item patterns that might be consistent with invalid responding behaviors (Ben-Porath, 2013; Burchett & Bagby, 2014).

The collection of survey completion time data through computer-based platforms for administering questionnaires has created a new way to screen for invalid responders (Meyer, 2010; Wise & DeMars, 2006). This procedure may work best at detecting non–content-based invalid responding. Non–content-based responding can occur when a respondent is inattentive or noncognitively engaged (Meade & Craig, 2012). Here, participants are unlikely to actually read the questions or to skim them so rapidly that there is insufficient time to provide an accurate reflection of their views or beliefs. Although completion time has been used in a national assessment of academic achievement (Lee & Jia, 2014), our review of the literature found no studies of bullying that examined the use of survey completion time to identify invalid responders. The combination of validity screening items and assessment of survey completion time might provide a more effective way to identify survey data that should be omitted from analyses.

The Current Study

Failure to screen samples for invalid responders has been found to lead to both exaggerated prevalence rates (Cornell et al., 2012; Cornell, Lovegrove, et al., 2014; Furlong, Sharkey, Bates, & Smith, 2008) and erroneous conclusions regarding associations between student conditions (e.g., adoption, disability status) and adjustment (such as drinking, fighting, low self-esteem, low school engagement; Fan et al., 2002, 2006). The current study extends this work by investigating the impact that validity screening has on the associations between reports of bullying victimization and student academic and socio-emotional adjustment. Student academics included measures of their GPA (grade point average) as well as reports of their affective and cognitive engagement in school. The socio-emotional domain was assessed through measures of depression and risk behaviors (e.g., reports of weapon carrying, fighting, and substance abuse). In evaluating these associations, we controlled for student characteristics (i.e., gender, race, and grade level) that have been shown to affect the prevalence of maladjustment (Bauman, Toomey, & Walker, 2013; Dempsey, Haden, Goldman, Sivinski, & Wiens, 2011; Kowalski & Limber, 2013) as well as the correspondence between bullying experiences and student adjustment (Kowalski & Limber, 2013; Reed, Nugent, & Cooper, 2015).

Two methods of identifying invalid responders were used. First, the survey included two validity screening items that have been used in past research on middle school students: ''I am telling the truth on this survey'' and ''How many of the questions on this survey did you answer truthfully?'' (Cornell et al., 2012; Cornell, Lovegrove, et al., 2014). Second, survey completion time was examined to identify surveys that were completed in an improbably brief time.

Method

Sampling and Procedures

Data for the current study were obtained from the Authoritative School Climate Survey, a statewide survey of school climate and safety conditions in Virginia public secondary schools, which was administered to 323 public high schools in the spring of 2014. Schools had two options for sampling students: (a) invite all students to take the survey, with a goal of surveying at least 70% of all eligible students (whole grade option), or (b) use a random number list to select at least 25 students in each grade to take the survey (random sample option). Schools were given these options in order to choose a more or less comprehensive assessment of their students. Schools choosing the random sample option were provided with a random number list along with instructions for selecting students (for more information, see Cornell, Huang, et al., 2014). All students were eligible to participate except those unable to complete the survey because of limited English proficiency or an intellectual or physical disability. The principal sent an information letter to parents of selected students that explained the purpose of the survey and offered them the option to decline participation

(passive consent). All surveys were administered through a secure online Qualtrics platform. Students completed the survey in classrooms under teacher supervision using a set of standard instructions, and each student was provided with a password that was unique to their school.

Forty-five schools using the whole-grade option obtained an estimated participation rate of 82.9% (21,530 of 25,983). In 254 schools using the random sample option, the estimated participation rate was 93.4% (30,482 of 32,631). The overall student participation rate was 88.7% (52,012 student participants from a pool of 58,613 students asked to participate). The current unscreened sample consisted of 52,012 students with 50.3% female. A total of 26.4% of the participants were in Grade 9, 25.8% in Grade 10, 24.7% in Grade 11, and 23.1% in Grade 12. The racial/ethnic breakdown was 55.1% White, 18.3% Black, 11.3% Hispanic, 3.9% Asian, 1.3% American Indian or Alaska Native, and 0.7% Native Hawaiian or Pacific Islander, with an additional 9.5% of students identifying themselves with having more than one ethnic group.

Measures

Bullying Victimization. Bullying victimization was measured with a five-item scale that included global ("I have been bullied at school this year"), physical ("I have been physically bullied or threatened with physical bullying at school this year"), verbal ("I have been verbally bullied at school this year"), social ("I have been socially bullied at school this year"), and cyber ("I have been cyber bullied at school this year") bullying with four response categories (i.e., 1 = *Never*, 2 = *Once or twice*, 3 = *About once per week*, 4 = *More than once per week*; Cornell, Shukla, & Konold, 2015). Prior studies (Baly, Cornell, & Lovegrove, 2014; Branson & Cornell, 2009; Cornell & Brockenbrough, 2004) have shown its correspondence to peer and teacher nominations of victims of bullying and have demonstrated good concurrent and predictive validity of this scale. In the current sample the composite bullying victimization scale yielded a reliability (Cronbach's alpha) estimate of .85 after screening out invalid respondents, as described below. Scale scores ranged from 5 to 20, with higher scores reflecting more bullying experiences.

Academic Achievement (i.e., GPA). The survey asked, "What grades did you make on your last report card?" The seven response options on this item ranged from "Mostly As" to "Mostly Ds and Fs." Student responses were coded so that students with "Mostly As" scored at 4, "Mostly As and Bs" scored at 3.5, "Mostly Bs" scored at 3, and so on, with a response of "Mostly Ds and Fs" scored at 1. The resulting scores were comparable to the standard four-point metric for GPA ($M = 3.06$, $SD = 0.80$).

Engagement. Student engagement in school was measured with six items and grouped into two factors, Affective Engagement (e.g., "I like this school" and "I feel like I belong at this school") and Cognitive Engagement (e.g., "I usually finish my homework" and "I want to learn as much as I can at school"; see Konold et al.,

2014). Each factor was measured by three items with four response categories (1 = *Strongly disagree*, 2 = *Disagree*, 3 = *Agree*, 4 = *Strongly agree*). The total score ranged from 3 to 12. Higher scores reflected greater levels of student engagement at school. A previous study with 39,364 middle school students (Konold et al., 2014) revealed that factor loadings for the Affective and Cognitive engagement scales ranged from .84 to .94 and from .68 to 81, respectively. Cronbach's alphas for the Affective and Cognitive scales in the current sample, after screening out invalid respondents, were .88 and .70, respectively.

Depression. Depression was measured by the six-item Orpinas Modified Depression Scale (Orpinas, 1993). Exemplary items include ''In the last 30 days how often . . . were you sad?'' ''In the last 30 days, how often . . . were you grouchy, irritable, or in a bad mood?'' Each had five categorical response options (i.e., 1 = *Never*, 2 = *Seldom*, 3 = *Sometimes*, 4 = *Often*, 5 = *Always*). Individual scale scores were obtained as the average of the six items. Scale scores ranged from 1 to 5, with higher scores reflecting greater levels of depression. The Modified Depression Scale has been validated in adolescents aged 10 to 18 with a good internal consistency of .74 (Orpinas, 1993). Internal consistency reliability (Cronbach's alphas) in the current sample after removing invalid respondents was .86.

Risk Behavior. The survey included six items from the Youth Risk Behavior Surveillance Survey (YRBS) to measure the prevalence of student risk behavior in the areas of fighting (''During the past 12 months, how many times were you in a physical fight on school property?''), carrying weapons (''During the past 30 days, on how many days did you carry a weapon such as a gun, knife, or club on school property?''), using marijuana (''During the past 30 days, how many times did you use marijuana?''), and consuming alcohol (''During the past 30 days, on how many days did you have at least one drink of alcohol?''). Students were asked about suicide ideation (''During the past 12 months, did you ever seriously consider attempting suicide?''), with a *Yes* or *No* response. They were asked about suicide attempts (''During the past 12 months, how many times did you actually attempt suicide?'') on a 5-point scale (1 = *0 times*, 2 = *1 time*, 3 = *2 or 3 times*, 4 = *4 or 5 times*, 5 = *6 or more times*). As with other studies that have used the YRBS items (e.g., David, May, & Glenn, 2013; Stack, 2014), variables were dichotomized (1 = yes) or not (0 = no) to indicate whether the respondents had engaged in the activity within the past 12 months or the past 30 days.

Validity Screening Items. Two validity screening items were included in the survey in order to identify students who admitted that they were not answering truthfully or who were not taking the survey seriously. One screening item (i.e., ''I am telling the truth on this survey'') was measured on a 4-point scale ranging from *Strongly disagree* to *Strongly agree*, and the second screening item (i.e., ''How many of the questions on this survey did you answer truthfully?'') was measured on a 5-point scale

ranging from *All of them* to *Only a few or none of them*. Students answering S*trongly disagree* or *Disagree* on the first item or students answering *Some of them* or *Only a few or none of them* were classified as invalid responders. Previous studies found that the use of these validity screening items can identify adolescents who give exaggerated reports of risk behavior and more negative perspectives of school climate than other adolescents (Cornell et al., 2012; Cornell, Lovegrove, et al., 2014).

Survey Completion Time. The online Qualtrics platform used to administer the survey was set to record the length of time students took to complete it. Survey completion time varied widely across respondents as some participants failed to close out of the system after completing the survey. A plot of the natural log of survey completion time among those completing the survey in less than 20 minutes revealed a negatively skewed distribution that indicated the presence of a small proportion of students who completed the survey much more quickly than most others. A two-component finite normal mixture model was fit to the natural log of response time to determine the threshold between students who completed the survey too fast and those who took adequate time to complete it. The two-class model was found to provide better fit (Bayesian information criterion = 11126.49) than the one-class model (Bayesian information criterion = 18137.52). Students who completed the survey so rapidly (i.e., less than 6.1 minutes) were found to be statistically anomalous in comparison to the other respondents (Cornell, Huang, et al., 2014). Furthermore, content evaluations by expert reviewers and volunteer readers confirmed that it was highly improbable for respondents to complete the survey below this time point. Among students who completed the survey in more than 6.1 minutes, the median completion time was 14.4 minutes, and approximately 90% of the surveys were completed between 8.3 and 42.8 minutes.

Data Analysis Plan

Invalid responders were identified through two validity screening methods. First, the two screening items (Cornell, Huang, et al., 2014) were used to identify high school students who reported not telling the truth on the survey. This resulted in the identification of $N = 3,579$ students (6.88% of the sample). Second, students identified as completing the survey too rapidly (less than 6.1 minutes) were also characterized as invalid responders ($N = 649$). This resulted in the identification of an additional $N = 406$ students (0.83% of the remaining sample). There were $N = 243$ invalid responders who were identified through both screening methods. In combination, $N = 3,985$ responders were members of the invalid group (7.66% of the total sample), leaving 48,027 in the valid group.

The impact of validity screening on associations between bully victimization and a variety of student adjustment outcomes was examined through a series of hierarchical regression models in which the nesting of students within schools was treated by including schools as a fixed effect (Huang, 2016). Linear regression was used for the

continuous student outcomes (i.e., GPA, affective and cognitive engagement, and depression), and logistic regression was used with the dichotomous outcomes (i.e., weapon carrying, fighting, alcohol use, marijuana use, suicidal thinking, and suicidal attempt). Step 1 of each model considered only the student and school fixed-effect control variables. Associations between bully victimization and the student outcomes were examined in Step 2 of each model. The third step of each model evaluated the effect of using both validity screening approaches on the associations of bully victimization and student outcomes through inclusion of a dichotomously coded validity screening variable (valid responders = 0 and invalid responders = 1) and a Bully victimization × Validity screening interaction term. In addition to evaluating the impact of validity screening through the use of both screening items and response time, we also consider the results that would have been obtained if only one of these two approaches were used. There were no missing data for bully victimization, GPA, affective and cognitive engagement; however, the proportion of missing data in other outcomes (i.e., risk behaviors and depression) was less than 0.58% (with the exception that 1.27% was missing for suicidal thoughts). All analyses were conducted through SPSS and STATA.

Results

Descriptive statistics for each of the student outcomes are presented for the total sample and for the valid and invalid respondent groups in Table 1. Results reveal that inclusion of the invalid responders in the total sample inflated the prevalence of all reported risk behaviors with the exception of suicidal thoughts, and deflated student reports of GPA, school engagement, and depression. In other words, after removing invalid respondents, the overall prevalence rates for risk behaviors were lower than those for the total sample that did not include screening for invalid respondents. Likewise, the sample means for depression, GPA, and school engagement were lower in the total sample in which screening was not used. Standard deviations were also smaller in the valid respondent group across all outcomes when evaluated in relation to the unscreened total sample. Additionally, inflation rates of the sample means were considerably higher for the binary risk behavior outcomes than for the outcomes that were measured on a continuous scale (i.e., GPA, engagement, and depression).

The last two columns in Table 1 present contrasts between the valid and invalid respondent groups for all investigated outcomes. The estimates were obtained through regression models that controlled for gender, race, grade level, and the fixed effect of school. All adjusted mean differences between groups across student outcomes were statistically significant ($p < .001$), with meaningful effect sizes (Cohen's d) that ranged from 0.21 to 0.84 in absolute value. The sole exception was the depression scale where the effect size was smaller ($d = 0.11$).

Associations among demographic student characteristics and respondent group types are presented in Table 2. The invalid group was found to be composed of more

Table 1. Descriptive Statistics and Group Contrasts Between Valid and Invalid Respondent Groups.

	Total sample (N = 52,012)		Valid responders (N = 48,027)		Invalid responders (N = 3,985)		Inflation rate	Adjusted difference	Cohen's d
	M	SD	M	SD	M	SD			
GPA	3.06	0.80	3.08	0.78	2.73	0.93	−0.88%	−0.26***	−0.30
Affective Engagement	8.60	2.19	8.68	2.13	7.67	2.65	−0.89%	−0.83***	−0.34
Cognitive Engagement	9.73	1.78	9.82	1.67	8.57	2.46	−0.97%	−1.14***	−0.54
Depression	2.37	0.98	2.38	0.96	2.20	1.21	−0.42%	−0.12***	−0.11
								Odds ratio	
Weapon Carrying	0.06	0.24	0.05	0.22	0.23	0.42	20.00%	4.56***	0.84
Fighting	0.10	0.30	0.08	0.28	0.27	0.45	25.00%	3.06***	0.62
Alcohol Use	0.27	0.44	0.25	0.44	0.39	0.49	8.00%	1.94***	0.37
Marijuana Use	0.17	0.37	0.15	0.36	0.35	0.48	13.33%	2.72***	0.55
Suicidal Thoughts	0.13	0.34	0.13	0.33	0.17	0.38	0.00%	1.47***	0.21
Suicidal Attempts	0.07	0.26	0.06	0.24	0.19	0.39	16.67%	3.21***	0.64

Note. Adjusted differences and odds ratios are in unstandardized form. They represent contrasts between the valid and invalid groups after controlling for gender, race, grade, and school as a fixed effect.
***p < .001.

Table 2. Gender, Ethnicity, and Grade Comparisons by Respondent Type.

	Valid respondents		Invalid respondents		
	n	%	*n*	%	Chi Square
Gender					305.09***
Female	24,667	51.4%	1,473	37.0%	
Male	23,360	48.6%	2,512	63.0%	
Ethnicity					632.02***
White	27,219	56.7%	1,437	36.1%	
Non-White	20,808	43.3%	2,548	63.9%	
Grade					44.81***
Grade 9	12,518	26.1%	1,225	30.7%	
Grade 10	12,466	26.0%	962	24.1%	
Grade 11	11,937	24.9%	890	22.3%	
Grade 12	11,106	23.1%	908	22.8%	

***$p < .001$.

males ($\chi^2 = 305.09$, $p < .001$), non-Whites ($\chi^2 = 632.02$, $p < .001$), and respondents from younger grades ($\chi^2 = 44.81$, $p < .001$).

Model estimates from the hierarchical regression models that evaluated the influence of validity screening on associations between bully victimization and student outcomes are presented in Table 3. After controlling for student and school characteristics in Step 1 of each model, associations between bully victimization and each of the investigated outcomes were statistically significant in the total sample (all ps $< .001$) and in expected directions. Higher reports of bully victimization were associated with lower GPAs ($B = -0.03$), lower affective ($B = -0.15$) and cognitive ($B = -0.05$) engagement, and higher reported depression ($B = 0.12$); as well as an increased likelihood of carrying a weapon ($B = 0.17$), fighting ($B = 0.17$), using alcohol ($B = 0.10$) and marijuana ($B = 0.11$), and suicidal thoughts ($B = 0.25$) and attempts ($B = 0.26$).

Step 3 introduced the effect of validity group membership as well as an interaction term that allowed for examination of the moderating effect of validity group membership on the relationship between bully victimization and the student outcomes. When invalid responders were identified on the basis of both screening items (SI) and response time (RT), group membership was found to have a statistically significant effect on all student-reported outcomes (all ps $< .001$), and group membership was also found to play a moderating role in the relationship between bullying and all student outcomes (all ps $< .05$) with the exception of student self-reported depression ($p > .05$). Effect sizes ranged from $\Delta R^2 = 1\%$ to 3%; see Step $3_{\text{SI and RT}}$ model results in the middle of Tables 3 and 4.

Associations between bully victimization and the student outcomes by valid and invalid group membership are presented for each outcome in Figure 1, and the moderating role of validity screening can also be seen in the coefficients presented in

Table 3. Linear Regression Coefficients for Predicting Continuous Outcomes of Student Adjustment.

	GPA	Affective Engagement	Cognitive Engagement	Depression
Step 1				
Male	−0.22***	0.11***	−0.54***	−0.43***
Grade	0.03***	−0.10***	−0.07***	0.05***
Non-White	−0.21***	−0.30***	−0.05**	−0.02*
School fixed effects				
R^2	.10***	.12***	.05***	.08***
Step 2				
Bullying Victimization (B)	−0.03***	−0.15***	−0.05***	0.12***
R^2	.11***	.15***	.06***	.18***
ΔR^2	.01***	.03***	.01***	.10***
Step 3$_{SI \ and \ RT}$				
Invalid (I)	−0.24***	−0.75***	−1.07***	−0.20***
B × I	0.03**	0.14***	−0.12***	−0.01
R^2	.12***	.16***	.09***	.19***
ΔR^2	.01***	.01***	.03***	.01***
Step 3$_{SI}$				
Invalid (I)	−0.29***	−0.81***	−1.11***	−0.22***
B × I	0.00	0.24***	0.02	−0.05***
R^2	.12***	.16***	.09***	.19***
ΔR^2	.01***	.01***	.03***	.01***
Step 3$_{RT}$				
Invalid (I)	0.13***	−0.41***	−0.93***	−0.25***
B × I	0.02	−0.02	−0.37***	0.13***
R^2	.11***	.15***	.07***	.18***
ΔR^2	.00	.00	.01***	.00

Note. Invalid = responders identified as invalid group (valid group as reference); B × I = bullying by invalid interaction; SI = invalid responders identified through screening items; RT = invalid responders identified through response time. Coefficients are presented in unstandardized form.
*$p < .05$. **$p < .01$. ***$p < .001$.

Table 3 for the continuous outcomes. For example, a one-unit increase on the bully victimization scale yields a decrease in affective engagement of 0.15 ($B_B = -0.15$) points for the valid group, whereas a one-unit increase in the bully victimization scale yields a decrease in affective engagement of 0.01 for the invalid group (i.e., ($B_B = -0.15$) + ($B_{BxV} = 0.14$) = −0.01). Consequently, as bullying experiences increase, the slope of the invalid group decreases more slowly than that for the valid group (see second graph in Figure 1). The moderating role of validity screening can also be seen in the coefficients obtained from the logistic regression models for the dichotomous outcomes; see Table 4. For example, a one-unit increase in bully victimization yields a log odds change of 0.17 for weapon carrying among valid responders, and a log odds change of (0.17 + 0.08 =) 0.25 among invalid responders. The odds for the
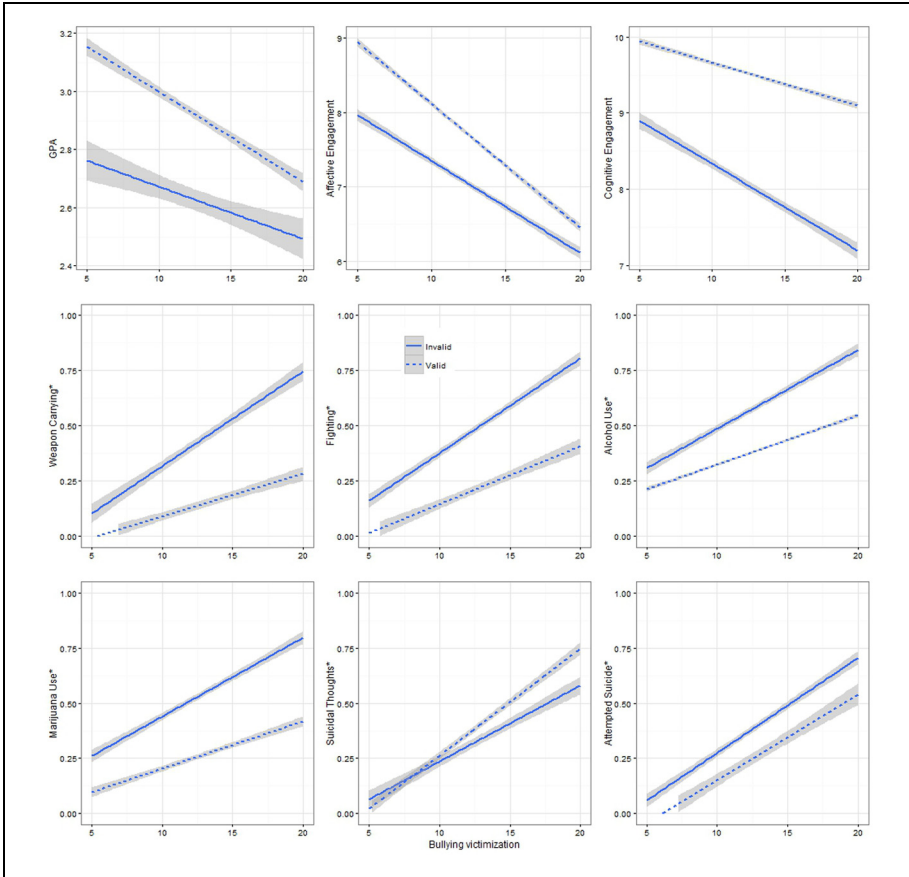
Figure 1. Associations between bully victimization and the predicted student outcomes by valid and invalid group membership.
*Indicates dichotomous outcomes.

valid and invalid groups were ($e^{0.17}$ =) 1.19 and ($e^{0.25}$ =) 1.28, respectively. The corresponding odds ratio (OR) of (1.28/1.19 =) 1.08 indicates that the odds of invalid responders reporting more weapon carrying are 1.08 greater than valid responders. Similar results were obtained for the outcomes of fighting (OR = 1.08), alcohol use (OR = 1.17), and marijuana use (OR = 1.08), in that invalid responders were more likely to report higher levels of risk behaviors (i.e., ORs > 1.0). By contrast, odds ratios for suicidal thoughts (OR = 0.80) and suicidal attempts (OR = 0.86) indicated that invalid responders were less likely to report these outcomes in comparison to valid responders. McFadden's (1978) pseudo-R^2 for the logistic regression models ranged from .06 to .18 across risk behaviors.

Results that would have been obtained by identifying invalid respondents through either screening items or response time alone are shown at the bottom of Tables 3

Table 4. Logistic Regression for Predicting Dichotomous Outcomes of Student Adjustment.

	Weapon Carrying		Fighting		Alcohol Use		Marijuana Use		Suicidal Thoughts		Suicidal Attempts	
	Beta	OR	Beta	OR	Beta	OR	Beta	OR	Beta	OR	Beta	OR
Step 1												
Male	1.01***	2.76	0.82***	2.26	0.09***	1.09	0.30***	1.35	−0.70***	0.50	−0.61***	0.54
Grade	0.11***	1.12	−0.16***	0.85	0.26***	1.30	0.21***	1.23	−0.04**	0.96	−0.08***	0.92
Non-White	0.34***	1.40	0.65***	1.91	−0.14***	0.87	0.30***	1.36	0.17***	1.18	0.46***	1.59
School fixed effects												
R^2	.09***		.08***		.04***		.04***		.04***		.05***	
Step 2												
Bullying (B)	0.17***	1.19	0.17***	1.18	0.10***	1.11	0.11***	1.12	0.25***	1.28	0.26***	1.29
R^2 (ΔR^2)	.15*** (.06***)		.13*** (.05***)		.05*** (.01***)		.07*** (.03***)		.12*** (.08***)		.16*** (.11***)	
Step 3$_{\text{SI and RT}}$												
Invalid (I)	1.36***	3.91	0.98***	2.66	0.56***	1.75	0.91***	2.47	0.28***	1.32	1.11***	3.03
B × I	0.08*	1.08	0.08*	1.08	0.16***	1.17	0.08**	1.08	−0.22***	0.80	−0.15***	0.86
R^2 (ΔR^2)	.18*** (.03***)		.14*** (.01***)		.06*** (.01***)		.08*** (.01***)		.13*** (.01***)		.17*** (.01***)	
Step 3$_{\text{SI}}$												
Invalid (I)	1.37***	3.95	0.99***	2.70	0.60***	1.82	0.94***	2.55	0.31***	1.37	1.13***	3.09
B × I	−0.01	0.99	0.01	1.01	0.06*	1.06	−0.02	0.98	−0.26***	0.77	−0.20***	0.82
R^2 (ΔR^2)	.17*** (.02***)		.14*** (.01***)		.06*** (.01***)		.08*** (.01***)		.13*** (.01***)		.17*** (.01***)	
Step 3$_{\text{RT}}$												
Invalid (I)	1.46***	4.31	0.95***	2.58	0.22*	1.24	0.73***	2.08	0.50***	1.64	1.25***	3.50
B × I	0.20**	1.22	0.19**	1.20	0.46***	1.58	0.30***	1.35	−0.18***	0.84	−0.01	0.99
R^2 (ΔR^2)	.15*** (.00)		.13*** (.00)		.05*** (.00)		.07*** (.00)		.13*** (.01***)		.16*** (.00)	

Note. OR = odds ratio; Invalid = responders identified as invalid group (valid group as reference); B × I = bullying by invalid interaction; SI = invalid responders identified through screening items; RT = invalid responders identified through response time.

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

and 4. Here again, group membership was found to have a statistically significant effect on all reported outcomes (all $ps < .05$) for both screening methods. The use of only screening items was also found to play a moderating role in the relationships between bullying and 5 of the 10 investigated student outcomes ($ps < .05$), and the use of only response time played a moderating role in 7 of the 10 student outcomes ($ps < .05$).

Discussion

Self-report systems of measurement are among the most frequently used tools for data collection in the social sciences. They are used to obtain the incidence and intensity of individual characteristics such as behaviors and personalities (Kooij et al., 2008; Shukla & Wiesner, 2015; Vazire, 2006). Despite their frequency of use, self-reports are rarely cross-checked for accuracy and are susceptible to invalid responses. In the psychoeducational literature, invalid responses are often characterized as resulting from insincere respondents, respondents that choose to purposefully distort (e.g., lie) their responses in order to provide more (or less) favorable ratings of their circumstances (Burchett et al., 2015), rebellious responders that purposefully provide a particular response pattern because they find it amusing (Fan et al., 2006), and careless or rapid responders who are inattentive to the survey items (Meade & Craig, 2012). Failure to identify and remove these respondents from analytic samples prior to analysis has been found to contaminate substantive conclusions regarding the prevalence rates of risk behaviors in younger samples (Cornell et al., 2012; Cornell, Lovegrove, et al., 2014; Fan et al., 2002, 2006). The current study extends this work by investigating the impact that validity screening has on reported prevalence rates of risk behaviors, means of student adjustment, as well as the relationships between bullying victimization and adjustment in an older sample of high school students.

Validity Screening Impact on Prevalence Rates and Reported Means

Consistent with previous research on the impact of validity screening in adolescents (Brener et al., 2003; Fan et al., 2006), the current investigation of high school students revealed that even a relatively small proportion of invalid respondents (7.66% of the total high school sample) had a significant impact on the reported prevalence rates of risk behavior and means of student adjustment outcomes when both screening items (SI) and response time (RT) were used to identify invalid responders. In some instances, invalid respondents inflated the prevalence rates of risk behaviors. For example, the prevalence rate for physical fighting increased from 8% in the valid responder group to10% in the invalid responder group, which is an inflation rate of 25%. By contrast, the invalid responder group deflated the sample mean of the continuous outcomes that were investigated (i.e., GPA, engagement, and depression), with relatively smaller deflation rates that raged from 0.42% to 0.97%. However, when these same between group contrasts were made after controlling for student-

and school-level covariates, the prevalence rates of risk behavior among invalid responders were significantly higher, as has been reported elsewhere in examinations of middle school students (Cornell et al., 2012; Cornell, Lovegrove, et al., 2014), than that of valid responders across all outcomes. A large effect size was obtained for weapon carrying; moderate effect sizes emerged for the outcomes of physical fighting, marijuana use, and suicidal attempts; and small effect sizes were observed for alcohol use and suicidal thoughts. Effect sizes were small to moderate for the academic and depressive outcomes, where reported means were smaller for the invalid responder group. In addition, variances across all investigated outcomes were significantly larger for the invalid respondent group, indicating a greater degree of heterogeneity in response patterns. In the aggregate, invalid responders were more likely to report higher levels of risk behavior and lower academic performance and depression than those students not identified as being invalid responders. One interpretation of these findings is that the inflation due to invalid responding was greatest for the most extreme or unusual behaviors (such as bringing a weapon to school) that have the lowest base rate in the general population. For outcomes such as GPA and depression, which vary substantially in the general population, there was a relatively small effect of invalid responding. This suggests that researchers should be most concerned about invalid responding when they are investigating behaviors with low base rates, as has also been reported elsewhere in investigations of self-reported suicidal attempts (Hom, Joiner, & Bernert, 2015; Plöderl, Kralovec, Yazdi, & Fartacek, 2011).

Validity Screening Impact on Associations Among Variables

In addition to differences that emerged with respect to reported prevalence rates or means, the impact of failing to screen for invalid responders is also likely to have an influence on relationships among variables of substantive interest that often form the basis for theory development and intervention research. In the current study, meaningful differences emerged when associations between bullying victimization and a variety of student adjustment outcomes were separately examined in the SI and RT invalid and valid groups. Consistent with previous research, bully victimization experiences were positively associated with the probability of involvement in risk behaviors (e.g., Brockenbrough et al., 2002; Klomek et al., 2010; Luk et al., 2010; Smalley et al., 2016; Van Geel et al., 2014a, 2014b) and depression (e.g., Hawker & Boulton, 2000; Reed et al., 2015), and negatively associated with student-reported GPA, cognitive engagement, and affective engagement in school (e.g., Beran & Li, 2008; Schneider et al., 2012) after controlling for student- and school-level covariates. However, these associations were moderated by valid versus SI and RT invalid group membership. Both the validity screening main effects and its interaction with bullying victimization were significant across all investigated outcomes, with the exception of depression.

As illustrated in Figure 1, the magnitudes of the differences between the valid and invalid groups were dependent on the level of reported bullying victimization. In most instances, larger group differences were observed for higher levels of bullying victimization experiences. In other words, invalid high school respondents that report being exposed to more bully victimization experiences have a much higher predicted probability of also claiming risk behavior (e.g., weapon carrying, physical fighting, alcohol use, and marijuana use) and much lower cognitive engagement and suicidal thoughts than the valid group responders. On the other hand, for those respondents who report lower levels of bully victimization, the differences between the two groups' predicted probabilities are not as pronounced. Two exceptions to these patterns were evident for the outcomes of GPA and affective engagement, where more pronounced differences between the valid and invalid groups were observed at lower levels of reported bullying victimization. One interpretation of these findings is that the invalid responders are not a homogenous group, but include a group of adolescents who tend to endorse both bully victimization and risk behaviors, but are less likely to endorse that it had an impact on their grades or engagement in school. In contrast, the valid responders who are victims of bullying may be registering a marked decline in GPA and positive feelings toward school.

The results of our analyses also shed light on the differential nature of cognitive and affective engagement in relation to bully victimization. Among invalid responders the relationship (i.e., slope) between bully victimization was largely the same for the affective and cognitive outcomes. However, the slope was much greater, and the relationship more pronounced, for bully victimization and affective engagement than for bully victimization and cognitive engagement, among valid responders. This pattern suggests that a student's affective attachment to school may be more vulnerable to bullying experiences.

Finally, student demographics (i.e., gender, grade, race) and the school fixed effects were all significantly associated with the adjustment outcomes of high school students, accounting for 5% to 12% of the total variance in student academic outcomes and depression. Consistent with past research, the invalid group was composed of significantly greater proportions of males, non-Whites, and earlier grade levels (Cornell et al., 2012).

Implications for Survey Research

The YRBS has been widely used in school violence research, but lacks this kind of mechanism to screen for potentially invalid respondents (e.g., Reed et al., 2015). Furlong et al. (2008) criticized the limited empirical evidence of reliability and validity of the YRBS items and concluded that including extreme response patterns in the YRBS sample would compromise the integrity of the respondent sample. By replicating the previous research findings in high school students, this study demonstrated the importance of validity screening in self-report surveys of adolescents. In the absence of screening, our results show the potential for researchers to reach

erroneous conclusions about the relationships between bullying victimization and other student outcomes. Future research employing self-report surveys with adolescents should consider incorporating some form of validity screening. This might include use of built-in validity screening items (Cornell et al., 2012; Cornell, Lovegrove, et al., 2014), inclusion of a validity scale consisting of rarely endorsed items (Furlong, Fullchange, & Dowdy, 2016; Goodwin, Sellbom, & Arbisi, 2013), tracking the response time (Lee & Jia, 2014; Meyer, 2010) with computer-administered surveys, or collection of external validity evidence on responses (Fan et al., 2002, 2006), to name a few. Results of our analyses revealed that the use of only one of the two investigated screening methods would be beneficial to researchers conducting self-report studies. When examined separately, both screening methods were statistically related to all student adjustment outcomes, and played a moderating role in many of the investigated associations between bullying and student adjustment.

More recently, Robinson-Cimpian (2014) proposed a sensitivity analysis to handle potential between-group disparity estimation bias in the situation that researchers have already collected survey data and other screening mechanisms are not feasible. Similarly, researchers recently proposed the use of latent class analysis on response inconsistency variables as a mean for identifying invalid responders (Shukla & Konold, under review); the virtue of this method is that it does not require the use of validity items, response time, or other external criteria be built into the survey design.

Limitation and Future Directions

Although built-in screening items could detect a considerable number of adolescents who are either willing to admit that they are not telling the truth or are answering the survey carelessly, it would not be able to detect adolescents who intentionally give distorted answers and do not admit it on the validity screen questions (Cornell et al., 2012). The time completion screening procedure will capture speedy completers, but not content-based invalid responders. Nevertheless, the procedures used in this current study were able to demonstrate differences between valid and invalid responders that are noteworthy. Future research should consider different screening approaches and their impact on substantive results. Future research is also needed to disentangle the different types of invalid responders that may be present and investigate their potential for producing differential substantial results. Finally, the anonymous survey makes it impossible to validate the adolescent's reports from outside resources. Therefore, we were unable to verify the extent to which invalid responders were accurately identified in the current study.

As for the associations between invalid responders and higher prevalence rates of risk behavior and lower academic performance, further studies with external evidence are needed to determine whether the invalid responders do engage in high-risk behavior or are simply claiming to do so. Likewise, we could not establish the causal direction of bullying victimization and adjustment outcomes due to the cross-

sectional and correlational nature of the study; that is, the study cannot determine whether involvement in risk behavior causes more bullying victimization or bullying victimization leads to more risk behaviors involvement and poor academic performance.

Authors' Note

Acknowledgments

Declaration of Conflicting Interests

Funding

References

Baly, M., Cornell, D., & Lovegrove, P. (2014). A longitudinal investigation of self- and peer reports of bullying victimization across middle school. *Psychology in the Schools*, *51*, 217-240. doi:10.1002/pits.21747

Bauman, S., Toomey, R. B., & Walker, J. L. (2013). Associations among bullying, cyberbullying, and suicide in high school students. *Journal of Adolescence*, *36*, 341-350. doi:10.1016/j.adolescence.2012.12.001

Ben-Porath, Y. S. (2013). Forensic applications of the Minnesota Multiphasic Personality Inventory–2 Restructured Form. In R. P. Archer & E. M. A. Wheeler (Eds.), *Forensic use of clinical assessment instruments* (pp. 63-107). New York, NY: Routledge.

Beran, T., & Li, Q. (2008). The relationship between cyberbullying and school bullying. *Journal of Student Wellbeing*, *1*, 16-33.

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

Brener, N. D., Billy, J. O., & Grady, W. R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *Journal of Adolescent Health*, *33*, 436-457. doi:10.1016/S1054-139X(03)00052-1

Brockenbrough, K. K., Cornell, D., & Loper, A. B. (2002). Aggressive attitudes among victims of violence at school. *Education and Treatment of Children*, *25*, 273-287. Retrieved from http://www.jstor.org/stable/42899706

Burchett, D., & Bagby, R. M. (2014). Multimethod assessment of response distortion: Integrating data from interviews, collateral records, and standardized assessment tools. In C. Hopwood & R. Bornstein (Eds.), *Multimethod clinical assessment* (pp. 345-378). New York, NY: Guilford.

Burchett, D., Dragon, W. R., Smith Holbert, A. M., Tarescavage, A. M., Mattson, C. A., Handel, R. W., & Ben-Porath, Y. S. (2015). ''False feigners'': Examining the impact of non-content-based invalid responding on the Minnesota Multiphasic Personality Inventory–2 Restructured Form content-based invalid responding indicators. *Psychological Assessment*. Advance online publication. doi:10.1037/pas0000205

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., Huang, F., Konold, T., Meyer, P., Shukla, K., Heilbrun, A., . . . Nekvasil, E. (2014). *Technical report of the Virginia Secondary School Climate Survey: 2014 results for 9th-12th grade students and teachers*. Charlottesville: University of Virginia, Curry School of Education.

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., Lovegrove, P. J., & Baly, M. W. (2014). Invalid survey response patterns among middle school students. *Psychological Assessment*, *26*, 277-287. doi:10.1037/a0034808

Cornell, D., Shukla, K., & Konold, T. (2015). Peer victimization and authoritative school climate: A multilevel approach. *Journal of Educational Psychology*, *107*, 1186-1201. doi:10.1037/edu0000038

David, E., May, A. M., & Glenn, C. R. (2013). The relationship between nonsuicidal self-injury and attempted suicide: Converging evidence from four samples. *Journal of Abnormal Psychology*, *122*, 231-237. doi:10.1037/a0030278

Dempsey, A. G., Haden, S. C., Goldman, J., Sivinski, J., & Wiens, B. A. (2011). Relational and overt victimization in middle and high schools: Associations with self-reported suicidality. *Journal of School Violence*, *10*, 374-392. doi:10.1080/15388220.2011.602612

Fan, X., Miller, B., Christensen, M., Bayley, B., Park, K., Grotevant, H., . . . 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

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

Furlong, M. J., Fullchange, A., & Dowdy, E. (2016). Effects of mischievous responding on universal mental health screening: I love rum raisin ice cream, really I do! *School Psychology Quarterly*. Advance online publication. doi:10.1037/spq0000168

Furlong, M., Sharkey, J., Bates, M. P., & Smith, D. (2008). 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

Goodwin, B. E., Sellbom, M., & Arbisi, P. A. (2013). Posttraumatic stress disorder in veterans: The utility of the MMPI-2-RF Validity Scales in detecting overreported symptoms. *Psychological Assessment*, *25*, 671-678. doi:10.1037/a0032214

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 and Allied Disciplines*, *41*, 441-455. doi:10.1111/1469-7610.00629

Hom, M. A., Joiner, T. E., Jr., & Bernert, R. A. (2015). Limitations of a single-item assessment of suicide attempt history: Implications for standardized suicide risk assessment. *Psychological Assessment*, *28*, 1026-1030. doi:10.1037/pas0000241

Huang, F. L. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *Journal of Experimental Education*, *84*, 175-196. doi:10.1080/00220973.2014.952397

Juvonen, J., & Graham, S. (2014). Bullying in schools: The power of bullies and the plight of victims. *Annual Review of Psychology*, *65*, 159-185. doi:10.1146/annurev-psych-010213-115030

Juvonen, J., Wang, Y., & Espinoza, G. (2010). Bullying experiences and compromised academic performance across middle school grades. *Journal of Early Adolescence*, *31*, 152-173. doi:10.1177/0272431610379415

Klomek, A. B., Sourander, A., & Gould, M. (2010). The association of suicide and bullying in childhood to young adulthood: A review of cross-sectional and longitudinal research findings. *Canadian Journal of Psychiatry*, *55*, 282-288. doi:10.1177/070674371005500503

Konishi, C., Hymel, S., Zumbo, B. D., & Li, Z. (2010). Do school bullying and student–teacher relationships matter for academic achievement? A multilevel analysis. *Canadian Journal of School Psychology*, *25*, 19-39. doi:10.1177/0829573509357550

Konold, T., Cornell, D., Huang, F., Meyer, P., Lacey, A., Nekvasil, E., . . . Shukla, K. (2014). Multilevel multi-informant structure of the Authoritative School Climate Survey. *School Psychology Quarterly*, *29*, 238-255. doi:10.1037/spq0000062

Kooij, J. S., Boonstra, A. M., Swinkels, S. H. N., Bekker, E. M., de Noord, I., & Buitelaar, J. K. (2008). Reliability, validity, and utility of instruments for self-report and informant report concerning symptoms of ADHD in adult patients. *Journal of Attention Disorders*, *11*, 445-458.

Kowalski, R., & Limber, S. (2013). Psychological, physical, and academic correlates of cyberbullying and traditional bullying. *Journal of Adolescent Health*, *53*, 513-520. doi:10.1016/j.jadohealth.2012.09.018

Lacey, A., & Cornell, D. (2013). The impact of teasing and bullying on schoolwide academic performance. *Journal of Applied School Psychology*, *29*, 262-283. doi:10.1080/15377903.2013.806883

Luk, J. W., Wang, J., & Simons-Morton, B. G. (2010). Bullying victimization and substance use among U.S. adolescents: Mediation by depression. *Prevention Science*, *11*, 355-359. doi:10.1007/s11121-010-0179-0

Lee, Y. H., & Jia, Y. (2014). Using response time to investigate students' test-taking behaviors in a NAEP computer-based study. *Large-scale Assessments in Education*, *2*, 1-24. doi:10.1186/s40536-014-0008-1

Lelkes, Y., Krosnick, J. A., Marx, D. M., Judd, C. M., & Park, B. (2012). Complete anonymity compromises the accuracy of self-reports. *Journal of Experimental Social Psychology*, *48*, 1291-1299. doi:10.1016/j.jesp.2012.07.002

McFadden, D. (1978). Quantitative methods for analyzing travel behaviour on individuals: Some recent developments. In D. Hensher & P. Stopher (Eds.), *Bahvioural travel modelling* (pp. 279-318). London, England: Croom Helm.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*, 437-455. doi:10.1037/a0028085

Meyer, J. P. (2010). A mixture Rasch model with item response time components. *Applied Psychological Measurement*, *37*, 521-538. doi:10.1177/0146621609355451

Miller, B. C., Fan, X., Grotevant, H. D., Christensen, M., Coyl, D., & van Dulmen, M. (2000). Adopted adolescents' overrepresentation in mental health counseling: Adoptees' problems or parents' lower threshold for referral? *Journal of the American Academy of Child & Adolescent Psychiatry*, *39*, 1504-1511. doi:10.1097/00004583-200012000-00011

Nakamoto, J., & Schwartz, D. (2009). 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

Orpinas, P. (1993). *Modified Depression Scale*. Houston: University of Texas Health Science Center at Houston.

Plöderl, M., Kralovec, K., Yazdi, K., & Fartacek, R. (2011). A closer look at self-reported suicide attempts: False positives and false negatives. *Suicide and Life-Threatening Behavior*, *41*, 1-5. doi:10.1111/j.1943-278X.2010.00005.x

Reed, K. P., Nugent, W., & Cooper, R. L. (2015). Testing a path model of relationships between gender, age, and bullying victimization and violent behavior, substance abuse, depression, suicidal ideation, and suicide attempts in adolescents. *Children and Youth Services Review*, *55*, 128-137. doi:10.1016/j.childyouth.2015.05.016

Robinson, J. P., & Espelage, D. L. (2011). Inequities in educational and psychological outcomes between LGBTQ and straight students in middle and high school. *Educational Researcher*, *40*, 315-330. doi:10.3102/0013189X11422112

Robinson-Cimpian, J. P. (2014). Inaccurate estimation of disparities due to mischievous responders several suggestions to assess conclusions. *Educational Researcher*, *43*, 171-185. doi:10.3102/0013189X14534297

Schneider, S. K., O'Donnell, L., Stueve, A., & Coulter, R. W. (2012). Cyberbullying, school bullying, and psychological distress: A regional census of high school students. *American Journal of Public Health*, *102*, 171-177. doi:10.2105/AJPH.2011.300308

Shukla, K., & Konold, T.R. (under review). Identifying invalid respondents in self-reports through latent profile analysis.

Shukla, K. D., & Wiesner, M. (2015). Direct and indirect violence exposure: relations to depression for economically disadvantaged ethnic minority mid-adolescents. *Violence and Victims*, *30*, 120-135. doi:10.1891/0886-6708.VV-D-12-00042

Smalley, K. B., Warren, J. C., & Barefoot, K. N. (2016). Connection between experiences of bullying and risky behaviors in middle and high school students. *School Mental Health*. Advance online publication. doi:10.1007/s12310-016-9194-z

Spirrison, C. L., Gordy, C. C., & Henley, T. B. (1996). After-class versus in-class data collection: Validity issues. *Journal of Psychology*, *130*, 635-644. doi:10.1080/00223980 .1996.9915037

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

Stack, S. (2014). Differentiating suicide ideators from attempters: A research note. *Suicide and Life-Threatening Behavior*, *44*, 46-57. doi:10.1111/sltb.12054

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

Tharp-Taylor, S., Haviland, A., & D'Amico, E. J. (2009). Victimization from mental and physical bullying and substance use in early adolescence. *Addictive Behaviors*, *34*, 561-567. doi:10.1016/j.addbeh.2009.03.012

Van Geel, M., Vedder, P., & Tanilon, J. (2014a). Relationship between peer victimization, cyberbullying, and suicide in children and adolescents: A meta-analysis. *JAMA Pediatrics*, *168*, 435-442. doi:10.1001/jamapediatrics.2013.4143

Van Geel, M., Vedder, P., & Tanilon, J. (2014b). Bullying and weapon carrying: A meta-analysis. *JAMA Pediatrics*, *168*, 714-720. doi:10.1001/jamapediatrics.2014.213

Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality*, *40*, 472-481. doi:10.1016/j.jrp.2005.03.003

Wise, S. L., & DeMars, C. E. (2006). An application of item response time: The effort-moderated IRT model. *Journal of Educational Measurement*, *43*, 19-38. doi: 10.1111/j.1745-3984.2006.00002